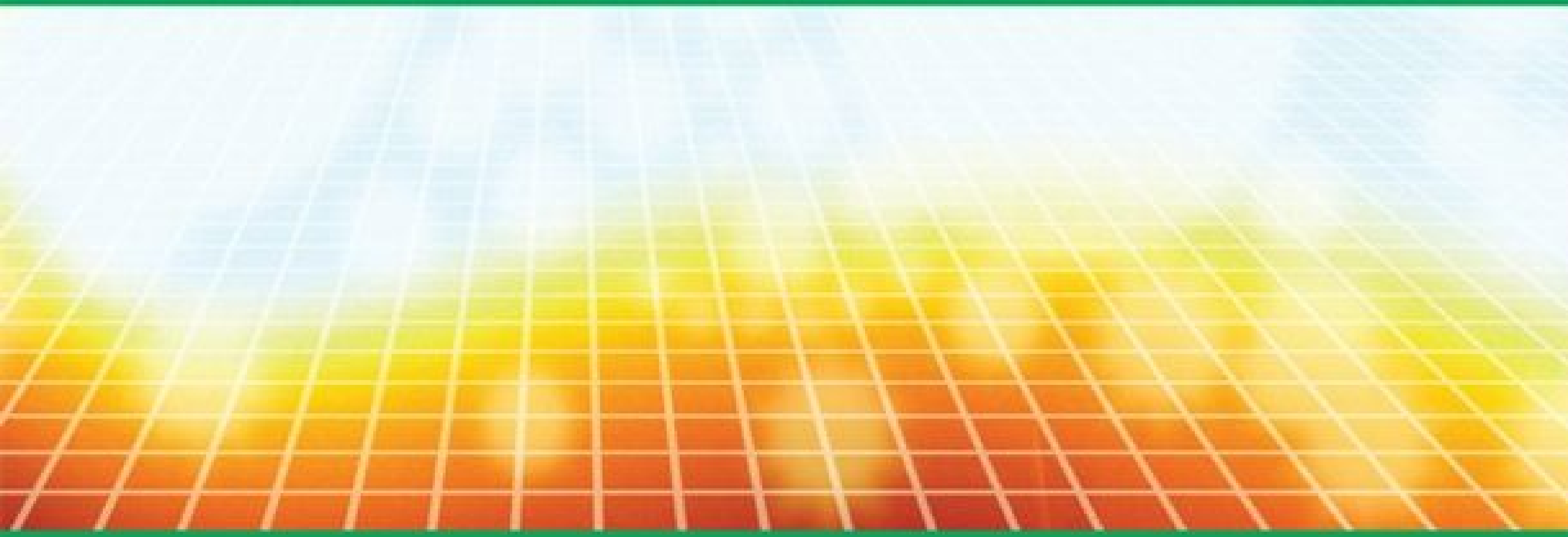
DSP

DIGITAL SIGNAL AND IMAGE PROCESSING SERIES



Information Fusion in Signal and Image Processing

Major Probabilistic and Non-Probabilistic Numerical Approaches

Edited by Isabelle Bloch

ISTE

WILEY

This page intentionally left blank

Information Fusion in Signal and Image Processing

This page intentionally left blank

Information Fusion in Signal and Image Processing

Major Probabilistic and Non-probabilistic Numerical Approaches

Edited by
Isabelle Bloch

ISTE

WILEY

First published in France in 2003 by Hermes Science/Lavoisier entitled "Fusion d'informations en traitement du signal et des images"
First published in Great Britain and the United States in 2008 by ISTE Ltd and John Wiley & Sons, Inc.

ISTE Ltd
6 Fitzroy Square
London W1T 5DX
UK
www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA
www.wiley.com

Library of Congress Cataloging-in-Publication Data

[Fusion d'informations en traitement du signal et des images English] Information fusion in signal and image processing / edited by Isabelle Bloch.
p. cm.
Includes index.
ISBN 978-1-84821-019-6
1. Signal processing. 2. Image processing. I. Bloch, Isabelle.
TK5102.5.I49511 2008
621.382'2--dc22
2007018231

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN: 978-1-84821-019-6

Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire.

Table of Contents

Preface

Over the past few years, the field of information fusion has gone through considerable and rapid change. While it is difficult to write a book in such a dynamic environment, this book is justified by the fact that the field is currently at a turning point. After a phase of questions, debates, and even mistakes, during which the field of fusion in signal and image processing was not well defined, we are now able to efficiently use basic tools (often imported from other fields) and it is now possible to both design entire applications, and develop more complex and sophisticated tools. Nevertheless, there remains much theoretical work to be done in order to broaden the foundations of these methods, as well as experimental work to validate their use.

The objectives of this book are to present, on the one hand, the general ideas of fusion and its specificities in signal and image processing and in robotics, and on the other hand, the major methods and tools, which are essentially numerical. This book does not intend, of course, to compete with those devoted entirely to one of these tools, or one of these applications, but instead tries to underline the assets of the different theories in the intended application fields.

With a book like this one, we cannot aspire to be comprehensive. We will not discuss methods based on expert or multi-agent systems (however, an example will be given to illustrate them), on neural networks and all of the symbolic methods expressed in logical formalism. Several teams work on developing such methods, for example, in France, the IRIT in Toulouse and the CRIL in Lens on logical methods, the LAAS in Toulouse on neuromimetic methods, the IMAG in Grenoble on multi-agent systems, and many others. Likewise, among the methods we will discuss, many interesting aspects will have to be left aside, whether theoretical, methodological or regarding applications because they would bring the reader beyond the comparative context we want him to stay in, but we hope that the cited references will help complete this presentation for readers who would wish to study these aspects further.

This book is meant essentially for PhD students, researchers or people in the industry, who wish to familiarize themselves with the concepts of fusion and discover its main theories. It can also serve as a guide to understanding theories and methodologies, developing new applications, discovering new research subjects, for example, those suggested by the problems and prospects mentioned in this book.

The structure is organized in two sets of chapters. The first deals with definitions (Chapter 1) and the specificities of the fields that are discussed: signal processing in Chapter 2, image processing in Chapter 3 and robotics in Chapter 4. The second part is concerned with the major theories of fusion. After an overview of the modes of knowledge representation used in fusion (Chapter 5), we present the principles of probabilistic and statistical fusion in Chapter 6, of belief function theory in Chapter 7, of fuzzy and possibilistic fusion in Chapter 8. The specificities of fusion in image processing and in certain robotics problems require taking into account spatial information. This is discussed in Chapter 9, since the fusion methods developed in other fields do not consider it naturally. An example of an application that relies on a multi-agent architecture is given in Chapter 10. The specific methods of temporal fusion, finally, are described in Chapter 11.

This book owes a great deal to the GDR-PRC ISIS and to their directors, Odile Macchi and Jean-Marc Chassery. Its authors were the coordinators of the workgroup on information fusion and the related actions. The GDR was the first initiative that led to bringing together the French community of people working on information fusion in signal and image processing, to build ties with other communities (man-machine communications, robotics and automation, artificial intelligence), to enrich ideas and it thus became the preferred place for discussion. This book would not have existed without the maturity acquired in this group. This book is also indebted to the comments and discussions of the FUSION Working Group (a European project) directed by Professor Philippe Smets (IRIDIA, Université Libre de Bruxelles), aimed at summarizing the problems and methods of data fusion in different fields, from artificial intelligence to image processing, from regulations to financial analysis, etc. It grouped together researchers from the IRIT in Toulouse, the IRIDIA in Brussels, Télécom-Paris, the CNR in Padua, the University of Granada, the University of Tunis, the University of Magdeburg, the ONERA, Thomson-CSF, Delft University, University College London. Chapter 1 in particular owes much to this group. Finally, the trust bestowed on us by Bernard Dubuisson, his motivation and his encouragements also helped a great deal in the completion of this book. This book is dedicated to the memory of Philippe Smets.

Isabelle BLOCH

Chapter 1

Definitions

1.1. Introduction

Fusion has become an important aspect of information processing in several very different fields, in which the information that needs to be fused, the objectives, the methods, and hence the terminology, can vary greatly, even if there are also many analogies. The objective of this chapter is to specify the context of fusion in the field of signal and image processing, to specify the concepts and to draw definitions. This chapter should be seen as a guide for the entire book. It should help those with another vision of the problem to find their way.

1.2. Choosing a definition

In this book, the word "information" is used in a broad sense. In particular, it covers both data (for example, measurements, images, signals, etc.) and knowledge (regarding the data, the subject, the constraints, etc.) that can be either generic or specific.

The definition of information fusion that we will be using throughout this book is given below.

DEFINITION 1.1 (Fusion of information). *Fusion of information consists of combining information originating from several sources in order to improve decision making.*

Chapter written by Isabelle BLOCH and Henri MAÎTRE.

This definition, which is largely the result of discussions led within the GDR-PRC ISIS[1] workgroup on information fusion, is general enough to encompass the diversity of fusion problems encountered in signal and image processing. Its appeal lies in the fact that it focuses on the combination and decision phases, i.e. two operations that can take different forms depending on the problems and applications.

For each type of problem and application, this definition can be made more specific by answering a certain number of questions: what is the objective of the fusion? what is the information we wish to fuse? where does it come from? what are its characteristics (uncertainty, relation between the different pieces of information, generic or factual, static or dynamic, etc.)? what methodology should we choose? how can we assess and validate the method and the results? what are the major difficulties, the limits?, etc.

Let us compare this definition with those suggested by other workgroups that have contributed to forming the structure of the field of information fusion.

Definition 1.1 is a little more specific than that suggested by the European workgroup FUSION [BLO 01], which worked on fusion in several fields from 1996 to 1999[2]. The general definition retained in this project is the following: gathering information originating from different sources and using the gathered information to answer questions, make decisions, etc. In this definition, which also focuses on the combination and on the goals, the goals usually stop before the decision process, and are not restricted to improving the overall information. They include, for example, obtaining a general perspective, typically in problems related to fusing the opinions or preferences of people, which is one of the themes discussed in this project, but this goes beyond the scope of this book. Here, improving knowledge refers to the world as it is and not to the world as we would like it to be, as is the case with preference fusion.

Some of the first notable efforts in clarifying the field were made by the data fusion work group at the US Department of Defense's Joint Directors of Laboratories (JDL). This group was created in 1986 and focused on specifying and codifying the terminology of data fusion in some sort of dictionary (Data Fusion Lexicon) [JDL 91]. The method suggested was exclusively meant for defense applications (such as automatically tracking, recognizing and identifying targets, battlefield surveillance) and focused on functionalities, by identifying processes, functions and techniques [HAL 97]. It emphasized the description of a hierarchy of steps in processing a system. The definition we use here contrasts with the JDL's definition and chooses another perspective, focusing more on describing combination and decision

1. www-isis.enst.fr.

2. This chapter greatly benefited from the discussions within this workgroup and we wish to thank all of the participants.

methods rather than systems. It is better suited to the diversity of situations encountered in signal and image processing. In this sense, it is a broader definition.

Another European workgroup of the EARSeL (European Association of Remote Sensing Laboratories) extended the JDL's definition to the broader field of satellite imagery [WAL 99]: the fusion of data constitutes a formal framework in which the data originating from different sources can be expressed; its goal is to obtain information of higher quality; the exact definition of "higher quality" will depend on the application. This definition encompasses most of the definitions suggested by several authors in satellite imagery, which are gathered in [WAL 99]. Definition 1.1 goes further and includes decisions.

The meaning of the word fusion can be understood on different levels. Other concepts, such as estimation, revision, association of data and data mining, can sometimes be considered as fusion problems in a broad sense of the word. Let us specify these concepts.

Fusion and estimation. The objective of estimation is to combine several values of a parameter or a distribution, in order to obtain a plausible value of this parameter. Thus, we have the same combination and decision steps, which are the two major ingredients of Definition 1.1. On the other hand, numerical fusion methods often require a preliminary step to estimate the distributions that are to be combined (see section 1.5) and the estimation is then interpreted as one of the steps of the fusion process.

Fusion and revision or updating. Revising or updating consists of completing or modifying an element of information based on new information. It can be considered as one of the fields of fusion. Sometimes, fusion is considered in a stricter sense, where combination is symmetric. As for revision, it is not symmetric and it draws a distinction between information known beforehand and new information. Here, we will be considering dynamic processes among others (particularly robotics), and it seems important for us to include revision and updating as part of fusion (for example, for applications such as helping a robot comprehend its environment). Revision involves the addition of new information that makes it possible to modify, or specify, the information previously available about the observed phenomenon, whereas updating involves a modification of the phenomenon that leads to modifying the information about it (typically in a time-based process).

Fusion and association. Data association is the operation that makes it possible to find among different signals originating from two sources or more those that are transmitted by the same object (source or target). According to Bar-Shalom and Fortman [BAR 88], data association is the most difficult step in multiple target tracking. It consists of detecting and associating noisy measurements, the origins of which are unknown because of several factors, such as random false alarms in detections, clutter, interfering targets, traps and other countermeasures. The main models used in this field are either deterministic (based on classic hypothesis tests), or probabilistic models (essential Bayesian) [BAR 88, LEU 96, ROM 96]. The most common method [BAR 88] relies on the Kalman filter with a Gaussian hypothesis. More recently, other estimation methods have been suggested, such as the Interactive Multiple Model estimator (IMM), which can adapt to different types of motion and reduce noise, while preserving a good accuracy in estimating states [YED 97]. This shows how the problems we come across can be quite different from those covered by Definition 1.1.

Fusion and data mining. Data mining consists of extracting relevant parts of information and data, which can be, for example, special data (in the sense that it has specific properties), or rare data. It can be distinguished from fusion that tries to explain where the objective is to find general trends, or from fusion that tries to generalize and lead to more generic knowledge based on data. We will not be considering data mining as a fusion problem.

1.3. General characteristics of the data

In this section, we will briefly describe the general characteristics of the information we wish to fuse, characteristics that have to be taken into account in a fusion process. More detailed and specific examples will be given for each field in the following chapters.

A first characteristic involves the type of information we wish to fuse. It can consist of direct observations, results obtained after processing these observations, more generic knowledge, expressed in the form of rules for example, or opinions of experts. This information can be expressed either in numerical or symbolic form (see section 1.4). Particular attention is needed in choosing the scale used for representing the information. This scale should not necessarily have any absolute significance, but it at least has to be possible to compare information using the scale. In other words, scales induce an order within populations. This leads to properties of commensurability, or even of normalization.

The different levels of the elements of information we wish to fuse are also a very important aspect. Usually, the lower level (typically the original measurements) is distinguished from a higher level requiring preliminary steps, such as processing, extracting primitives or structuring the information. Depending on the level, the constraints can vary, as well as the difficulties. This will be illustrated, for example, in the case of image fusion in Chapter 3.

Other distinctions in the types of data should also be underlined, because they give rise to different models and types of processing. The distinction between common and rare data is one of them. Information can also be either factual or generic. Generic knowledge can be, for example, a model of the observed phenomenon, general rules, integrity constraints. Factual information is more directly related to the observations. Often, these two types of information have different specificities. Generic information is usually less specific (and serves as a "default") than factual information, which is directly relevant to the particular phenomenon being observed. The default is considered if the specific information is not available or reliable, otherwise, and if the elements of information are contradictory, more specific information is preferred. Finally, information can be static or dynamic, and again, this leads to different ways of modeling and describing it.

The information handled in a fusion process is comprised, on the one hand, of the elements of information we wish to fuse together and, on the other hand, of additional information used to guide or assist the combination. It can consist of information regarding the information we wish to combine, such as information on the sources, on their dependences, their reliability, preferences, etc. It can also be contextual information regarding the field. This additional information is not necessarily expressed using the same formalism as the information we wish to combine (it usually is not), but it can be involved in choosing the model used for describing the elements of information we wish to fuse.

One of the important characteristics of information in fusion is its imperfection, which is always present (fusion would otherwise not be necessary). It can take different forms, which are briefly described below. Let us note that there is not always a consensus on the definition of these concepts in other works. The definitions we give here are rather intuitive and well suited to the problem of fusion, but are certainly not universal. The different possible nuances are omitted on purpose here because they will be discussed further and illustrated in the following chapters for each field of fusion described in this book.

Uncertainty. Uncertainty is related to the truth of an element of information and characterizes the degree to which it conforms with reality [DUB 88]. It refers to the nature of the object or fact involved, its quality, its essence, or its occurrence.

Imprecision. Imprecision involves the content of the information and therefore is a measurement of a quantitative lack of knowledge on a measurement [DUB 88]. It involves the lack of accuracy in quantity, size, time, the lack of definition on a proposal which is open to different interpretations or with vague and ill-defined contours. This concept is often confused with uncertainty because both these imperfections can be present at the same time and one can cause the other. It is important to be able to tell the difference between these two terms because they are often antagonistic, even if they can be included in a broader meaning for uncertainty. On the contrary, other classifications with a larger number of categories have been suggested [KLI 88].

Incompleteness. Incompleteness characterizes the absence of information given by the source on certain aspects of the problem. Incompleteness of the information originating from each source is the main reason for fusion. The information provided by each source is usually partial, i.e. it only provides one vision of the world or the phenomenon we are observing, by only pointing out certain characteristics.

Ambiguity. Ambiguity expresses the possibility for an element of information to lead to two interpretations. It can be caused by previous imperfections, for example, an imprecise measure that does not make it possible to distinguish two situations, or the incompleteness that causes possible confusion between objects and situations that cannot be separated based on the characteristics exposed by the source. One of the objectives of fusion is to erase the ambiguities of a source using the information provided by the other sources or additional knowledge.

Conflict. Conflict characterizes two or more elements of information leading to contradictory and therefore incompatible interpretations. Conflict situations are common in fusion problems and are often difficult to solve. First of all, detecting conflicts is not always simple. They can easily be confused with other types of imperfections, or even with the complementarity of sources. Furthermore, identifying and classifying them are questions that often arise, but in different ways depending on the field. Finally, solutions come in different forms. They can rely on the elimination of unreliable sources, on taking into account additional information, etc. In some cases, it can be preferable to delay the combination and wait for other elements of information that might solve the conflicts, or even not go through with the fusion at all.

There are other, more positive characteristics of information that can be used to limit the imperfections.

Redundancy. Redundancy is the quality of a source that provides the same information several times. Redundancy among sources is often observed, since the sources provide information about the same phenomenon. Ideally, redundancy is used to reduce uncertainties and imprecisions.

Complementarity. Complementarity is the property of sources that provide information on different variables. It comes from the fact that they usually do not provide information about the same characteristics of the observed phenomenon. It is directly used in the fusion process in order to obtain more complete overall information and to remove ambiguities.

The tools that can be used to model the different kinds of information and to measure the imperfections of the information, as well as redundancy and complementarity, will be described in Chapter 6.

1.4. Numerical/symbolic

There has been a great deal of discussion in the fusion community regarding the duality between numerical and symbolic fusion. The objective in this section is not to go over the details of these discussions, but rather to present the different levels on which this question can be expressed. By cleverly describing these levels, it is often possible to silence these debates. The three levels we will distinguish here involve the type of data, the type of process applied to the data and the role of representations. They are discussed in detail in the following sections.

1.4.1. *Data and information*

By numerical information, we mean information that is directly given in the form of numbers. These numbers can represent physical measurements, gray levels in an image, the intensity of a signal, the distance given by a range-finder, or the response to a numerical processing operator. They can be either directly read inside the data we wish to fuse or attached to the field or the contextual knowledge.

By symbolic information, we mean any information given in the form of symbols, propositions, rules, etc. Such information can either be attached to the elements of information we wish to fuse or to knowledge of the field (for example, proposals on the properties of the field involved, structural information, general rules regarding the observed phenomenon, etc.).

The classification of information and data as numerical or symbolic cannot always be achieved in a binary way, since information can also be hybrid, and numbers can represent the coding of information of non-numerical nature. This is typically the case when evaluating data or a process, or when quantifying imprecision or uncertainty. In such cases, the absolute values of the numbers are often of little importance and what mostly counts is where they lie on a scale, or the order they are in if several quantities are evaluated. The term "hybrid" then refers to numbers used as symbols to represent an element of information, but with a quantization, which makes it possible to handle them numerically. These numbers can be used for symbolic as well as numerical information.

1.4.2. *Processes*

In the context of information processing, a numerical process refers to any calculation conducted with numbers. In information fusion, this covers all of the methods that combine numbers using formal calculations. It is important to note that this type of process does not necessarily formulate any hypotheses regarding the type of information represented by numbers. At the beginning, information can be either numerical or symbolic in nature.

Symbolic processes include formal calculation on propositions (for example, logic-type methods or grammars, more details of which can be found in [BLO 01]), possibly taking into account numerical knowledge. Structural methods, such as graph-based methods, which are widely used in structural shape recognition (particularly for fusion), can be included in the same category.

We use the phrase hybrid process for methods where prior knowledge is used in a symbolic way to control the numerical processes, for example, by declaring propositional rules that suggest, enable or on the contrary prohibit certain numerical operations. Typically, a proposition that defines in which cases two sources are independent can be used to choose how probabilities are combined.

1.4.3. *Representations*

As shown in the two previous sections, representations and their types can play very different roles. Numerical representations can be used for intrinsically numerical data but also for evaluating and quantizing symbolic data. Numerical representations in information fusion are often used for quantifying the imprecision, uncertainty or unreliability of the information (this information can be either numerical or symbolic in nature) and therefore to represent information on the data we wish to combine rather than the data itself. These representations are discussed in greater detail in the chapters on numerical fusion methods. Numerical representations are also often used for degrees of belief related to numerical or symbolic knowledge and for degrees of consistency or inconsistency (or conflict) between the elements of information (the most common case is probably the fusion of databases or regulations). Let us note that the same numerical formalism can be used to represent different types of data or knowledge [BLO 96]: the most obvious example is the use of probabilities to represent data as different as frequencies or subjective degrees of belief [COX 46].

Symbolic representations can be used in logical systems, or rule-based systems, but also as *a priori* knowledge or contextual or generic knowledge used to guide a numerical process, as a structural medium, for example, in image fusion, and of course as semantics attached to the objects handled.

In many examples, a strong duality can be observed between the roles of numerical and symbolic representations, which can be used when fusing heterogenous sources. Examples will be given in different fields in the following chapters.

1.5. Fusion systems

Fusion generally is not an easy task. If we simplify, it can be divided into several tasks. We will briefly describe them here because they will serve as a guide to describing theoretical tools in the following chapters. Let us consider a general fusion problem with m sources $S_1, S_2, \ldots, S_m$, and where the objective is to make a decision among n possible decisions $d_1, d_2, \ldots, d_n$. The main steps we have to achieve in order to build the fusion process are as follows:

1) modeling: this step includes choosing a formalism and expressions for the elements of information we wish to fuse within this formalism. This modeling can be guided by additional information (regarding the information and the context or the field). Let us assume, to give the reader a better idea, that each source S_j provides an element of information represented by M_i^j regarding the decision d_i. The form of M_i^j depends of course on which formalism was chosen. It can, for example, be a distribution in a numerical formalism, or a formula in a logical formalism;

2) estimation: most models require an estimation phase (for example, all of the methods that use distributions). Again, the additional information can come into play;

3) combination: this step involves the choice of an operator, compatible with the modeling formalism that was chosen, and guided by the additional information;

4) decision: this is the final step of fusion, which allows us to go from information provided by the sources to the choice of a decision d_i.

We will not go into further detail about these steps here because it would require discussing formalisms and technical aspects. This will be the subject of the following chapters.

The way these steps are organized defines the fusion system and its architecture. In the ideal case, the decision is made based on all of the M_i^j, for all of the sources and all of the decisions. This is referred to as global fusion. In the global model, no information is overlooked. The complexity of this model and of its implementation leads to the development of simplified systems, but with more limited performances [BLO 94].

A second model thus consists of first making local decisions for each source separately. In this case, a decision $d(j)$ is made based on all of the information originating from the source S_j only. This is known as a decentralized decision. Then, in a second step, these local decisions are fused into a global decision. This model is the obvious choice when the sources are not available simultaneously. It provides answers rapidly because procedures are specific to each source, and can easily be adapted to the addition of new sources. This type of model benefits from the use of techniques from adaptive control and often uses distributed architectures. It is also referred to as decision fusion [DAS 96, THO 90]. Its main drawback comes from the fact that it poorly describes relations between sensors, as well as the possible correlations or dependences between sources. Furthermore, this model very easily leads to contradictory local decisions ($d(j) \neq d(k)$ for $j \neq k$) and solving these conflicts implies arbitration on a higher level, which is difficult to optimize, since the original information is no longer available. Models of this type are often implemented for real-time applications, for example in the military.

A third model, "orthogonal" to the previous one, consists of combining all of the M_i^j related to the same decision d_i using an operation F, in order to obtain a fused form $M_i = F(M_i^1, M_i^2, \ldots, M_i^m)$. A decision is then made based on the result of this combination. In this case, no intermediate decision is made and the information is handled within the chosen formalism up until the last step, thus reducing contradictions and conflicts. This model, just like the global model, is a centralized model that requires all of the sources to be available simultaneously. Simpler than the global model, it is not as flexible as the distributed model, making the possible addition of sources of information more difficult.

Finally, an intermediate, hybrid model consists of choosing adaptively which information is necessary for a given problem based on the specificities of the sources. This type of model often copies the human expert and involves symbolic knowledge of the sources and objects. It is therefore often used in rule-based systems. Multi-agent architectures are well suited for this model.

The system aspect of fusion will be discussed further in an example in Chapter 10.

1.6. Fusion in signal and image processing and fusion in other fields

Fusion in signal and image processing has specific features that need to be taken into account at every step when constructing a fusion process. These specificities also require modifying and complexifying certain theoretical tools, often taken from other fields. This is typically the case of spatial information in image fusion or in robotics. These specificities will be discussed in detail in the case of fusion in signal, image and robotics in the following chapters.

The quality of the data to be processed and its heterogenity are often more significant than in other fields (problems in combining expert opinions, for example). This causes an additional level of complexity, which has to be taken into account in the modeling, but also in the algorithms.

The data is mostly objective (provided by sensors), which separates them from subjective data such as what can be provided by individuals. However, they maintain a certain part of subjectivity (for example, in the choice of the sensors or the sources of information, or also of the acquisition parameters). There is also some subjectivity in how the objectives are expressed. Objective data is usually degraded, either because of imperfection in the acquisition systems, or because of the processes to which it is subjected.

In fact, one of the main difficulties comes from the fact that the types of knowledge that are dealt with are very heterogenous. They are comprised not just of measurements and observations (which can be heterogenous themselves), but also of general cases, typical examples, generic models, etc.

The major differences with other application fields of information fusion first stem from the fact that the essential question (and therefore the objective of fusion) is not the same. In signal and image processing, it consists essentially, according to Definition 1.1, of improving our knowledge of the world (as it is). This implies the existence of a truth, even if we only have access to a partial or deformed version of it, or if it is difficult to obtain, as opposed to the fusion of preferences (the way we want the world to be), the fusion of regulations (the way the world should be), or voting problems, where typically there is no truth, etc. [BLO 01].

1.7. Bibliography

[BAR 88] BAR-SHALOM Y., FORTMANN T.E., *Tracking and Data Association*, Academic Press, San Diego, California, 1988.

[BLO 94] BLOCH I., MAÎTRE H., "Fusion de données en traitement d'images: modèles d'information et décisions", *Traitement du Signal*, vol. 11, no. 6, p. 435-446, 1994.

[BLO 96] BLOCH I., "Incertitude, imprécision et additivité en fusion de données: point de vue historique", *Traitement du Signal*, vol. 13, no. 4, p. 267-288, 1996.

[BLO 01] BLOCH I., HUNTER A. (ED.), "Fusion: General Concepts and Characteristics", *International Journal of Intelligent Systems*, vol. 16, no. 10, p. 1107-1134, October 2001.

[COX 46] COX R.T., "Probability, Frequency and Reasonable Expectation", *Journal of Physics*, vol. 14, no. 1, p. 115-137, 1946.

[DAS 96] DASARATHY B.V., "Fusion Strategies for Enhancing Decision Reliability in Multi-Sensor Environments", *Optical Engineering*, vol. 35, no. 3, p. 603-616, March 1996.

[DUB 88] DUBOIS D., PRADE H., *Possibility Theory*, Plenum Press, New York, 1988.

[HAL 97] HALL D.L., LLINAS J., "An Introduction to Multisensor Data Fusion", *Proceedings of the IEEE*, vol. 85, no. 1, p. 6-23, 1997.

[JDL 91] *Data Fusion Lexicon, Data Fusion Subpanel of the Joint Directors of Laboratories Technical Panel for C^3*, F. E. White, Code 4202, NOSC, San Diego, California, 1991.

[KLI 88] KLIR G.J., FOLGER T.A., *Fuzzy Sets, Uncertainty, and Information*, Prentice Hall, Englewood Cliffs, 1988.

[LEU 96] LEUNG H., "Neural Networks Data Association with Application to Multiple-Target Tracking", *Optical Engineering*, vol. 35, no. 3, p. 693-700, March 1996.

[ROM 96] ROMINE J.B., KAMEN E.W., "Modeling and Fusion of Radar and Imaging Sensor Data for Target Tracking", *Optical Engineering*, vol. 35, no. 3, p. 659-673, March 1996.

[THO 90] THOMOPOULOS S.C.A., "Sensor Integration and Data Fusion", *Journal of Robotics Systems*, vol. 7, no. 3, p. 337-372, 1990.

[WAL 99] WALD L., "Some Terms of Reference in Data Fusion", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, p. 1190-1193, 1999.

[YED 97] YEDDANAPUDI M., BAR-SHALOM Y., PATTIPATI K., "IMM Estimation for Multi-Target-Multisensor Air Traffic Surveillance", *Proceedings of the IEEE*, vol. 85, no. 1, p. 80-94, 1997.

Chapter 2

Fusion in Signal Processing

2.1. Introduction

There has been a significant evolution in sensors available today in terms of performance and quality as well as the associated signal processing. This constant progress, from the perspective of both hardware and software, provides us with increasingly dense and complex elements of information, that differ in nature and reliability, for example, the multi-mode radar, capable of performing several tasks such as detecting, tracking or identifying targets.

Whether in the field of military applications, with the improved performances of portable devices, where speed, range, maneuverability, stealth, signal jamming and group movements have a direct impact on the surveillance system's efficiency, or in other fields of signal processing, there are major demands: a surveillance or diagnosis system must have a reactivity close to real-time, without loss of performance, and must offer as quickly as possible a situation assessment, with a reliability and an accuracy known to the operator. The use of a single type of sensor quickly became obsolete and the multi-sensor approach, associated with information fusion, progressively became prevalent for the creation of a comprehensive system to assist decision making.

This multi-sensor approach introduced new concepts, many of them inherent to how the systems functioned, such as control, decision making and communications management, in order to co-ordinate the various components and to ensure a certain consistency. Because of disparities in response time, accuracy or operating conditions between the sensors, managing such a system is complex in many regards.

Chapter written by Jean-Pierre LE CADRE, Vincent NIMIER and Roger REYNAUD.

The major concepts are directly related to information processing. Data fusion systems rely mostly on a series of modeling, estimation, retiming and data association, combination (or fusion itself) of elements of information, and then decision making or supervision steps. Going from the knowledge of a bit of information to a mathematical representation that renders it usable constitutes the information modeling stage. The retiming and data association phase is preliminary to the combination or fusion phase of multi-source information. The first three phases are usually clearly uncorrelated from the decision making phase, which consists of expressing compromise problems (costs, risks, etc.). These concepts allow us to achieve improvements due to the complementarity and redundancy of the pre-existing information and of the measurements. A system's efficiency then results from the complexity of the resulting system, from the reliability of the model, from the retiming and association techniques, from the clever combination of the information, and finally from the decisions that are made.

At the same time, information and communications systems are expected to assist and co-operate with the operators of the application field (the users) with the goal of reaching a decision. There are functions that are entirely automated on a local scale over which the operator has no element of control because these functions are reliable and/or accurate enough. On the other hand, the system as a whole has to be interactive with the user, who has to be able to control certain parts of the system by modifying, for example, confidence levels on whether a set of considered hypotheses is complete, or by defining in real-time a balance between different decision criteria. The system should also be capable of providing complementary information, upon request from the user, for example, on the level of conflict between elements of information.

One of the fundamental ideas has to do with the meaning of information and the combination mechanisms in a broad sense. The modeling that is chosen has to be suited accurately to the meaning of the information that is actually available. This accuracy in modeling causes problems of heterogenity or hybridism in the representation of data. This leads to the suggestion of modeling and heterogenous fusion mechanisms where the concept of reliability between the meaning of the information actually available, and the meaning of the mathematical representation is essential.

The question of focusing more on the combination mechanisms rather than the semantics, or vice versa, divides researchers in this field. In the field of signal processing, the trend among authors has been to emphasize mechanisms based on the idea that the process's quality essentially relies on the quality of the mechanisms involved. Probability theory, based on a strong sense of modeling, gives us well-known and most importantly well-controlled mechanisms (simulated annealing, hypothesis test, multi-model Kalman filtering, etc.). From this perspective, probability theory is therefore the "right" theoretical framework which has been particularly well studied by a number of researchers, despite certain drawbacks regarding the reliability of the semantic representations when there is little information, but the semantic aspect remains fundamental. It is therefore useful to rely on other forms of representing information in order to increase the model's reliability by considering information of smaller meaning, or by adding mechanisms for sorting, windowing, etc., to authorize this semantic information to be taken into account.

A certain number of difficulties in data fusion are caused by generic problems that are independent of theoretical frameworks.

The first problem is how knowledge, or the lack of knowledge, is modeled (meaning of the information and semantic representation). As we will see, there are several methods.

The second generic problem involves the method of information fusion and the choice of mechanisms for information management. The problems we are discussing here involve reliability and/or data association. These are questions related to the concept of uncertainty. The "right" method for combining information necessarily takes into account the imprecision relative to each source. Let us note that the choice of mechanisms strongly depends on how knowledge is modeled because either information is reliably modeled and the combinations are rather simple, or the model lacks in reliability and, in that case, additional focus is needed on the mechanisms in order to take into account the reliability problems during the combination phase.

Finally, a third difficulty lies in the choice of evaluation criteria for the quality of classifiers. This is because performance in terms of proper classification rates is not, by itself, a sufficient criterion, hence the necessity of evaluating a classifier's robustness, in other words how well performances rate when the model strays from reality.

2.2. Objectives of fusion in signal processing

Let us recall Definition 1.1 from the previous chapter: fusion consists of combining information originating from several sources in order to improve decision making. In the field of signal processing, the goal of information fusion is to obtain a system to assist decision making, whose main quality (among others) is to be robust when faced with various imprecisions, uncertainties and forms of incompleteness regarding the information sources.

The basic fusion mechanism is described in Chapter 1. It is comprised of four sequential phases, i.e. a modeling phase, an estimation phase, the actual combination phase and a decision phase. A fusion system is then comprised of a collection of different basic mechanisms depending on the problem we are dealing with. We will now discuss the three major categories of problems that information fusion techniques attempt to solve in the field of signal processing.

2.2.1. *Estimation and calculation of a law* a posteriori

In the context of mobile robotics, for which the general concepts of fusion will be described in Chapter 4, the navigation of a mobile robot is a basic problem for a fusion system. It is well-known today that the solution to this problem is obtained from the competition between two sub-systems [ABI 92, STE 95]:

– an almost continuous navigation, using dead reckoning, based on a behavioral model and on data provided by different proprioceptive sensors;

– a retiming operation at regular intervals based on the observation of checkpoints or control points located near the mobile robot.

Dead reckoning uses sensors such as a gyrometer, an accelerometer, a steering wheel angle measurements, a pedometer and an odometer (based on an angular encoder connected to a wheel). The exclusive use of dead reckoning works through the integration of data using a dynamic model and cannot prevent the estimated trajectory from straying from the actual trajectory. It is therefore necessary to observe the real world at regular intervals, using sensors such as cameras, distance measurements, acoustic or optical barriers, GPS (Global Positioning System) in order to register the estimated trajectory with the real world. The most commonly used fusion mechanism consists of combining various elements of information through an extended Kalman filter that works in three phases: the first phase is a short-term prediction based on dead reckoning navigation by proprioceptive data integration[1]; when exteroceptive[2] data is accessible, the second phase consists of providing an estimate of its own location based on this data; the final phase of this iterative process is a fusion categorized as a revision or an update, which is conducted using a weighted interpolation of the distributions between the position predicted from the proprioceptive data and the position estimated from the exteroceptive data (see Figure 2.1, as well as section 4.2.2).

The use of an adequate Kalman filter [CHU 91] provides an optimal estimate of the internal state involving the moving object's navigation, in the context of stochastic dynamic systems theory [GEL 84, GOP 93]. The predicted and estimated positions are provided by one, two or three-dimensional probability density distributions. The greatest difficulty lies not in predicting the moving object's future position, which can be modeled very reliably, but in the mechanism for estimating the position that depends on the environment and the final accuracy desired. The environment can be completely structured (in other words filled with markers leading to a precise reconstruction of the position), partly structured (there are a certain number of markers that can be used for regular retiming, the difficulty being to find them and use them to infer

1. Proprioceptive: able to measure an attribute involving its own state.

2. Exteroceptive: able to measure an attribute involving an external object that is present.

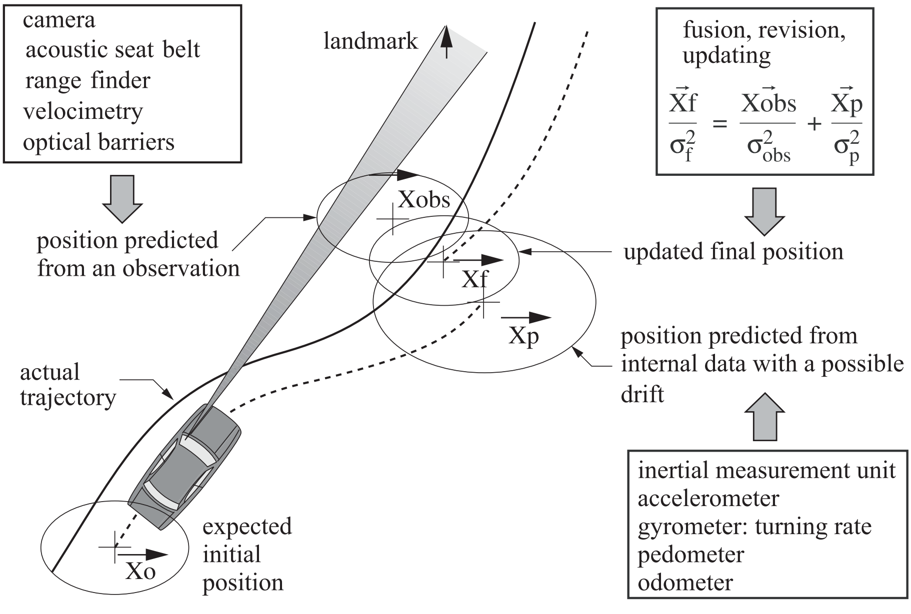


Figure 2.1. *Navigation and localization*

the vehicle's position) or non-structured. The final accuracy depends on the number of markers and on their respective configurations [BON 96, ROM 98].

The dual problem is the tracking of maneuvering moving objects that are co-operative or non-co-operative (Figure 2.2). In a tracking system [APP 98, BAR 88, BAR 93], the proprioceptive data from different moving objects maneuvering inside the scene is not available. The dead reckoning navigational sub-system is then replaced by a predictor, based on an evolutionary model that makes it possible to estimate the position of the moving object between two consecutive observations. The moving objects in question possess maneuvering capabilities and the difficulty lies in the choice of the adequate evolutionary model at a given time. The different sensors can also be located in remote sites, causing the data to be out of synchronization. We then have to define a mechanism that allows for the data to be synchronized within a grain of time that also has to be defined.

The process then works as follows: the data is acquired at each site, a local pre-processing phase sorts and validates the data before it is sent to the decision center, where each valid element of data is translated in a centralized co-ordinate system and associated with a track. At this stage of the process, we are dealing with an iterative mechanism identical to the one we saw in navigation, a prediction phase based on a

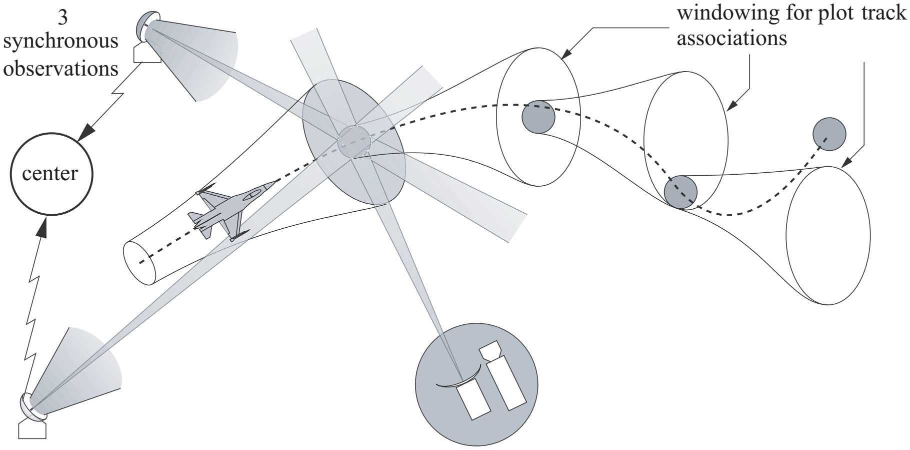


Figure 2.2. *Multi-sensor tracking of a maneuvering target*

model, followed by an update phase, based on observed data, that can be implemented using an extended Kalman filter.

If the different models used are close to reality and if the partial decisions (data validation choosing an evolutionary model, associating validated data with a track) were right, then the problem consists of combining several distributions involving a quantity in order to infer a plausible value for the resulting distribution which takes into account all of the imprecisions. If one of the partial decisions is incorrect, a conjunctive or weighted mean combination no longer has any physical meaning and more elaborate mechanisms are required to account for the problem's uncertainties. This last comment shows the importance of windowing mechanisms, which are designed to prohibit combination when the distributions are incompatible. This is relevant to fusion techniques only insofar as we are discussing the robustness of a mechanism with respect to modeling defects, the use of not perfectly reliable data, to taking into account uncertainties which are native or induced by a series of partial decisions.

The distribution combination mechanism may or may not authorize the combination of distributions, depending on the scenario.

A first windowing mechanism authorizes the combination:

– when one of the distributions is much more precise than the other, the combination must then behave as a conjunctive mechanism, such that the resulting distribution generally behaves like the most precise of the distributions;

– when the accuracies of the two distributions are close, a dissymmetric fusion of the type revision or update has to be implemented. This mechanism then makes it possible to manage some sort of a compromise between the confidence we have in the prediction mechanism and the confidence we have in the mechanism which, based on an observation, allows us to go from an estimate to the position. This continuous adjustment is fundamental if we hope to obtain an almost optimal and hence robust solution with respect to imprecisions on measurements and models.

When the first windowing mechanism does not authorize the combination, there is no actual combination. Data that is not assigned to any track is not discarded. It takes part in a mechanism for creating a new track. We can therefore consider that we are performing a non-symmetric disjunctive combination at the level of track management.

These comments on the estimation and calculation of a law *a posteriori* show that fusion in signal processing, even if it is still oriented towards statistical and probabilistic techniques, relies on the same basic physical or logical principles as in other fields.

2.2.2. *Discriminating between several hypotheses and identifying*

In a large number of identification problems, we have, on the one hand, information characterizing each hypothesis, class or type to recognize and, on the other hand, information extracted from observations. These two elements of information are provided for a set of attributes that can be seen as different explanations of the real situation, and which have to be exploited together. The information characterizing the classes will be referred to as *a priori* information, since it specifies what we can expect for the values of the attributes, conditionally to each hypothesis, before obtaining an observation. As for the observations (perceptive information), they are measurements of these attributes. This approach is maintained at every information level. Thus, an observation can be obtained from a possibly complex, previous process. At any rate, the imperfection of each observation has to be defined, whether it is a crude, low-level measurement or a high level perspective of the situation.

We will use indifferently the words sensor, observer or source of information when referring to any instrument capable of providing information on an object or an attribute assumed to be part of a continuous or discrete set. The discrete set of hypotheses within which we will have to discriminate is referred to as the frame of discernment.

In the example of Figure 2.3, we have, as our input, three *a priori* distributions involving the speed of a moving object conditionally to the type to which the moving object belongs (above). The graph in the top-right corner shows the measurement of the moving object's speed provided by a Doppler radar and the associated imprecision.

We can either plot the four distributions on the same graph, or produce the convolution of the distribution involving the measurement with each *a priori* distribution and plot only the first order moment concerning the measurement on the resulting graph (below). The intersection provides three likelihood values conditionally to each type of moving object. The concept of likelihood, in this context, refers to what we believe is true.

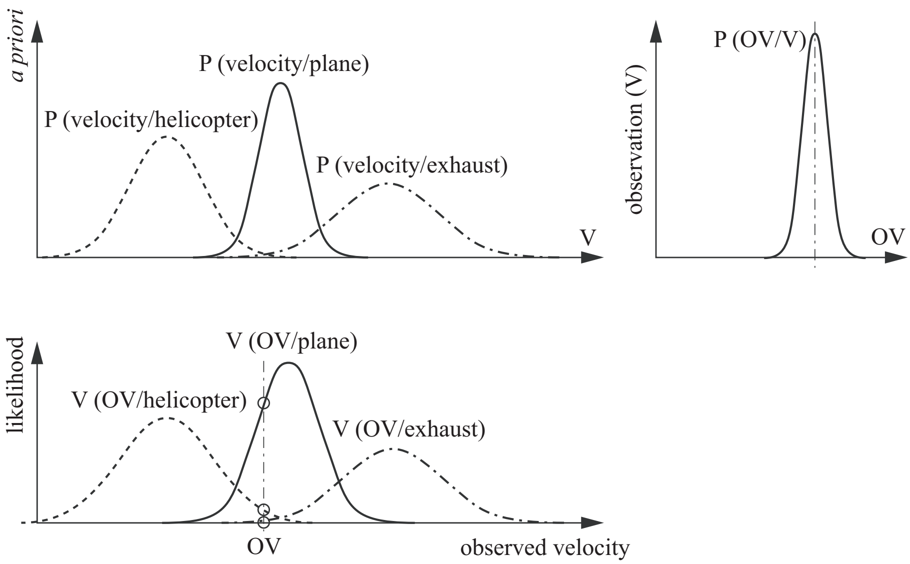


Figure 2.3. *Discriminating among several hypotheses*

Let us assume that, at this stage, we have to pick one hypothesis out of three. Either we have information at our disposal regarding the probability of a moving object in this area and we are capable of producing *a posteriori* probabilities based on these *a priori* probabilities and on the likelihoods after a phase of multiplicative combination, followed by a renormalization (see Chapter 6). It is then natural to pick the type that has the highest *a posteriori* probability. If this information is not available, we pick the type with the highest likelihood. In both mechanisms, we have just added elements of uncertainty that can be expressed using a confusion matrix in a probabilistic approach, in other words there is a probability of picking type 1 when the type was actually H1, H2, H3 or was just a false alarm. It is then possible to use the concept of Bayesian risk to make a decision that minimizes a cost function we have to define.

Today, in the context of a multi-sensor system, we have to minimize the probability of making a false decision while maintaining the highest possible level of operational efficiency for the system. Each sensor can work according to different modes competing with each other and, for each mode, a sensor can provide different attributes. For each attribute, the system contains a set of *a priori* distributions conditionally to a set of hypotheses in competition with each other.

Once the measurement involving an attribute has been validated, we calculate the likelihood of each hypothesis for which an associated conditional distribution is available. For all of the measured attributes, we can define an encompassing set of competing hypotheses, which we will refer to as the frame of discernment. For each of the hypotheses in question, either we have its likelihood conditionally to the measurement of an attribute, or we implant new mechanisms to extend the likelihoods to all of the hypotheses in the frame of discernment. At this stage, the likelihoods are fused to provide an overall likelihood of each hypothesis in the frame, conditionally to the set of attribute measurements that have been conducted. It is then possible to implant a decision mechanism.

We now consider problems involving reliability and/or data association. These are questions related to concepts of uncertainty. The "right" method of combining elements of information necessarily takes into account the imprecisions and uncertainties related to each source. It is worth noting that the choice of mechanisms strongly depends on how knowledge is modeled, since either information is accurately modeled, and the combinations are rather simple, or the modeling is not accurate, in which case the focus should be placed on the mechanisms for taking into account reliability problems.

Different techniques have been developed these past years, particularly in the field of tracking [BAR 88, BLO 88a, BLO 88b, BLO 89, REI 79, SIN 74].

Let us assume that we are attempting to track a moving object (a target) using a sensor. Let us also assume that the sensor is noisy, causing a certain number of false alarms. The risk here is to take a false alarm into account in order to retime the target's state vector. Once the track has been initialized, a prediction window (ellipsoidal with three sigmas, for example) is available at the time t. The measurements appearing in this window are validated. The other measurements are directly considered as false alarms and are discarded.

There are two types of methods we can use:

– MHT (Multiple Hypothesis Tracking) [REI 79] in which each validated measurement is associated with a track. By studying the likelihood over time of each of these tracks, it is possible to weed out some of them. The hypotheses corresponding to different tracks are managed using a tree diagram. Combinatorial aspects limit the size of the solvable problems. This method is therefore adapted to cases with a limited number of false alarms;

– PDAF (Probabilistic Data Association Filter) [BAR 74, BAR 80] in which all of the validated measurements are assigned to the track. In this case, we conduct a weighted mean combination, in agreement with the theorem of total probabilities. In this case, we hope for a uniform spatial probabilistic distribution of false alarms. These can thus have a statistically isotropic influence and therefore be filtered over the course of the iterations in time. This method is therefore adapted to cases with higher numbers of false alarms.

What should be understood at this level is that it is necessary to jointly take into account both the estimation and mechanisms for managing uncertainties, by explicitly displaying a measurement of what is believed to be true for each uncertainty that has to be managed (for example, the association between a validated measurement and a track). Several measurements have been considered, the most common of which are:

– Fisher information [FIS 12], which relies on the inverse of a covariance matrix [MAN 92];

– Shannon information, obtained from the likelihood algorithm of a probability distribution [MCI 96];

– Kullback-Leibler information [KUL 59] or cross entropy, which measures the distance between two probability distributions. A discrimination gain [KAS 96, KAS 97] can be calculated between the density predicted when no observation is made on the target and the density predicted if one particular sensor is handling it.

2.2.3. *Controlling and supervising a data fusion chain*

Another generic method for designing operational systems is to supervise the data processing chain. This chain is assumed to be adaptive, for example, the behavior of moving targets are governed by three competing dynamic models and a mechanism needs to be implemented to deal with the competition between these three models. Two types of methods are found in other works: by alternately switching from one model to another according to criteria that need to be defined [ACK 70] or by making the different models interact in a probabilistic framework [BLO 89]. More generally, the objective is to control the sequence of the various processes by assuming that other processes are conducted in parallel, then by deciding afterwards which process is optimal, or by defining a processing chain comprised of several steps, each step itself controlled by a set of competing models. We then have to supervise which model controls the current processing step.

This problem of dynamically affecting resources is not, strictly speaking, specific to the topic of data fusion. It exists wherever a sufficiently large number of parameters have to be supervised in order for a system to function in an optimal or sub-optimal way. This is the case in particular in the field of multi-agent systems [FER 88, GAS 92]. However, there are many specificities to a multi-sensor system.

They involve the basic mechanisms of information combination, or the choice of data that has to be retrieved (complementarity or redundancy). The allocation of multi-sensor resources is the optimization of the overall performances of a set of sensors or measuring instruments, according to operational criteria or depending on the mission of this set. Another, more concise and pragmatic definition is given by [MCI 96]: "a multi-sensor system generally has to answer four questions: which sensor should I use? for what purpose (mode)? where should I direct it? when should I begin?"

This set of sensors is also characterized by six major functions, in the case of military applications, by using the information on targets as input and the control of the sensors as output [BUE 90]:

– events are predicted in order to evaluate the periods of time during which events occur that require sensors to observe them;

– predicting the sensor's state makes it possible to model the performances of the sensor in order to determine its abilities to accomplish the tasks it was assigned to do;

– arranging targets by order of priority is done for all of the targets, according to information needs and urgency, depending on criteria based on threat (in defense), opportunity (in attack), or surveillance;

– the assignment of sensors to targets is determined based on the previous two functions (prediction on the sensor and target ranking), in order to quantify the usefulness [POP 89], the adequacy of each possible assignment (an optimal assignment is obtained by maximizing this usefulness);

– assignment control makes it possible to organize and program the sensor's various tasks over time;

– the interface with the sensors makes it possible to dispatch the orders to the various sensors.

These functions give an overview of how a sensor system works, particularly in terms of the sequence in time.

A systems architecture mostly involves how the sensors are organized with respect to each other, particularly in terms of communications, but also depending on how the information is processed. The choice of an architecture immediately leads to choosing the system's control, as well as its co-ordination. The architectures of multi-sensor systems used to be strongly centralized. These architectures had the advantage of providing information on different levels of abstraction, but the system is then vulnerable to possible breakdowns of the central processor, which has to process an increasingly large and heterogenous volume of data. Needs have evolved towards a more independent system. System control has therefore become more delocalized: it is either semi-distributed, allowing for partial fusions of information at different intermediate levels, with a final decision based on the processed information or distributed, making it possible to make many decisions locally and independently. If the system has to be efficient, the implementation of control rules is difficult and many problems occur because of local decision conflicts between agents. A comparative study in tracking mode of these three types of architectures can be found in [BLA 86].

Modeling the sensor resource [BUE 90] can help realize which characteristics play a role:

– the passive mode (the sensor merely receives information), the active mode (transmitting and receiving) or the protected mode (the transmitted and received information is in the form of a pulse);

– the direction; the frequency (changing the frequencies used by active sensors is a significant need in a military context);

– the type of wave, pulse or continuous; the power (greater range and quality of measurement, particularly in noisy or jammed environments);

– the size of the beam, thin or wide;

– the illumination time (identification requires a longer time than simple detection).

On some sensors, there are four types of control available [WAL 90]:

– global (this control mode is used to establish the default values of the sensor's parameters);

– sectoral (the surveillance volume is partitioned into different sectors, for which the sensor's parameters can be adapted);

– targeted (the parameters are adapted based on the various targets that are present);

– the last mode involves the search for targets (the precise volume or other attributes are specified).

Data retrieval plays a particular role. The first mode is the *push* mode, which means that the data processing system expects the data and processes them as they come along. The sensors continuously send observation sequences and it is up to the system to manage the waiting queue (the time sequence). The drawback of such a system is the lack of reactivity because the elements of information arrive in a pre-defined order (generally based on the sensors' acquisition times). The other data retrieval mode is the *pull* mode. In this case, the system sends sensors requests, in other words information queries, specifying among other things which target the sensor should be aimed at and the observation time. This way, the system controls the information it needs and the information retrieval sequence can be different for each target. Over the course of a tracking function, additional information requests are sent when the target is maneuvering. In a classification process, if the target's speed is very high, and if the system is hesitating between a missile and a plane, the following request may involve wingspan and the decision will then be immediate. In the *pull* mode, an operator can act on the sequence to modify a sensor's state. In reference to the seconds-long human reaction time, this is called a long loop, as opposed to an automated system response.

However, control distribution generates very stringent constraints on the management of telecommunications. The communications network can be either partly or fully connected. In [GRI 92], a study on propagation and fusion of estimates in the nodes of a multi-sensor network was conducted under three constraints: there had to be no unique fusion center, communications were imposed from one node to another and nodes had no overall information regarding the network, they only knew the nodes to which they were connected. The goal was to find the optimal estimate to propagate, using all the available and useful information, while minimizing redundancy.

Another approach to communications management in multi-sensor systems involves setting up intelligent resource allocation based on information theory. In a decentralized system, the objective is to quickly find a receiver for whom the information it receives will maximize the change in its entropy. An intelligent mechanism is compared with the standard round-robin mechanism in a multi-sensor tracking system [DEA 97b] by relying on the information filter, and on a multi-sensor identification system [DEA 97a, GRE 96] by using a decentralized Bayesian algorithm. Results show that the average and maximum waiting periods for communications can be reduced. In identification, the number of targets processed is greater because of this algorithm. In tracking, the system made it possible to reduce communications, while still obtaining more specifics about the target, and more significantly for targets following a uniform straight line trajectory or performing major maneuvers.

2.3. Problems and specificities of fusion in signal processing

By stating the three main types of objectives, we have stumbled upon a certain number of sub-problems specific to fusion in signal processing. We will now discuss a few of these basic subjects to show how they can be handled and solved.

2.3.1. *Dynamic control*

The complexity of an actual application is reflected not only in the volume of data involved, but also and most importantly in the variety of mechanisms that need to be taken into account in order to obtain a feasible solution. Supervising the whole system consists of providing a plan regarding the fusion strategy that needs to be implemented. There are currently two levels that coexist.

The first level is to provide contextual combination mechanisms in order to be able to take into account context changes defined on the overall supervisor's level. The objective is to bring back down to the combination level itself dissymmetric weightings involving different elements of information being combined. Because the system is, by nature, sub-optimal, we have to reach a compromise between the different quantities involved. The basic case is when a prediction and an observation are available. This is the case shown in Figure 2.1. The use of an extended Kalman filter amounts to managing the respective weights attributed to the present observation and the prediction function of the past and the prediction model. This filter is usually implemented recursively. The time intervals between observations are not regular and this compromise naturally has to be adapted to each new observation. The weight has to express, of course, the confidence placed in each element of information.

Obviously, the problem gets more complicated when the number of sources of information is greater than two. The simplest way to handle the process is to couple several sensors mechanically on a same platform. Many such systems exist, for example, two coupled cameras, a camera and a range-finder, a camera and a light projection system. The same applies to signal processing. The system shown in Figure 2.4 is a mechanical coupling between a radar and an FLIR which produces infrared labels. When a plane is locked onto by a radar, it will perform maneuvers to evade tracking. The usual way to do this is to use the artifacts of the single radar tracking algorithm. When the trajectory is a straight line, the algorithm makes a compromise between the weights of the observations and tracks, while maintaining a direction change detector active. This detector reacts once it has 10 samples on the target. When the pilot begins his maneuver, the plane is positioned so as to reflect as little energy as possible and triggers a counter-measure system. The role of the infrared labels, in normal mode, is to confirm the target's direction given the shape of the exhaust stream and, in direction changing mode, the variation in shape of the exhaust stream can be used to conduct an early detection and to provide information on changes in direction. In a way, the data originating from one of the sensors supervises the overall process.

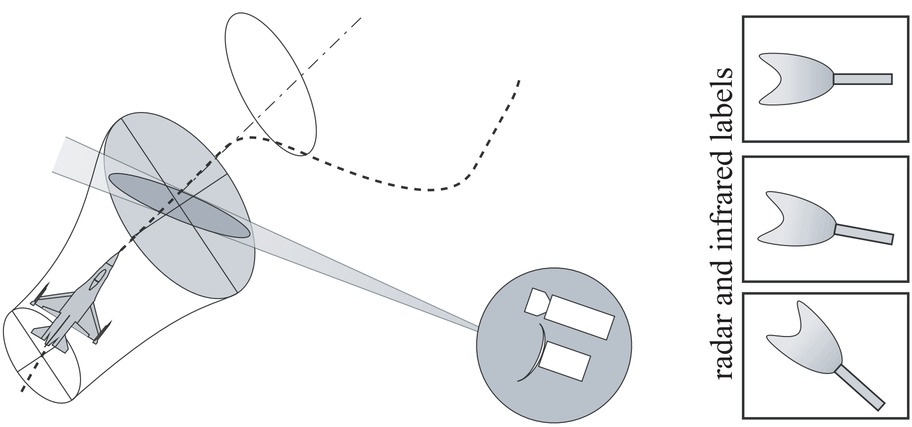


Figure 2.4. *Single platform coupling of two sensors*

More generally, in the context of the coexistence of several sensors, a confidence level is assigned to each sensor conditionally to each hypothesis of the frame of discernment [APP 91]. This level is constantly adapted based on contextual knowledge and must reflect the sensor's reliability given the context (temperature, degree of humidity, partial masking, etc.). This level can be associated with the validity of simultaneously using a set of sensors, the reliability of a sensor depending on the context and not on hypotheses [NIM 98].

The second level consists of extracting the various degrees or probabilities of confidence or reliability, based either on *a priori* information or on attribute measurements that have already been conducted. The method then relies on the operator's expertise, on knowledge acquired about the scene and on the occurrence of outside events. In the end, we need a final action that can be expressed in operational terms, such as resource management with respect to mission objectives or priority orders.

Figure 2.5 shows the three stage supervision of a change in orientation of sensors, in the case of an absence of data caused by masked terrain.

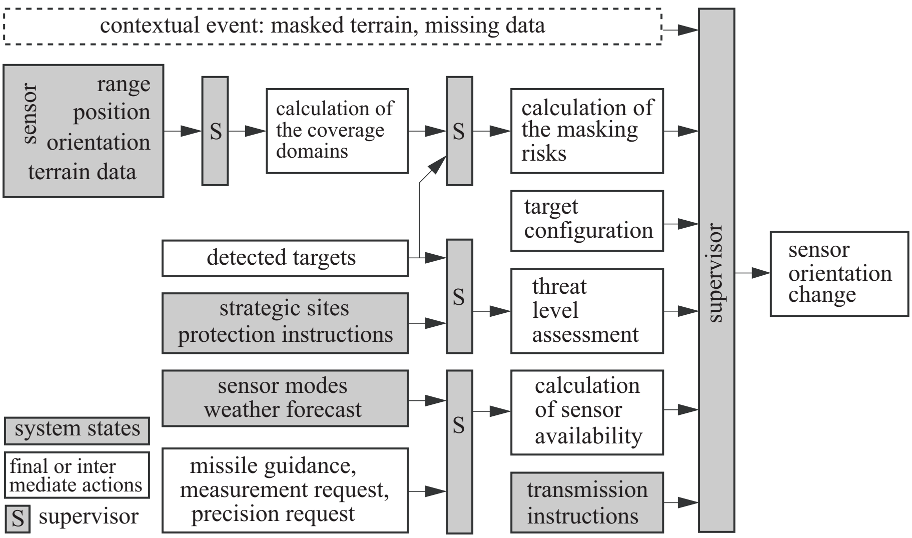


Figure 2.5. *Controlling the orientation of sensors in the case of masked terrain*

As a result, this high-level control is strongly related to the problem of synchronizing and integrating data. A geographically distributed system involves a communications network comprised of slow and fast channels. The transmission time can vary, meaning that the observation may reach the fusion center in non-chronological order.

Thus, we have the following questions:

– how long do we have to wait for the missing data?

– when they reach the processing center, should they be integrated to the fusion process?

– how does this data contribute to the quality tracking operation?

– how does the quantity of information change as it grows older?

As with missing data of contextual origin, the method can rely on a decision tree, whose final action is to choose between gaining information by using the delayed data, or gaining time by discarding this same data. There are necessarily several criteria to this choice because it involves the content of the data delayed by evaluating the expected information gain for each track, but the complexity of the resulting situation without this data plays a role and requires evaluating the risk of confusing tracks if these tracks are close, as well as evaluating the risk of losing the track.

When a multi-sensor system is operating, for scene surveillance, often the quality of the result depends mostly on time. In a multi-target environment, recognizing and identifying the objects in the scene has to be done in as short a time as possible, without waiting for all of the data that can be provided by the sensors. Because accessing the data takes time, one of the solutions is to operate in the *pull* mode, or data request mode, and to choose the smallest (the most discriminating) set of attributes for differentiating objects from one another.

For airspace surveillance, there are many types of models and data available for classification and identification. Thus, physical attributes that do not change, such as the wingspan or the length of a plane, can sometimes help in directly recognizing the target, or at least in obtaining a measurement and therefore formulate hypotheses on the target's class, which will later be confirmed or proven wrong based on other attribute measurements. Furthermore, many types of sensors can provide a wide variety of information, depending on what mode is selected and thus be used to measure different attributes. The variety of *a priori* information that can be obtained about a target when trying to classify it can be used to think in advance. This way, it is possible to recognize an object faster by taking into consideration the discriminatory capabilities of the different attributes conditionally to each class.

If, for each attribute, we have at our disposal the membership function for each class, the choice of the quickest attribute for characterizing an object is the one that leads to the result with as little ambiguity as possible. We then have to define the membership functions that represent the different classes. The degree of separation between two membership functions [CON 01, CON 02] can then be used to implement a mechanism for making it easier to choose the first attributes to search for (the selection of the sensor and of its mode), corresponding to each attribute. The *a priori* information is obtained either from learning or by constructing a database and is stored in the form of membership functions. At each iteration, the result is a list of attributes, arranged in a particular order, and the supervisor in charge of allocating the system's entire resources then provides a measurement corresponding to one of the attributes. The performance evaluation requires a simulation with the introduction of noise and a statistical validation over a large number of trials, whose quality criterion is the number of requests necessary to obtain an object's class. The order used for arranging these requests can be pre-defined, random or based on a designed mechanism. Finally, the number of requests needed to obtain the right classification of the object can bring improvements up to, for example, three out of eight total requests, without affecting non-recognition and poor recognition rates. This overall orientation of suggesting modular algorithmic subdivisions, whose output is no longer a single decision but instead an ordered list (in this case the requests), is increasingly sought after. This is because it naturally authorizes the resource allocation stage to operate in a more serene way, thus allowing it to fully play its part, with a significant decrease in conflicts over resources.

In the end, many research fields involve the problem of organizing a processing chain generally consisting of modules with sequential subdivisions and other competing subdivisions. What is specific to the field of fusion is the fact that it is necessary to evaluate in different places the amount of redundant or complementary information in a sub-module's input in order to know how this particular data can be integrated into the process. This systems architecture, where the local process control is performed from an evaluation of the information content in the inputs, is no longer sufficient today. It remains fundamental for dynamic process control, but it is accompanied by process supervision performed by each sub-system.

It is considered today that, in the same way as sensors have their own operating modes, algorithmic subdivisions also have different processing modes and that it is possible either to switch locally, or under the control of a supervisor, from one mode to another, or to locally make some of the processes compete with each other, then to aggregate, combine or fuse all of the outputs in order to provide either an optimal output or a summed-up output of the information acquired. We are no longer discussing only fusion, but the more general field of artificial intelligence. In order to achieve an operational status, all of the various subdivisions have to interact with each other according to one or several plans, such as those suggested in [GAR 97] (Figure 2.6).

The most common simple plan works by using a supervision mechanism based on competing processing modes, in a fusion architecture comprised of low-level signal processing phases and of two types of streams: a continuous stream of data and a stream for controlling processing modes on demand [CHE 97]. An accurate way to describe reality is to take into account, on the processing mode level, the concept of object behaviors by using different models expressing behavior variations in these objects [REY 96].

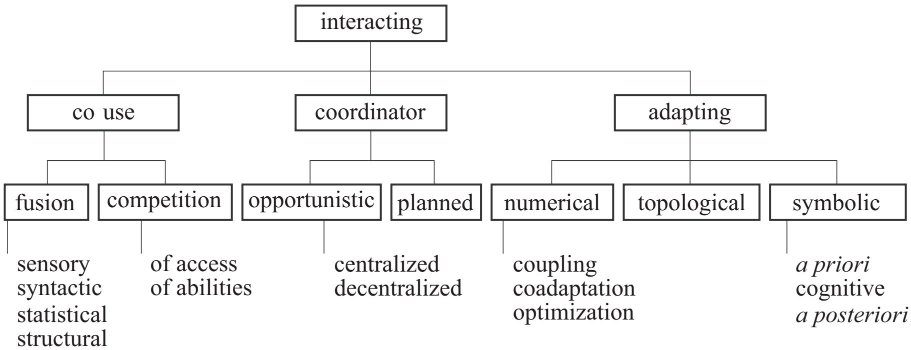


Figure 2.6. *Taxonomy of the different interaction mechanisms*

In any case, it is necessary to be able to refer to "degraded" modes, in the sense that when there is a strong uncertainty in the choice of the current behavior and supervision operates by switching modes, this uncertainty can lead to an operational malfunction. In order to remain robust towards this type of uncertainty, the use of degraded modes that include the behaviors of more accurate and "refined" modes makes it possible to manage the amount of uncertainty that exists at a given time regarding behavioral knowledge.

2.3.2. *Quality of the information*

In the field of military intelligence, the STANAG 2022 gives us a precise definition of the variable that allows us to qualify information. The first parameter is related to the quality of the source. The available information is essentially symbolic since it is obtained either directly from human sources, or from an analysis of sensor signals performed by a human operator. As a result, the quality of a source is only relative and somewhat subjective. This quality can be expressed on an alphabetical scale from A when the source is completely reliable, to D when a source is not very reliable or not at all; the letter E indicates that the reliability cannot be estimated. In a context of essentially numerical information, this same quality can be assessed but this time using a numerical coefficient that can be in the form of a probability, a possibility or a Dempster-Shafer mass. This coefficient will be calculated based on the context, which can be provided by weather conditions, the possibility of jamming or decoys, or also by the sensor's operating conditions. This coefficient will be included naturally in the fusion algorithms, in order to favor a source over another if the former has more favorable operating conditions [NIM 02].

A second parameter that can be used to assess the quality of information is given by the credibility of that information. In the field of intelligence, this credibility is provided by a numerical scale from 1, for information confirmed by other sources, to

5 for unlikely information; the number 6 denotes information for which the accuracy cannot be estimated. Likewise, in a numerical context, this quality can be assessed by taking into account the sensor's definition, the signal-to-noise ratio and it will be a representation of the likelihood of the state we are trying to estimate.

2.3.3. *Representativeness and accuracy of learning and a priori information*

For each quantity we wish to fuse, we need *a priori* information characterizing each type of object as well as one or several elements of perceptive information. This pair of elements cannot be represented using the same formalism, probabilistic, possibilistic, belief function, etc. Assessing the compatibility of the observations and the hypotheses of the frame of discernment has to be done with heterogenous elements of information, which requires the implementation of hybrid fusion mechanisms. Even the *a priori* information involving the attributes does not have to be homogenous. This point has been discussed by many authors. Expressing how a possibilistic representation can still hold some meaning when suggesting to switch over to a representation in probabilistic context is at the core of the theory explaining the meaning of each representation, and was in fact its foundation [BLO 96]. A first idea is to switch to a homogenous context, which reflects to a certain extent the accuracy with respect to the real world and leads either to a loss of information, or to including information without justification. In the case of non-homogenous combination between probability and possibility distributions, it can be useful to rely on the strengths of these two representations: the probabilistic representation holds a strong meaning for the part of the medium where the value is high; the possibilistic representation holds a strong meaning for the part where it is equal to zero. Hence, it is possible to suggest a double representation system where each of the two brings its strength for the part of the medium where it is strong [NIF 00].

Finally, we should point out the incompleteness of knowledge obtained through learning.

2.4. Bibliography

[ABI 92] ABIDI M., GONZALES C., *Data fusion in robotics and machine intelligence*, Academic Press, New York, 1992.

[ACK 70] ACKERSON G.A., FU K.S., "On state estimation in switching environment", *IEEE Transactions on Automatic Control*, vol. AC-15, no. 1, p. 10-17, 1970.

[APP 91] APPRIOU A., "Probabilités et incertitude en fusion de données multi-senseurs", *Revue scientifique et technique de la Défense*, p. 27-40, 1st trimester 1991.

[APP 98] APPRIOU A., "Uncertain data aggregation in classification and tracking processes", in B. BOUCHON-MEUNIER (ed.) *Aggregation and fusion of imperfect information*, p. 231-260, Physica-Verlag, 1997.

[BAR 74] BAR-SHALOM Y., “Extension of the Probabilistic Data Association Filter to multi-target environments”, *Symposium on non-linear estimation*, San Diego California, p. 16-21, 1974.

[BAR 80] BAR-SHALOM Y., FORTMANN T.E., SCHEFFE M., “Joint Probabilistic Data Association For multiple targets in clutters”, *Conference on Information, Sciences and Systems*, Princeton, NJ, 1980.

[BAR 88] BAR-SHALOM Y., FORTMANN T.E., *Tracking and Data Association*, Academic Press, 1988.

[BAR 93] BAR-SHALOM Y., LI X.R., *Estimation and Tracking Principles, Techniques and Software*, Artech House, Norwood, 1993.

[BLA 86] BLACKMAN S.S., *Multiple-Target Tracking with Radar Application*, Artech House Inc., Norwood, Massachusetts, 1986.

[BLO 88a] BLOM H. A.P., “An Efficient Filter for Abrupty Changing Systems”, *Proceeding of the 23rd conference on decision and control*, 1988.

[BLO 88b] BLOM H.A.P., BAR-SHALOM Y., “The Interactive Multiple Model Algorithm for Systems with Markovian Swithching Coefficients”, *IEEE Transactions on Automatic Control*, vol. 33, no. 8, 1988.

[BLO 89] BLOM H.A.P., BAR-SHALOM Y., “Tracking a Maneuvring target Using Input Estimation Versus the Interacting Multiple Model Algorithm”, *IEEE Transaction on Aerospace and Electronic Systems*, vol. AES 25, no. 2, 1989.

[BLO 96] BLOCH I., “Incertitude, imprécision et additivité en fusion de données: un point de vue historique”, *Revue Traitement du Signal*, vol. 13, no. 4, p. 267-288, 1996.

[BON 96] BONNIFAIT P., GARCIA G., “Continuous Localization of Outdoor Mobile Robots using Odometry and Relative Azimuth Angles of known Landmarks”, *Symposium on Robotics and Cybernetics*, vol. CESA’96, p. 30, 1996.

[BUE 90] BUEDE D., WALTZ E., “Issues in Sensor Management”, *Proc. 5th IEEE Int. Symp. on Intelligent Control*, vol. 2, Philadelphia, p. 839-842, 1990.

[CHE 97] CHELLA A., GESÚ V.D. *et al.*, “DAISY: a Distributed Architecture for Intelligent System”, *Computer Architecture for Machine Perception, CAMP 97*, p. 42-50, 1997.

[CHU 91] CHUY C., CHEN G., *Kalman Filtering with Real Time Application*, Springer-Verlag, 1991.

[CON 01] CONTAT M., NIMIER V., REYNAUD R., “An Ordering Method to Select a Sensor and its Mode in a Multitarget Environment”, *4th International Conference On Information Fusion, Fusion 2001*, Montreal, Quebec, ISIF, 2001.

[CON 02] CONTAT M., NIMIER V., REYNAUD R., “Request Management using Contextual Information for Classification”, *5th International Conference On Information Fusion, Fusion 2002*, Annapolis, Maryland, ISIF, 2002.

[DEA 97a] DEAVES R. H., GREENWAY P., BULL D.R., BRIDGES M., NICHOLSON D., “An Evaluation of Communications Management in a Simulated Decentralized Identity Fusion System”, *Proceedings*, vol. 3068, SPIE, p. 64-75, 1997.

[DEA 97b] DEAVES R.H., NICHOLSON D., GREENWAY P., "An Evaluation of Communications Management in a Decentralized Target Tracking System", *Proceedings*, vol. 3068, SPIE, p. 261-271, 1997.

[FER 88] FERBER J., GHALLAB M., "Problématique des Univers Multi-Agents Intelligents", *Actes des journées PRC-IA*, Toulouse, 1988.

[FIS 12] FISHER R., "On a absolute criterion for fitting frequency curves", *Messenger of Math*, vol. 41, p. 155, 1912.

[GAR 97] GARBAY C., OPPIZI O., Interaction Système/Environnement pour l'interpretation des Signaux et des Images, Research report GdR ISIS, 1997.

[GAS 92] GASSER L., *An overview of Distributed Artificial Intelligence*, Kluwer Academic Publisher, 1992.

[GEL 84] GELB A., *Applied optimal estimation*, MIT Press, Cambridge, 1984.

[GOP 93] GOPAL M., *Modern Control System Theory*, New Age Intern. Publishers, 1993.

[GRE 96] GREENWAY P., DEAVES R.H., BULL D., "Communications Management in Decentralised Data Fusion Systems", *Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems*, IEEE, 1996.

[GRI 92] GRIME S., DURRANT-WHYTE H.F., HO P., "Communication in Decentralized Data-Fusion Systems", *American Control Conference*, Chicago, 1992.

[KAS 96] KASTELLA K., MUSICK S., "The Search for Optimal Sensor Management", *proceeding*, vol. 2759, SPIE, p. 318-329, 1996.

[KAS 97] KASTELLA K., "Discrimination Gain to Optimize Detection and Classification", *IEEE Trans. on Systems, Man, and Cybernetics, Part A: Systems and Humans*, vol. 27, no. 1, p. 112-116, 1997.

[KUL 59] KULLBACK S., *Information theory and statistics*, Wiley, New York, 1959.

[MAN 92] MANYIKA J., DURRANT-WHYTE H.F., "An Information-Theoretic Approach to Management in Decentralized Data Fusion", *Proceedings*, vol. 1828, SPIE, p. 202-213, 1992.

[MCI 96] MCINTYRE G., HINTZ K.J., "An Information Theoretic Approach to Sensor Scheduling", *Proceedings*, vol. 2755, SPIE, p. 304-312, 1996.

[NIF 00] NIFLE A., REYNAUD R., "Double Representation of Information and Hybrid Combination for Identification Systems", *International Conference on Information Fusion*, ISIF, 2000.

[NIM 98] NIMIER V., "Introduction d'information contextuelle dans des algorithmes de fusion multicapteurs", *Traitement du Signal*, vol. 14, no. 5, p. 543-553, 1998.

[NIM 02] NIMIER V., "Soft Sensor Management fo Multisensor Tracking Algorithm", in A.K. HYDER *et al.* (ed.), *Multisensor Fusion*, p. 365-380, Kluwer, 2002.

[POP 89] POPOLI R., BLACKMAN S., BROIDA T., A Utility Theory Method for Allocation of the Radar Electronically Scanned Antenna, Technical report, Hughes, 1989.

[REI 79] REID D.B., "An Algorithm for Tracking Multiple Targets", *IEEE Transactions on Automatic Control*, vol. AC24, no. 6, p. 843-854, 1979.

[REY 96] REYNAUD R., MAURIN T., "Control and supervision of a fusion processing line by competition among models", *Symposium on Robotics and Cybernetics*, CESA'96, IMACS, p. 494-499, 1996.

[ROM 98] ROMBAUT M. *et al.*, La perception et la fusion de données pour véhicules intelligents, Synthetic report, GDR ISIS, 1998.

[SIN 74] SINGER R.A., SEA R., HOUSEWRIGHT K.B., "Derivation and Evaluation of Improved Tracking Filters for Use in a Dense Multitargets Environments", *IEEE Transactions on Information theory*, vol. IT 20, no. 4, p. 423-432, 1974.

[STE 95] STELLE E. *et al.*, "Position estimation for a mobile robot using Data Fusion", *Symposium on Intelligent Control*, IEEE, p. 565, 1995.

[WAL 90] WALTZ E., LLINAS J., *Multisensor Data Fusion Approach*, Artech House Inc., Norwood MA, 1990.

Chapter 3

Fusion in Image Processing

In the same way as the previous chapter described the specificities of fusion techniques applied to signal processing, this chapter will focus on the specificities of fusion in image processing. We will go back to the general definitions provided in Chapter 1 and discuss them in this particular context. We wish to emphasize the specific nature of images and their representation in fusion problems, and insist on what makes fusion in image processing different from most of the other application fields in fusion.

3.1. Objectives of fusion in image processing

Images appeared of course very early on as important sources of information for existing information fusion systems and data fusion systems have used images. Let us consider, for example, a comprehensive tracking application for ecological situations. It requires remote sensing to provide weather information. The data provided by the image can then be integrated in a physical model by estimating, for each pixel, the cloud cover. We can include in a thermodynamics balance equation the level of water vapor estimated this way for each point inside the image.

However, this is not the type of application that has led to the original field of *image fusion*. We should look instead at the practice of image interpretation experts, where we will find the models that image processors have tried to copy, from widely different areas of society. Here are two examples, but we could easily find many other and more diverse cases.

Chapter written by Isabelle BLOCH and Henri MAÎTRE.

A radiologist in a hospital environment makes his decisions based on many photographs that often give different points of view of a same anatomical structure. These photographs can be spread out over a view box and the doctor makes his opinion based on closely and alternately examining several points of view. Based on a hypothesis inferred from one of the photographs, he confirms his interpretations using the others, clarifies it by cross-checking and including additional points of view, or on the contrary rules it out based on contradictory information provided by one of these points of view. This situation is typical of medical imaging, which is a field where the acquisition techniques are growing more diverse: X-rays, magnetic resonance, nuclear imaging and ultrasound imaging, each one leading to a variety of possible modes depending on the acquisition protocols. It also finds support in the efforts made by all hospital structures to group all of the image sources together in the same ward or to have all of the images converge on a single console where the diagnosis will take place. These efforts have progressively led to the introduction of integrated archiving and consultation systems in hospitals (PACS[1], for example).

For our second example, we choose a remote sensing expert whose task is to interpret a complex scene. He has a large number of images at his disposal, provided by various sensors, for example, images in the visible spectrum in different ranges of wavelengths, or infrared images, or also radar images. Each source gives him information on a particular aspect of the scene, thus allowing him to come up with a scenario. Again, the expert works by confronting different representations, combining them either to support his idea or to rule it out. His ability is the result of a considerable amount of training and is increasingly complex as the image sources diversify and grow in number. However, both satellite applications, for which many sensors complete each other's information, and airborne applications, for which very different sources are used (maps, cadasters, land occupation maps, geological or agricultural surveys, elevation models, etc.), definitely tend to progress towards a greater complexity of image sources.

In this context, image fusion appears as a task in itself, distinct from data fusion because it is not clear whether it is possible to design an operational framework in which every element of information would have its place, as was the case with the satellite image that allowed us to include the "cloud cover" measurement in an overall plan involving, for example, the evapotranspiration of vegetation cover and weather conditions. In a broader context of image processing, image fusion is used to help decision making in a complex and usually poorly formalized situation, in which the different images provide an element of "truth" that contributes, in collaboration and in opposition with other sources, to an overall interpretation. Therefore, by developing automatic image fusion methods, preparing and shortening the human elaboration and expertise phase, and possibly in situations with a large number of image sources, our

1. *Picture Archiving and Communication Systems.*

wish is to manage the multiplicity without sacrificing the potential contributions of a complex combination, even beyond what human experts can achieve today.

Therefore, the task of fusion in image processing is closely related to decision making. On the other hand, it has little to do with the phase it is often associated with geometric registration of images, but this phase is today generally considered to be unavoidable, and many image fusion studies simply perform it and leave to a human operator the task of implementing the decision making.

The objective of registration is to exactly overlap the pixels corresponding to a same object observed in different images. This phase can be made easier if there is a recognized absolute frame of reference to describe the scene. This is the case, for example, in mapping or geography applications, which rely on geocoded frames of reference, as well as for medical applications, for which conventional anatomical frames of reference have been established. There are many kinds of registration techniques (see, for example, [ZIT 03]), based on different principles: correlation, dynamic programming, optical flow, elastic deformation, etc. (see [MAI 91] and [MAN 94] for summaries of the methods used in aerial and satellite imagery and medical imaging, respectively).

For which objectives are we likely to use image fusion? First of all, for improving the three main tasks of shape recognition, detecting, recognizing and identifying.

Detection. This consists in this case of validating the presence or absence of the object we are searching for: presence of a vehicle on a road, or of a stenosis in a blood vessel. This is sometimes combined with the other objective of tracking objects detected in a sequence of pictures.

Recognition or classification. A detected object is associated with one of the categories of known or expected objects based on photometric, geometric or morphological criteria. This operation can be conducted on objects on very different levels, from the pixel to complex sets of image components.

Identification. A detected and recognized object is identified when it is associated with a single prototype in its category. Thus, once a vehicle has been detected with infrared imaging, recognition can determine its type: truck, motorcycle or car, and the conclusion of the identification will be that the vehicle is the milkman's truck, which is the typical object monitored in this type of image ...

However, applications other than shape recognition also require the implementation of fusion methods. These operations can take place during the recognition process, but at a more preliminary stage, and do not necessary lead to a decision.

Segmentation. This constitutes a more focused objective than classification, since it intends to extract determined objects as precisely as possible. It can consist simply of using the complementarity of the sources of information, in order to better identify the limits of the image's homogenous components. For example, by fusing the precise graph provided by aerial imaging with the network's topological structure and its approximate geometry provided by a map, we can come up with an excellent description of the road network.

Reconstruction. The multiplicity of points of view is an advantage for three-dimensional reconstruction of observed scenes and, although this reconstruction may consist of the typical methods (such as in stereovision, or in tomography), in other situations in which acquisition is not controlled as well, it is only possible to reconstruct an approximate three-dimensional information that empirically combines the various available aspects.

Detection of change. This type of decision typically involves images taken at different dates, whether they are a map and an image, or multi-data images for tracking crops or a pathology. It may also consist of sequences of multi-source images (at a faster rate than multi-data images).

Updating knowledge of a phenomenon or a scene. Unlike in the previous case, the decision consists here of using the information provided by different sources (possibly multi-data) in order to modify or complete prior knowledge, for example, completing a road network with new traffic circles to update a map.

Some of these different decision problems are similar to the combination of experts, since each image can be considered as an expert giving his opinion according to his abilities. However, in general, the information in fusion problems of experts is more scattered than with images. Learning is therefore more difficult because there is less data, although the user often has less constraints over the algorithmic costs of the methods. With images, the amount of data to be fused is both an advantage when it comes to learning and a drawback for the computational load.

If we compare the problem of image fusion with that of data fusion based on aggregation and multi-criteria optimization, we notice that one of the main differences lies in the fact that for the latter, the goal is to find a solution that best satisfies a set of generally stringent constraints, whereas in image processing, each source provides (fairly explicitly) a level of satisfaction (for belonging to a category, for example, which can then be considered a criterion) and the decision rather consists of choosing the best one (the best category, for example).

3.2. Fusion situations

Depending on the applications, fusion problems can occur in different situations, in which the types of information elements are not the same. The main fusion situations in image processing are the following.

Several images from the same sensor. This consists, for example, of several channels on the same satellite, or multi-echo images in MRI, or also of image sequences for scenes in motion. The data in those cases is relatively homogenous because it corresponds to similar physical measurements.

Several images from different sensors. This is the most common case, in which the different physical principles of each sensor allow the user to have complementary perspectives of the scene. They can consist of ERS and SPOT images, MRI or ultrasound images, etc. The heterogenity is then much greater, since the various sensors do not deal with the same aspects of the phenomenon. Each image gives a partial image with no information on the characteristics they are not meant to observe (for example, an anatomical MRI yields no functional information and the resolution of a PET scan is too low for a precise view of the anatomy).

Several elements of information extracted from a same image. In this situation, different types of information are extracted from an image using several sensors, operators, classifiers, etc., that rely on different characteristics of the data and attempt to extract different objects, often leading to very heterogenous elements of information to fuse. The extracted information can involve the same object (fusion of contour detectors, for example) or different objects and the goal is then to find an overall interpretation of the scene and consistency between the objects. The elements of information can be on different levels (very local, or more structural when studying spatial relations between objects).

Images and another source of information. By another source of information, we mean, for example, a model, which may be particular like a map, or generic like an anatomical atlas, a knowledge base, rules, information provided by experts, etc. The elements of information are again in very different forms, both in nature and in their initial representation (images in the case of a map or a digital atlas, but also linguistic descriptions, databases, etc.).

3.3. Data characteristics in image fusion

The specifics of fusion in image processing make it difficult to take advantage of the progress made in other fields of information fusion. One of the reasons is the complexity of the data and knowledge, which make it impossible to attempt to find a comprehensive system to combine in a single relation all of the image's components.

The complexity is partly due to the volume of information to process (for example, a single MRI image of the brain takes up 8 to 16 megabytes). These large volumes of data, on which statistical learning is often possible, are one of the reasons behind the widespread use of probabilistic and statistical methods in image fusion.

The complexity is also a result of the strong heterogenity of information to combine, whether it is images taken from different sensors, images and models, or different characteristics extracted from one or several images. This disparity is found both in the nature of the information and in their representation. The data can be frequent, for example, when dealing with a typical case for a given application, for which it is possible, for example, to obtain statistical data. It can also be scarce (for example, pathological images) and in this case it is much more difficult to model them in a statistical way. The combination of these two types of data is common in image processing. Furthermore, they can be factual (typically, a photography of a scene at a given time) or generic (a model, rules, general knowledge about the application). Combining elements of information with different specificities often leads to conflict problems to solve. In image processing, this is not an easy task because factual information is not always reliable and accurate enough for it to systematically be given the priority over less specific and more generic information that may allow exceptions.

The combination of information is often guided or constrained by additional information regarding the information to combine, the context and the field of application. It is also a source of strong heterogenity. One example of additional information on the information is the reliability of a source, either overall or conditional to the objects observed. This is a very common case in the classification of multi-source images, where an image may be reliable for one class but not for another. Here are a few examples of additional information about the subject and the context:

– rivers are dark in SPOT's XS3 channel (information relating the type of acquisition with an observation);

– the CSF is dark in MRI images in T1;

– roads cross each other to form intersections (integrity constraint).

This generic information is used to guide the fusion process. The last example is a typical case of a rule with exceptions. The rule gives the most general case, but is not true in the case of dead-ends, for example.

Active fusion is one of the means for reducing complexity by choosing at every instant the best information to fuse. This choice can be performed based on a partial result of fusion obtained at a previous stage, on information measurements, on outside information likely to guide the fusion, on the identification of ambiguities that have to be cleared up, etc.

In image processing, fused information is necessarily tainted with imperfections (uncertainty, imprecision, incompleteness, ambiguity, conflict, etc., according to the distinctions proposed in Chapter 1). These imperfections originate on different levels, from the observed phenomena to the processes. For example, the smooth transition from healthy tissue to pathological tissue is an imprecision caused by the physiological phenomenon. Likewise, similar characteristics between two different kinds of tissue show up on images that measure this type of characteristic, resulting in doubts as to whether a particular point belongs to one kind of tissue or the other, which is an uncertainty due to both the phenomenon and the sensor. The delocalization of the spatial information, which is due to the fact that all of the information contained in a volume is grouped together in a same pixel, is caused by the sensor and its resolution, and constitutes an imprecision regarding the location of the information in the image (partial volume effect). Gibbs phenomena at the level of clear transitions, occurring in MRIs or radar imagery, for example, are a source of imprecision caused by the digital reconstruction algorithms used on the images. The representation of (symbolic) information in schematic form (with maps or atlases) is a source of both imprecisions and uncertainties. These are then magnified by the primitives extracted from the images, used as the basis for the fusion. The most familiar example is contour detection using Gaussian filters on different scales: when increasing the Gaussian's standard deviation, we get a higher certainty regarding the presence of contours, but we lose accuracy regarding their location. This antagonism between accuracy and certainty has been well identified as a characteristic trait of the approach in shape recognition [SIM 89]. This antagonism often gives rise to contradictions in image fusion, since there are several measurements available for one event: if the data is accurate, then it is probably uncertain, and might contradict itself; if the certainty is increased, this often comes at the price of more imprecision, which renders the data less informative if this imprecision is too great. Fusion therefore requires a decision system for explicitly managing uncertainty and imprecision in order to avoid inconsistencies.

Imprecision is not a feature specific to the data, but it can be related to the objectives and goals, especially if they are expressed in a vague linguistic form.

Finally, the spatial nature of information, specific to image processing, deserves particular attention. Its introduction in fusion methods, often inspired by other fields lacking this spatial nature, is not immediate and yet necessary in order to ensure the spatial consistency of the results. Imprecision is also present at this level. On a low level, it consists of problems of registration or partial volume, for example. At a higher level, it consists, for example, of relations between objects that can be intrinsically vague or poorly defined (such as a relation like "to the left of").

As with other fusion applications, redundancy and complementarity between the images we wish to fuse are assets in reducing imperfections such as uncertainty and imprecision, clearing up ambiguities, completing information, solving conflicts. Here are a few examples of complementarity in image fusion:

– involving the information itself: hidden parts that can be different in depth images or aerial images;

– involving the type of information: anatomical information or functional information for the same subject in different imaging modes;

– involving the quality of the information: two images of the same type, but different acquisition parameters can lead to elements of information of different qualities for different structures.

Redundancy is caused more obviously by the fact that the images we wish to combine describe the same scene. For certain fusion problems, such as group studies in functional imaging, redundancy (which areas are activated by all of the subjects) and complementarity (where are the differences) are subjects of study in themselves.

Conflict is a very delicate matter, as with other applications of information fusion, as we discussed in Chapter 1. With images, examples of conflicts that are only apparent and easily confused with complementarity occur when an image is not capable of distinguishing two classes whereas another one can. Imprecision and uncertainty are also sources of conflict. For example, a poorly localized contour can cause a conflict between several contour detectors. Conflicts due to the different specificities of the elements of information to combine are common in image and model fusion applications. For example, recognition of brain structures by fusion of MRI images and data found in an anatomical atlas must deal with variability among individuals, which is often not represented in the atlas, or also the possible presence of pathologies in the patient's images, which are not found in the generic model. Similar problems occur in the fusion problems of aerial and satellite images with digital maps. In this case, the conflict can be due to an imprecise drawing of the map, to modifications of the scene not included on an older map, etc.

3.4. Constraints

There are several types of constraints specific to image processing that have to be taken into account.

From the perspective of the fusion system's architecture, decentralized systems are rarely imposed. The most common case is that of *off-line* fusion, in which all the elements of information are available simultaneously. Centralized systems can then be used.

Real-time constraints are fairly rare, except with surveillance or multimedia applications, in which they are destined to play a growing role. We will come across such constraints again, but in a much stronger form, in the parts of this book that describe fusion in robotics, for example.

On the other hand, spatial consistency constraints are very stringent and constitute an important subject of research in image fusion. An increasing amount of studies focus on taking into account spatial information, either on a local level by way of the spatial context, or on a more structural level by way of spatial relations between the structures or the objects of the scene.

The fact that the data is growing in volume can exert some constraints on the calculation time. Thus, handling single pixels is limited to simple operations. More complicated operations often require including information of a higher level and a more structural representation of the information.

The complexity and amount of data dealt with often requires a choice of the information and knowledge to fuse. This choice is of course guided mainly by relevance criteria with regards to the decision objectives, but also by criteria involving the difficulty in accessing and representing information and knowledge, as well as their quality.

Finally, evaluation is a crucial problem. In image fusion problems, there is usually a "truth" or a "right" solution, but it is often difficult to find. Thus, evaluating and validating a fusion method can only be done by using simulations, phantom acquisitions, or by comparing with a manual decision. This situation is different from problems of vote or social choice, where there is no truth and the goal is to find a "best" solution expressed as a compromised, based on equity and ethics criteria.

3.5. Numerical and symbolic aspects in image fusion

If the numerical aspect of information handled in image processing is obvious, its symbolic aspect deserves a little more attention.

The symbolic information can be bound to the data we wish to fuse (for example, visual information in a map or an anatomical atlas, attributes calculated using the data or objects previously extracted from the data) or to the knowledge of the field. Typically, the information on the field is often represented by rules, structural representations such as graphs, often used in shape recognition in images, constraints that need to be taken into account in the algorithms. Structural information can, for example, specify that a road network can be represented by a graph using roads and intersections, or it can express in the form of propositions general rules about the scene such as "brain ventricles are always inside the white matter", etc. Structural information can also be represented in the form of icons and therefore in a way similar to images, for example, in the case of digital maps or anatomical atlases. The geometric representations of the structures are then combined with the nature or the semantics of these structures. However, the information in this field can also be purely numerical when it consists, for example, of specifics about the acquisition, such as the wavelengths used in satellite imagery, or acquisition times in medical imaging. Hybrid representations, in which numbers are used as symbols with a quantification, are used in image fusion to quantify the quality of a sensor, to evaluate symbolic data or the confidence in a measurement, the reliability of an image for certain classes or structures, etc. From a processing perspective, this hybrid aspect arises, for example, when a proposition establishing that the recognition of a structure only depends on the local context leads to an adequate modeling in a Markovian framework. Symbolic representations are also used as *a priori* knowledge, or contextual or generic knowledge to guide the numerical process. They also serve as a structural medium, for example, when fusing images and maps [MOI 95].

3.6. Bibliography

[MAI 91] MAÎTRE H., "Utilisation de l'imagerie aérienne et satellitaire pour l'aménagement du territoire", *Cours du 8ème Congrès AFCET de Lyon*, Tutorial 4, 1991.

[MAN 94] MANGIN J.F., FROUIN V., BLOCH I., LOPEZ-KRAHE J., BENDRIEM B., "Fast Nonsupervised 3D Registration of PET and MR Images of the Brain", *Journal of Cerebral Blood Flow and Metabolism*, vol. 14, p. 749-762, 1994.

[MOI 95] MOISSINAC H., MAÎTRE H., BLOCH I., "Graph Based Urban Scene Analysis Using Symbolic Data", *SPIE Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision II*, vol. 2486, Orlando, Florida, p. 93-104, April 1995.

[SIM 89] SIMON J.C., *From Pixels to Features*, V-X, North Holland, Amsterdam, 1989.

[ZIT 03] ZITOVA B., FLUSSER J., "Image registration methods: a survey", *Image and Vision Computing*, vol. 21, p. 977-1000, 2003.

Chapter 4

Fusion in Robotics

4.1. The necessity for fusion in robotics

Robots are mechanical objects whose purpose it is to replace humans either in tasks that may be simple, repetitive or strenuous (a welding robot for automobile manufacturing), or in environments that may be dangerous (demining robots, contaminated environment robots) or inaccessible (Martian robots, underwater robots, micro-robots for surgery) [ROB 02]. One of the characteristics of robots is that they maneuver in the same type of environment as a human being, i.e. a three-dimensional space comprised of elements in situations that can change over time.

A robot's ability to move is related to its ability to perceive the environment, in order to know it (exploration), to move inside it (localization) and to interact with it (action) [ABI 92, KOS 93].

Perception is usually achieved by way of sensors embedded in the robot itself, much like a human being is equipped with eyes, skin to allow them to sense objects, with a balance mechanism (the inner ear) to allow them to maneuver inside their environment [FRA 00]. As with human beings, there are many different kinds of sensors and, usually, they make it possible to obtain complementary information on the robot's state, its situation with respect to the environment, and the state of that environment.

The perception system is closely related to the mission the robot must achieve and to the prior knowledge it has of the environment.

Chapter written by Michèle ROMBAUT.

In the course of exploration, the robot's mission is to maneuver inside the environment in order to construct a map of it. This map may have different semantic levels, cells dividing the environment and which have to be recognized as occupied or empty, or even be used for recognizing and locating objects with more complex structures, such as a hole in a mechanical structure in which a pin has to be inserted, a vehicle on a road, or a door in a room.

In the course of localization, the robot has at its disposal a map of the environment that generally corresponds to reality in terms of accuracy (objects that are displaced or not well positioned), but also of reliability: the existence of an object that does not appear in the map (false positive) or the absence of an object listed on the map (false negative). Sometimes, the environment is comprised of objects that can also move in the environment. In that case, it is obviously impossible to have a map beforehand representing the environment at every instant. However, if evolutionary models of these objects are available, as well as their states at a given time, it is possible to predict the map's state in a "near" future, i.e. in a short time horizon compared to the time constant of the observed systems.

In the course of the action, based on the knowledge of its current state and of the objective it has to accomplish, the robot generates a trajectory that is feasible in its environment, in other words a control to transmit to its actuators and the observations needed to "properly" proceed with this trajectory [KOS 93].

4.2. Specific features of fusion in robotics

4.2.1. *Constraints on the perception system*

By definition, the robot must maneuver and operate inside its environment, which leads to risks. Therefore, it is necessary for the information obtained with the perception system to be accurate and reliable enough to ensure the safety of the robot, of the environment and of the other users in this environment, particularly human beings.

Sensors are embedded in the robot and their numbers are therefore limited, as well as their performances in time and in space, in size and in energy resources. As a result, the data obtained from the sensors is necessarily limited and the lack of information is what is dealt with, rather than redundancy.

4.2.2. *Proprioceptive and exteroceptive sensors*

Just as a human being needs to know its own state before considering a movement (sitting, standing, carrying weight, off balance), it is necessary to know a certain number of parameters regarding the robot's static and dynamic state. In order to do this, the robot has to be equipped with proprioceptive sensors such as an inertial measurement unit, accelerometers, odometers, etc. In many cases, there are also mechanical models of the robot available, which make it possible to complete the knowledge of the system's state. For example, the direct model of a manipulator robot gives us the position and orientation of its effector (tool) depending on the position value of each of its axes. Likewise, knowledge of a robot's kinematic model, based on the rotation speed of the wheels, can be used to evaluate its speed.

Before and during a movement, the robot has to localize itself with respect to its environment. In order to do this, it is necessary for it to perceive objects using sensors that are adequate to this environment, which are referred to as exteroceptive sensors. Under water, for example, the use of sonar will be preferred; for driving at night, an infrared camera will be used. The books by [CHA 98, ZHA 97] give a rather thorough presentation of everything pertaining to perception in intelligent vehicles.

4.2.3. *Interaction with the operator and symbolic interpretation*

A robot's purpose is to maneuver in space instead of or alongside humans in order to replace or to help them in achieving their mission. The extent of the interactions with the human operator vary depending on the cases. For example, when driving an automobile or teleoperating, the human operator stays in direct control of the robot (or of the vehicle). The role of the embedded perception system is then to transmit information to this operator in order to facilitate their control task. On the contrary, when performing more automated tasks, for example, with welding robots or Martian robots for which real-time control is impossible, the information perceived is essentially used by the robot itself. The information transmitted to the human operator only corresponds to status reports on the progress of the mission, on alarms or on the robot's situation.

Depending on their use (human or robot operator), the semantic nature of the information will be quite different, i.e. more symbolic for the human operator (the object is close, the speed is high), and more numerical if they are directly used by the robot (distance from the tool point to the closest point on the table = 20 cm).

4.2.4. *Time constraints*

The robot maneuvers inside the environment at a certain speed. The environment can also change with time (other robots or vehicles) at varying speeds. In order for the robot to achieve its mission, the frequency of observation of its state and the state of its environment needs to be high enough (Shannon's theorem). For safety reasons, it also has to be able to perceive "events" that require a rapid response or action.

Therefore, the processing of the information obtained from the sensors has to be quick and efficient in order to provide information in time, even if its quality is reduced. Many research teams are working on architectures dedicated to this type of application.

As for the methods that have been developed, several solutions can be suggested.

Data extrapolation. When the data is acquired, it is dated, either by a clock common to all of the sensors, or by different clocks that are synchronized periodically. Based on a sequence of data and using an evolutionary model, also called a dynamic model, it is often possible to extrapolate the observed variable, even if its quality deteriorates over time, from lack of observations. This is typically what is done with tracking filters, such as the Kalman filter [BAR 88, KOK 94, STE 98]. The prediction can be used to obtain an element of information with an error greater than that of the initial data, but which corresponds to the current date. This method is described at length in section 2.2.1.

Focusing. Depending on the context, we might need to obtain certain elements of information that seem more important than others. We would then have to focus the system's perception abilities on these elements, even if that means overlooking others that may also be available. For example, when a moving robot is about to go through a doorway, the areas where the two jambs are located are observed in priority, even if a watch is maintained to detect possible obstacles in front of the vehicle. The technique is used in particular in the case of information in the form of an image, where only a part of the image is subjected to significant processing. Focusing requires being able to identify what is relevant based on the estimate of the situation, or on action objectives that have been defined. Depending on what is relevant, a perception strategy will then have to be defined in terms of the sensors used, the choice of the data and the algorithms that are to be used. This method has been widely used at the LASMEA laboratory in Clermont-Ferrand for aiming, based on the results of image analysis, a laser beam capable of measuring the distance between two automobiles [CHE 96, TRA 93]. Finally, the technique is the same as that presented for the identification of military targets described in section 2.2.3.

Sensor based control. This method consists of foregoing the interpretation of data, which is often very costly in terms of time and computing power, by computing the control directly from the information obtained from the sensors before they undergo any extensive processing. Depending on the robot's situation and the mission it has to achieve, a control strategy is defined based on the sensor's measurements. For example, if a moving robot is located in a hallway and has to maneuver between the two walls that confine it, we can impose the order that the distances between each of the walls and the robot have to be identical. The robot's movements are then directly calculated based on the difference in the measurements by the sensors that provide these two distances. This method can help substantially in speeding up the control loop, but it requires, however, an assessment of the situation at a higher semantic level at times when the strategy is changed.

4.3. Characteristics of the data in robotics

4.3.1. *Calibrating and changing the frame of reference*

The data that is processed in the context of robotics is, for the most part, geometric in nature. It essentially involves distances as well as their successive derivatives (speed, acceleration, radius of curvature, etc.). We will have to localize a frame of reference associated with the robot with respect to the frame associated with its environment (absolute localization), or with respect to the same frame related to the robot but with the position it was in a few moments before (relative localization). We will also have to localize the frames associated with fixed or moving obstacles with respect to the robot's frame.

One frame of reference at least is always assigned to the robot, only one if it is not deformable, two or more if it is. For example, with a manipulator robot, we define the basic frame, which does not move in the environment and the tool (or effector) frame, which allows us to localize the operating part of the robot in the situation where the mission could be performed. With regards to manipulator robots, we should mention the book by Khalil [KHA 99] which covers the basics of modeling. The same techniques can be applied to mobile robotics.

Sensors located on the robot or in the environment also have their own frame of reference, based on which the measurement is defined. If we want the robot to be able to use the measurement, it has to be referenced to the frame it is associated with, which is usually different from that of the sensor's. Using a calibration method that is rarely trivial, the goal is to define the transformation relating the two frames, i.e. robot and sensor. This transformation is not perfectly known, leading to measurement errors. These errors, which are difficult to estimate, are usually neglected when the data is used.

Figure 4.1 shows an example of a situation where a robot is located in an inside environment. $R(k)$ is the situation of the frame associated with the robot at the time k. The robot is equipped with two sensors associated with the frames R_{sensor_1} and R_{sensor_2}. We will assume that a map of the robot's environment is available, which is associated with the frame R_{map}. This environment contains two walls, the characteristics of which (length, orientation) can be known, as well as their localization on the map using the frames they are associated with, i.e. R_{wall_1} and R_{wall_2}. The entire system can be referenced with respect to a universe frame R_u. Let us assume that we wish to localize the robot at the time k with respect to the universe frame. This means we have to find the geometric transformation $T_{R_u}^{R(k)}$ relating these two frames.

We know the map's localization with respect to the universe frame $T_{R_u}^{R_{map}}$ as well as the localization of the walls with respect to the map $T_{R_{map}}^{R_{wall_i}}$. The sensors give us the localization of the walls with respect to the sensors $T_{R_{sensor_j}}^{R_{wall_i}}$ and the calibration provides us with the localization of the sensors with respect to the frame associated with the robot $T_{R(k)}^{R_{sensor_j}}$. We can then infer that:

$$T_{R_u}^{R(k)} = \left(T_{R_u}^{R_{map}}\right) \cdot \left(T_{R_{map}}^{R_{wall_i}}\right) \cdot \left(T_{R_{sensor_j}}^{R_{wall_i}}\right)^{-1} \cdot \left(T_{R(k)}^{R_{sensor_j}}\right)^{-1}$$

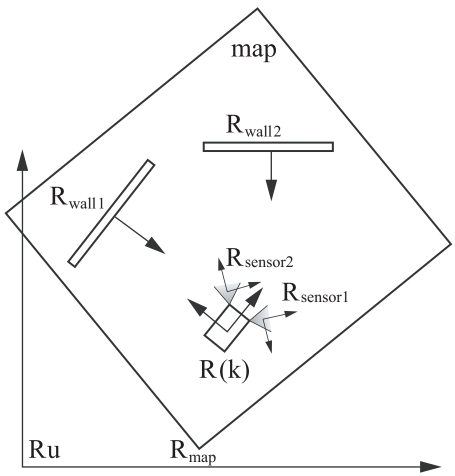


Figure 4.1. *Localization of a moving robot in its environment which is described in a map*

This simple example underlines the typical problems that have to be solved in robotics:

– the determination of transformations $T_{R(k)}^{R_{sensor_j}}$ by calibration;

– the evaluation of the transformation $T_{R_{sensor_j}}^{R_{wall_i}}$ based on sensor information, which requires signal and image processing;

– matching the "walls" that are seen with those that are known and on the map;

– the fusion of information from different sensors that see the various walls.

4.3.2. *Types and levels of representation of the environment*

In order to define a strategy for action or perception, it is often necessary to rely on an interpretation of the situation based on symbolic data with a high semantic level. For example, we can be faced with situations such that:

– the effector is close to the pin;

– the robot is reaching the end of the hallway;

– the vehicle is being overtaken.

The situation is characterized by symbolic data evaluated from usually numerical data obtained from the sensors. This evaluation requires the use of models that describe the symbols based on their numerical values:

– geometric description of the pin;

– meaning of “close”;

– geometric description of a hallway;

– meaning of “reaching the end of the hallway”;

– localization of a vehicle in the lanes;

– meaning of “being overtaken”.

The confidence that can be assigned to the various symbolic assertions depends on the quality of the measurements (accuracy, reliability) and of the symbol description models.

4.4. Data fusion mechanisms

Data fusion in robotics relies on the use of traditional tools (probabilities, crisp or fuzzy sets, belief theory). Applications in robotics are different from other applications, especially when it comes to the structure of the fusion and more generally of the perception.

First of all, we have to properly identify the objectives of the fusion in order for them to be consistent with the mission the robot has to achieve. For example, when the robot is located far away from the elements that comprise its environment, no effort will be made to precisely localize them since they do not pose an immediate threat. However, we need to be sure that the space chosen for movement is clear of obstacles. The objectives of perception can change over time, requiring that the fusion system, as well as the sensors and the processing algorithms, are able to adapt to the situation.

The aspect of time is fundamental in robotics. Many fusion algorithms use evolutionary models such as differential equations, recursive equations, graphs and sequential logic rules. It is then possible to fuse the knowledge acquired previously and registered with the present moment with newly acquired information. Some methods also use single or multiple target tracking algorithms [BAR 88, KOK 94, NIM 00].

Finally, the security aspect has to be taken into account because the information obtained from the fusion is generally used for moving the robot. If it is erroneous or not accurate enough, it can lead to the robot colliding with its environment. Therefore, it is essential to rate the data in terms of accuracy and certainty. If the quality of the data is insufficient, the perception and control strategies of the robot have to be adapted.

4.5. Bibliography

[ABI 92] ABIDI M.A., GONZALEZ R.C., *Data Fusion in Robotics and Machine Intelligence*, Academic Press, 1992.

[BAR 88] BAR-SHALOM Y., FORTMANN T., *Tracking and Data Association*, vol. 179, Academic press, 1988.

[CHA 98] CHAVAND F., COLLE E., *Perception de l'environnement en robotique*, Hermès, Paris, 1998.

[CHE 96] CHECCHIN P., Segmentation d'images de profondeur, PhD Thesis, Blaise Pascal University, Clermont-Ferrand, 19 December 1996.

[FRA 00] FRANKEL C., BEDWORTH M.D., "Control, Estimation and Abstraction in Fusion Architectures: Lessons from Human Information Processing", *International Conference on Information Fusion*, vol. 1, Paris, p. MoC5 3–10, 2000.

[KHA 99] KHALIL W., *Modélisation, identification et commande des robots*, Coll. Robotique, 2nd Ed., Lavoisier, 1999.

[KOK 94] KOKAR M., KIM K., "Centralized Multisensor Fusion Algorithms for Tracking Applications", *Control engineering practice*, vol. 2, no. 5, p. 803–809, 1994.

[KOS 93] KOSAKA A., MENG M., KAK A.C., "Vision-Guided Mobile Robot Navigation Using Retroactive Updating of Position Uncertainty", *IEEE International Conference on Robotics and Automation*, Atlanta USA, 1993.

[NIM 00] NIMIER V., BASTIÈRE A., COLIN N., MORUZZIS M., "MILORD, an Application of Multifeature Fusion for Radar NCTR", ISF, Ed., *International conference on information fusion*, Paris, p. WeD1-17–WeD1-24, 2000.

[ROB 02] "Les nouveaux robots", special edition, *La Recherche*, February 2002.

[STE 98] STEINBERG A.N., BOWMAN C.L., WHITE F.E., "Revision to the JDL Data Fusion Model", *Joint NATO/IRIS Conference*, Québec, Canada, 1998.

[TRA 93] TRASSOUDAINE L., Solutions multisensorielles temps réel pour la détection d'obstacles sur route, PhD Thesis, Blaise Pascal University, Clermont-Ferrand, 16 February 1993.

[ZHA 97] ZHAO Y., *Vehicle Location and Navigation Systems, Intelligent Transportation Systems*, Artech House, Inc., 1997.

Chapter 5

Information and Knowledge Representation in Fusion Problems

5.1. Introduction

In this chapter, we will briefly present the different modes for representing information and knowledge used in fusion, as well as how they are integrated into systems. Because numerical representations rely on the theories of probability, belief functions, fuzzy sets and possibility, they will be discussed again in greater detail in Chapters 6, 7 and 8. Knowledge-based systems, which can be used to structure information, knowledge and inference modes in order to combine them, will be presented only in broad strokes. They will not be discussed in detail in this book, but an example of a multi-agent system will be presented in Chapter 10. Symbolic approaches, as well as reasoning modes in different logics, will only be mentioned. They go beyond the scope of numerical fusion; however, their properties would deserve more attention in information fusion in signal and image processing.

5.2. Processing information in fusion

As we said in Chapter 1, we consider the word information in the broadest sense. Thus, the term information can be applied to any element that might be coded in order to be stored, processed or broadcast [DUB 01]. In signal and image processing, it often consists of information related to real worlds (observations, measurements, generic knowledge regarding real phenomena, etc.), but it can also consist of virtual worlds, in the expression of a user's goals and preferences.

Chapter written by Isabelle BLOCH and Henri MAÎTRE.

This generic concept is usually divided into two categories, i.e. knowledge, in reference to classes of objects, and data, which corresponds to cases, facts, or particular objects.

According to the distinction suggested in [DUB 01], information can be comprised of data, facts, and then involves a particular, well-determined situation (*there is a forest in that place*). It can relate to existence (or some other property) of an indeterminate situation (*there are some areas covered by forests*). Comprised of statistical data, prototypes or typical examples, it involves a set of particular situations. Finally, it consists of classes of situations and take the form of constraints, generic rules, with or without exceptions, and general knowledge.

The information handled in fusion is most often imperfect. These imperfections are in fact one of the reasons for fusion. These imperfections manifest themselves in multiple forms: ambiguity, bias, noise, incompleteness, imprecision, uncertainty, inconsistency and conflict, etc. We also have to mention the varying and evolving nature of information that relates to the dynamic world (see Chapters 4 and 11).

We have already mentioned in detail these imperfections in Chapter 1 and then specified them in Chapters 2 for the information handled in signal processing, 3 in images and 4 in robotics. To sum up, they are caused by:

– the observed phenomena;

– the limits of the sensors;

– the reconstruction and processing algorithms;

– noise;

– the lack of reliability (often a result of the previous limits);

– the representation mode;

– the knowledge and concepts involved.

What matters in fusion processes is to include these imperfections in the representations and in the reasoning modes.

The questions related to fusion are similar to those related to information processing in general [DUB 01]. Thus, the objective is to:

– represent the information (in order to express it in a useful form);

– store, retrieve and make the information explicit;

– use the information to decide and act;

– communicate the information.

The difficulty in solving these problems is of course greater because of these imperfections. There are three ways of reacting to this:

1) a first attitude consists of eliminating imperfections as best as possible. This involves, for example, improving sensors and increasing the number of acquisitions;

2) a second possible action is to tolerate the imprecision by producing robust algorithms and programs, and by combining them with procedures for repairing failures;

3) the third possibility is to try to reason with the imperfection. In this case, it is considered as a type of knowledge or information and taking it into account requires modeling it, developing approximate modes of thought, and using meta-knowledge, i.e. knowledge about these imperfections.

In this book, we will prefer the third approach, which explicitly involves techniques of information fusion and decision making.

5.3. Numerical representations of imperfect knowledge

The major numerical theories that allow us to represent imperfect knowledge and to use them as the basis for our approach are:

– probabilities (Chapter 6);

– belief functions (Chapter 7);

– fuzzy sets and possibilities (Chapter 8).

In probabilistic representations, language is comprised of probability distributions in a frame of reference. They allow us to rigorously take into account random or stochastic uncertainties. It is more difficult to take into account other forms of imperfections, both formally and semantically. Bayesian inference, often used in fusion in the subjects we are concerned with, serves as the basis for abductive reasoning (the different types of inferences are presented in section 5.6).

Belief function theory (or the Dempster-Shafer theory [SHA 76]) relies on a language defined by functions (referred to, in this context, as mass, belief and plausibility functions) over every subset of the frame of discernment. Using representations, we can take into account at the same time imprecision and uncertainty (including its subjective form), ignorance, incompleteness and have access to conflict. Inference based on Dempster's rule achieves conjunctive aggregation of the combined information.

In fuzzy set and possibility theories [DUB 80, DUB 88, ZAD 65, ZAD 78], language is comprised of fuzzy subsets of the frame of reference or possibility distributions over the frame of reference. It allows us to represent qualitative, imprecise, vague information. Inference is done according to logical rules (or their equivalent in numerical form), essentially by deductive reasoning that may be qualitative.

We will discuss these three theories again in detail in the following chapters, but for the moment, this is what matters:

– they do not model exactly the same concepts or the same aspects of the information;

– they do not have the same semantics;

– they do not have the same power of representation;

– they do not have the same reasoning power.

In particular, the first two points make it illusory and misleading to want to compare their performances on the same applications[1].

These remarks are also a motivation for hybrid representation techniques, which allow us to simultaneously represent elements of information with different types of imperfections. It is also possible to define the probabilities of fuzzy events, of the belief functions of fuzzy subsets, etc. However, these approaches are still rarely used in information fusion.

5.4. Symbolic representation of imperfect knowledge

Artificial intelligence is traditionally defined (in Minsky's and McCarthy's works, for example) from two points of view:

– from the cognitive point of view: this consists of constructing computable models of cognitive processes, in other words programs that can simulate human performances;

– from the computer science and engineering point of view: this consists of assigning to computers tasks that would be considered intelligent if performed by a human, in other words extending the abilities of computers.

Artificial intelligence generated symbolic representations of knowledge. The field of knowledge representation is characterized by:

– the definition of a representation as a set of syntactic and semantic rules to describe an element of knowledge;

– logical representations (the expressivity depends on the logic used; see section 5.6);

– compact representations (only the relevant and characteristic properties are stated explicitly);

– ease of use;

– what is important is actually explicit.

1. This partly explains the contradictory conclusions found on this subject in other works.

Most of the data handled in fusion in the fields of signal and image processing is analog or digital. Analog descriptions require a complete description of the world. Switching over to logical representations, which are cheaper and more compact, requires converting analog representations into symbolic representations.

Symbolic representations have requirements on several levels:

– the ontological level: all of the important concepts have to be taken into account;

– the epistemic level: we should not have to express what is not known;

– the computational level: the representation should allow for an efficient computation of the properties expressed.

The first two levels induce constraints on the language and the third on the inference mechanisms.

The knowledge representation (symbolic) community focuses on non-monotonic reasoning, automatic reasoning, logic descriptions, subjective representations (preferences, wishes, etc.), ontologies, etc. [REI 91]. Closer to what concerns us, it also focuses on learning, the integration and fusion of knowledge bases, decision and diagnosis, temporal and spatial reasoning, the representation of actions and planning. There are definitely directions to explore in this direction for the fusion that concerns us.

5.5. Knowledge-based systems

The evolution of image processing, from the lowest level of signal processing to the interpretation of complex scenes, quite naturally leads the user to taking into account knowledge beyond merely the image signal. As a consequence, this brings knowledge management techniques in contact with image processing. These are methods developed in artificial intelligence and which have been referred to as: expert systems, knowledge-based systems (KBS), multi-agent systems, etc.

These techniques have had a strong influence on the development of image processing techniques. In particular, they contributed to important projects that were part of European programs. We suggest discussing here the lessons that can be learned from these methods.

The objectives of these methods are as follows:

– representing knowledge in a declarative and not just procedural way as in the most common image processing algorithms;

– separating knowledge into different categories:
  - factual knowledge separated from operational knowledge,
  - particular knowledge separated from general knowledge;

– using the same knowledge to achieve different objectives;

– creating universal, and therefore reusable inference mechanisms;

– keeping records (explanations) of the inferences obtained.

Such systems can be seen as extensions of the usual expert systems, by allowing the use of different modes of knowledge representation and reasoning.

The most common method used is shown in Figure 5.1, including the major functions:

– the supervision system in charge of the overall system management, and of the use of rules (when and how to use them);

– an interfacing system that allows the user to specify particular rules for a given problem, or to set the objectives for a session;

– an inference engine that activates the processing rules, and generates the system's behavior;

– and of course a knowledge base that contains two different categories of knowledge: the facts on one side (in the case of an image processor, this is where the images are), and the rules on the other side, indicating how the facts should be combined in order to infer new knowledge.

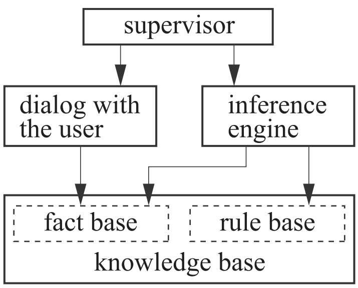


Figure 5.1. *The major functions of a KBS*

Building a KBS requires agents with three distinct roles:

– the user fills the system with data to be processed;

– the expert builds the knowledge base;

– the developer builds the inference engine and the reasoning strategy.

The most commonly used model in image processing is based on the "blackboard", i.e. an area shared by all of the system's users to drop in information obtained while the system is operating. The blackboard of an image processing expert system is often the image itself, or at least a multi-layered extension of the image, with each layer referenced on the image and bearing the results of a category of operators or experts applied to a specific area inside the image.

The different types of knowledge represented in these systems are divided into the following categories:

– declarative (how things are);

– procedural (how things are done);

– episodic (related to the previous experience);

– and meta-knowledge (knowledge about knowledge).

Generally, it is extremely difficult to change from one type of knowledge to another (and often even impossible) and therefore these categories of knowledge are, by nature, quite different and may all be necessary.

Control consists of searching for paths between initial knowledge and goals, using techniques called forward chaining (applying inference rules when new data is declared, with the consequences possibly triggering new inference rules), or backward chaining (applying inference rules when new requests are stated, in which case the premises of these rules that have not been verified generate new rules).

The most common examples of KBS include:

– production rules;

– frames;

– semantic networks;

– systems with uncertainty: Mycin [SHO 76], etc.

Production rule systems (of the type *if... and/or... then...*) are systems that are easy to adapt or extend, and the way they operate and their results can easily be explained. They have the drawback of having a fragmented representation of knowledge which causes a lack of efficiency. Their power of expression depends mostly on the type of logic used. For example, first-order or predicate logic make it possible to handle variables and quantifiers, whereas in propositional logic, everything is constant.

Frames constitute a declarative form of KBS in which a list of attributes or properties is supplied with characteristics and values for these characteristics. They are useful in describing general concepts, classes of objects. Classes with different granularities can be handled using links of hierarchy, inheritance, specialization and instantiation. Most of the time, these systems are static, but some dynamics can be introduced by assigning procedures to attributes.

Semantic networks rely on a graphic representation of the knowledge base, in which the nodes represent concepts and objects, and arcs represent relations. The inference rules are based on inheritance properties when taking an arc from one class to a more specific class. These networks are often used in natural language processing, for example.

In image viewing and processing, for which we operate in environments that are not completely specified and only partly known, we have examples of knowledge-based systems essentially designed for supervising programs [CLE 93, NAZ 84, THO 93] and for interpreting images [DES 90, GAR 89, HAN 78, MAT 86, MCK 85]. Quite generally, they are comprised of the following components: isolated points of the image, contours, areas. Semantic relations also come from artificial intelligence: *Part-Of*, *Is-a*, *Instance-of* are the most common. It is much more complicated to implement more complex structures, though they are quite useful to image processing, such as concepts of hidden areas, shadows, multi-scale representation and frequential components. Supervision strings together the expert's task based on the objectives and the results that have already been obtained, and the distinction is made between:

– systems guided by data, which put together, for example, pieces of contours to form lines, then lines to form objects, etc., until they end up with a known object;

– and systems guided by goals, which start with the hypothesis of the result and recursively search for the previous steps necessary to the presence of this hypothesis.

The prototype of such a system is well represented by Hanson and Rieseman's VISION system [HAN 78, HAN 88]. A slightly similar method is found in many other systems. The processing phase is separated in levels, as illustrated in Figure 5.2: typically a high level (level 3) which contains the symbolic descriptions and recognized objects, then an intermediate level (level 2) which contains areas, lines, shapes, and finally the low level (level 1) which contains images and their pixels. Operators use the data on the level $n-1$ to create information on the level n. The other way round, requests are made by the level n on the level $n-1$ in order to accomplish a certain number of tasks necessary to decision making. Also, processes are constantly operating within each level to organize, arrange in order and fuse the information.

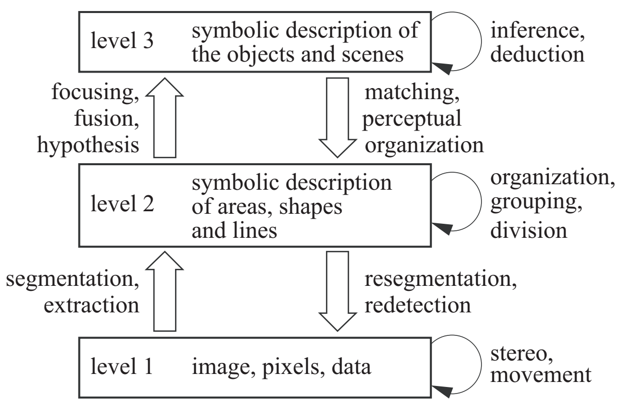


Figure 5.2. *Diagram explaining how a KBS works. This representation is based on VISION [HAN 88]*

In vision, specific tasks of focusing and adapting (with their attentional mechanisms, for revision or repairs and consistency management), co-operation and fusion (confrontational, augmentative, integrative) and co-ordination (deliberative, reactive, optimal) are added to KBS. These methods are discussed in detail in [GAR 01] and will not be covered here.

An example of a multi-agent system will be given in Chapter 10.

5.6. Reasoning modes and inference

There is a wider variety of inference modes used in KBS than in traditional knowledge bases. These modes are divided into the following categories:

– deductive reasoning, which provides consequences based on facts (for example, if the fact base contains A and the proposition $A \rightarrow B$, then we can infer B);

– contraposition allows us to reason on non-observations (for example, if we have $A \rightarrow B$ and non-B, we can infer non-A);

– abductive reasoning attempts to find the causes explaining the observations (for example, based on $A \rightarrow B$ and the observation of B, we infer that A is a possible cause of B);

– inductive reasoning allows us to infer rules from regular or usual observations (for example, if we have B every time we have A, we can infer $A \rightarrow B$);

– projection provides consequences based on actions (if the fact base contains the proposition $A \rightarrow B$ and we perform A, we expect B to occur);

– planning establishes which actions to perform in order to achieve goals (if we want B and the base contains $A \rightarrow B$, then we infer the action A).

These last two inference modes are particularly developed in embedded KBS, such as those used in mobile robotics [SAF 02].

Information fusion often requires the help of different reasoning modes, in order to better grasp and represent the nuances and subtleties of human reasoning.

In monotonic reasoning, obtaining more information naturally leads to more conclusions: if A is inferred from a base KB, we will also infer A from $KB \cup B$. Traditional, propositional and first-order logic resort to this mode of reasoning.

In non-monotonic reasoning, new information can invalidate previous conclusions. In the presence of imperfect knowledge and information, as is the case in information fusion, sources of non-monotonicity essentially come from the hypotheses and restrictions that are applied. These are necessary to be able to reason, but can be questioned if new information or elements of knowledge become available. These hypotheses that are sources of non-monotonicity include:

– the use of typical properties;

– possible exceptions;

– the closed world hypothesis, which will be discussed particularly in Chapters 7 and 8.

Non-monotonic logics rely on concepts of preferences (which world, which situation is more “normal” than others, what are the preferred goals if not all of them can be met, etc.), belief changes or revisions (the well-known AGM postulates [ALC 85]) and of course a certain number of postulates that manage non-monotonicity, referred to as rationality postulates. “Cautious” monotonicity, which is an example of such a postulate, expresses the fact that if a base KB can be used to infer $A \rightarrow B$ and C, then it can be used to infer $A \wedge C \rightarrow B$.

The concepts of contingency, i.e. of necessary or possible truth, are not well represented in traditional logic. Manipulating such objects requires the introduction of modalities in logic. Modal logics [CHE 80, HUG 68] allow us to reason on propositions A (A is true), $\Box A$ (A is necessary), $\Diamond A$ (A is possible). Numerical forms of these concepts are found in belief function theory in terms of belief and plausibility (see Chapter 7), and in possibility theory in terms of necessity and possibility (see Chapter 8). We will not discuss in detail the logical meanings of these concepts in this book.

Finally, the concepts of imprecision and uncertainty, which we have already discussed extensively, can be represented in fuzzy and possibilistic logic, which will be briefly described in Chapter 8.

5.7. Bibliography

[ALC 85] ALCHOURRÓN C.E., GÄRDENFORS P., MAKINSON D., “On the Logic of Theory Change: Partial Meet Contraction and Revision Functions”, *Journal of Symbolic Logic*, vol. 50, p. 510-530, 1985.

[CHE 80] CHELLAS B., *Modal Logic, an Introduction*, Cambridge University Press, Cambridge, 1980.

[CLE 93] CLÉMENT V., THONNAT M., “A Knowledge-Based Approach to Integration of Image Processing Procedures”, *CVGIP: Image Understanding*, vol. 2, p. 166-184, 1993.

[DES 90] DESACHY J., “ICARE: An Expert System for Automatic Mapping from Satellite Imagery”, in L. F. PAU (ed.) *Mapping and spatial modelling for navigation*, vol. F65 of *NATO-ASI*, Springer Verlag, Berlin, 1990.

[DUB 80] DUBOIS D., PRADE H., *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, 1980.

[DUB 88] DUBOIS D., PRADE H., *Possibility Theory*, Plenum Press, New York, 1988.

[DUB 01] DUBOIS D., PRADE H., “La problématique scientifique du traitement de l’information”, *Information-Interaction-Intelligence*, vol. 1, no. 2, 2001.

[GAR 89] GARNESSON P., GIRAUDON G., MONTESINOS P., “MESSIE: un système multi-spécialistes en vision, application à l’interprétation en imagerie aérienne”, *RFIA-AFCET’89*, Paris, France, p. 817-832, 1989.

[GAR 01] GARBAY C., “Architectures logicielles et contrôle dans les systèmes de vision”, in J.M. JOLION (ed.) *Les systèmes de Vision*, Chapter 7, p. 197-252, Hermès, 2001.

[HAN 78] HANSON A., RIESEMAN E., *Vision: A Computer System for Interpreting Scenes*, p. 303-333, Academic Press, 1978.

[HAN 88] HANSON A., RISEMAN E., WILLIAMS T., “Sensor and Information Fusion from Knowledge Based Constraints”, *SPIE Sensor Fusion*, vol. 931, p. 186-196, 1988.

[HUG 68] HUGHES G.E., CRESSWELL M.J., *An Introduction to Modal Logic*, Methuen, London, 1968.

[MAT 86] MATSUYAMA T., “Knowledge-Based Aerial Image Understanding Systems and Expert Systems for Image Processing”, *International Geoscience and Remote Sensing Symposium*, Zürich, p. 1026-1038, 1986.

[MCK 85] MCKEOWN D.M., HARVEY W.A., MCDERMOTT J., “Rule-Based Interpretation of Aerial Imagery”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, no. 5, p. 570-585, 1985.

[NAZ 84] NAZIF A., LEVINE M.D., “Low Level Image Segmentation: An Expert System”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 4, p. 555-577, 1984.

[REI 91] REICHGELT H., *Knowledge Representation: An AI Perspective*, Ablex Publishing, 1991.

[SAF 02] SAFFIOTTI A., Artificial Intelligence for Embedded Systems, Technical report, AASS, University of Örebro, Sweden, 2002.

[SHA 76] SHAFER G., *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[SHO 76] SHORTLIFFE E., *Computer Based Medical Consultation: Mycin*, Elsevier, 1976.

[THO 93] THONNAT M., CLÉMENT V., DEN ELST J.V., Supervision of Perception Tasks for Autonomous Systems: The OCAPI Approach, Report no. 2000, INRIA, Sophia Antipolis, France, 1993.

[THO 02] THONNAT M., “Knowledge-Based Techniques for Image Processing and for Image Understanding”, in J.P. ROZELOT and A. BIJAOUI (ed.) *Voies nouvelles pour l’analyse des données en sciences de l’univers*, p. 189-236, EDP Sciences, 2002.

[ZAD 65] ZADEH L.A., “Fuzzy Sets”, *Information and Control*, vol. 8, p. 338-353, 1965.

[ZAD 78] ZADEH L.A., “Fuzzy Sets as a Basis for a Theory of Possibility”, *Fuzzy Sets and Systems*, vol. 1, p. 3-28, 1978.

This page intentionally left blank

Chapter 6

Probabilistic and Statistical Methods

6.1. Introduction and general concepts

Probabilistic methods essentially deal with the uncertainty of information. They rely on solid and well-mastered mathematical theories in signal and image processing, such as Bayesian decision theory, estimation theory, entropy measurements, etc., thus making it one of the preferred tools for fusion.

Information and its imperfections (mostly those whose nature can be expressed in terms of uncertainty) are modeled using probability distributions or statistical measurements. We will see in section 6.2 how this formalism can be used to measure information. We will then describe the different stages of the fusion process: modeling and estimation in section 6.3, Bayesian combination in section 6.4, then Bayesian combination seen as an estimation problem in section 6.5. The most common rules of decision making are presented in sections 6.6 and 6.7. The following sections give examples of applications and other theoretical tools are discussed, in the fields of multi-source classification in image processing in section 6.8, then of target motion analysis in signal processing in section 6.9.

6.2. Information measurements

If we have a set of l sources of information I_j, a first task often consists of transforming it into a smaller and therefore easier to process subset, without losing any information.

Chapter written by Isabelle BLOCH, Jean-Pierre LE CADRE and Henri MAÎTRE.

The approach in principal component analysis, which projects each source of information on the eigenvectors of the covariance matrix, is often used, in order to obtain l new decorrelated sources, arranged in decreasing order of energy. Truncating the set to keep only the first l' ($l > l'$) sources can often be done while preserving most of the original set's energy.

However, in practice, this method quickly shows its limitations in image processing, for example, because it cannot take into account complex dependences between images or the spatial variations of dependences.

In order to express the information contributed by adding a new source of information I_{l+1} to an already known set $\{I_1, \ldots, I_l\}$, the preferred approach is that suggested by Shannon, which relies on the concepts of information and entropy [KUL 59, MAI 96]. Based on the joint probability of the first l sources $p(I_1, \ldots, I_l)$ (estimated most of the time by using frequencies of occurrence, for example, based on the multi-dimensional histogram of gray levels in an image), the entropy (or mean information per pixel in the case of images) of the first l sources is defined by:

$$H\big(I_1, \ldots, I_l\big) = -\sum p\big(I_1, \ldots, I_l\big) \log p\big(I_1, \ldots, I_l\big), \tag{6.1}$$

and the entropy contributed by the $(l+1)^{\text{th}}$ source is expressed, either depending on the entropies, or depending on the probabilities, as:

$$\begin{aligned} H\big(I_{l+1} \mid I_1, \ldots I_l\big) &= H\big(I_1, \ldots, I_{l+1}\big) - H\big(I_1, \ldots, I_l\big) \\ &= -\sum p\big(I_1, \ldots I_{l+1}\big) \log p\big(I_{l+1} \mid I_1, \ldots, I_l\big). \end{aligned} \tag{6.2}$$

For two sources, we thus define redundancy[1] between them as:

$$R(I_1, I_2) = H(I_1) + H(I_2) - H(I_1, I_2), \tag{6.3}$$

and the complementarity of the source I_2 with respect to I_1, i.e. the mean quantity of information that has to be added to I_2 in order to have I_1:

$$C\big(I_1 \mid I_2\big) = H\big(I_1 \mid I_2\big), \tag{6.4}$$

which leads us to the following relation:

$$H\big(I_1\big) = R\big(I_1, I_2\big) + C\big(I_1 \mid I_2\big). \tag{6.5}$$

Analogous methods could be considered in a non-probabilistic framework, by relying, for example, on fuzzy entropy [LUC 72]. For the moment, the formalism is less well developed in this direction.

1. This redundancy unfortunately cannot be generalized to more than two sources without potentially losing its positivity property.

In fusion, we usually work with highly redundant sources to confirm an uncertain decision and complementary images to broaden the range of decisions. Complementary sources can lead to either conflicting or consensual decisions.

In image processing, the concept of entropy has been extended to characterize not only how spread out the measurements are in the measurement space, but also the spatial consistency of the measurements, by taking into account occurrence probabilities of certain pixel configurations, in the context of either classification [MAI 94, MAI 96], or Markov fields [TUP 00, VOL 95].

The concepts of overall entropy are not always well suited for fusion problems and the concepts of entropy conditional to the classes to recognize are often preferable: they achieve a finer analysis of the information that each source provides for each class and are therefore better suited for problems in which a source is better for certain classes and worse for others. Although the formal definition of such concepts poses no particular difficulty, they are rarely used in fusion and would probably deserve to be further investigated.

6.3. Modeling and estimation

The most commonly used theory in other works is by far probability theory, associated with Bayesian decision theory [DUD 73]. It models information as a conditional probability, for example, the probability for a pixel to belong to a particular class, given the images available. Thus, the measurement introduced in section 1.5 can be written as follows:

$$M_i^j(x) = p\big(x \in C_i \mid I_j\big). \quad [6.6]$$

This probability is calculated based on characteristics $f_j(x)$ of the information extracted from the sources. For example, with images, they can consist of the simplest of cases of the considered pixel's gray level or of more complex information requiring preliminary processing. Equation [6.6] then no longer depends on the entire source I_j and is instead written more simply as:

$$M_i^j(x) = p\big(x \in C_i \mid f_j(x)\big). \quad [6.7]$$

In signal and image processing, in the absence of strong functional models for describing the observed phenomena, the probabilities $p(f_j(x) \mid x \in C_i)$, or more generally $p(I_j \mid x \in C_i)$ (which represents the probability, conditional to the class C_i, of the information provided by the source I_j), are learned from frequencies of occurrence on testing areas (or by learning on these areas the parameters of a given law) which gives us the probabilities in equations [6.6] and [6.7] by applying Bayes' rule.

6.4. Combination in a Bayesian framework

In the Bayesian model, fusion can be achieved in equivalent ways on two levels:

– either on the modeling level and we then calculate probabilities in the form:

$$p\big(x \in C_i \mid I_1, \dots, I_l\big), \tag{6.8}$$

using Bayes' rule:

$$p\big(x \in C_i \mid I_1, \dots, I_l\big) = \frac{p\big(I_1, \dots I_l \mid x \in C_i\big)p\big(x \in C_i\big)}{p\big(I_1, \dots, I_l\big)}, \tag{6.9}$$

where the different terms are estimated by learning;

– or from Bayes' rule itself, where the information provided by a sensor updates the information regarding x, which is estimated according to the previous sensors (this is the only usable form if the elements of information are available one after the other and not simultaneously):

$$\begin{aligned} &p\big(x \in C_i \mid I_1, \dots, I_l\big) \\ &\quad = \frac{p\big(I_1 \mid x \in C_i\big)p\big(I_2 \mid x \in C_i, I_1\big) \cdots p\big(I_l \mid x \in C_i, I_1, \dots, I_{l-1}\big)p\big(x \in C_i\big)}{p\big(I_1\big)p\big(I_2 \mid I_1\big) \cdots p\big(I_l \mid I_1, \dots, I_{l-1}\big)}. \end{aligned}$$

Very often, because of the complexity of learning using several sensors and the difficulty of gathering enough statistics, these equations are simplified under the independence hypothesis. Again, criteria have been suggested for verifying the validity of these hypotheses. The previous formulae then become:

$$p\big(x \in C_i \mid I_1, \dots, I_l\big) = \frac{\prod_{j=1}^{l} p\big(I_j \mid x \in C_i\big)p\big(x \in C_i\big)}{p\big(I_1, \dots, I_l\big)}. \tag{6.10}$$

This equation clearly shows the combination of information as a product, hence a conjunctive fusion. It is worth noting that the *a priori* probability plays exactly the same role in the combination as each of the sources with which it is also combined by a product.

6.5. Combination as an estimation problem

Another way of seeing probabilistic fusion consists of considering that each source yields a probability (of belonging to a class, for example) and that fusion consists of combining these probabilities, in order to find the overall probability of belonging to a class. This point of view amounts to considering that fusion is an estimation problem and is a way of using combination operators other than the product. In particular, the mean, weighted mean and consensus methods are often used [COO 88, COO 91, FRE 85]. Robust estimators can also be used, in order to reduce or eliminate the influence of outliers. Finally, methods provided by regionalized variable theory [MAT 70], such as kriging or universal kriging, can also be used in this context.

6.6. Decision

The last step involves the decision, for example, choosing which class a point belongs to. This binary decision can be paired up with a measurement of the quality of this decision, which can possibly lead to its rejection. The most commonly used rule in probabilistic and Bayesian decision is the *a posteriori* maximum:

$$x \in C_i \text{ if } p\big(x \in C_i \mid I_1, \ldots, I_l\big) = \max\big\{p\big(x \in C_k \mid I_1, \ldots, I_l\big),\ 1 \leq k \leq n\big\},$$

but many other criteria have been developed by probabilists and statisticians, in order for them to find the best way to adapt to the user and to the context of his decision: maximum likelihood, maximum entropy, maximum marginal probability, maximum expected value, minimum risk, etc. However, the large diversity of these criteria leaves the user hard-pressed to justify a choice and brings him further away from the objectivity initially sought by these methods.

6.7. Other methods in detection

The field of detection by multi-sensor fusion has been studied at length and has led to several methods. A distinction is made between centralized detection, in which measurements made by different sensors are considered as a vector on which the decision is made, and decentralized detection, in which each sensor yields a binary response (detection or not), and these answers are then combined by a fusion operator.

In the first case [VAR 97], the decision rules often rely on average risk, maximum profit, minimum risk, all taken from the Bayesian approach, but also on criteria such as the Neyman-Pearson criterion, which consists of maximizing the detection probability for a given probability of having a false alarm. This implicitly assumes that the false alarm is considered as the worst error, which is not always the case depending on the applications (for example, in the case of humanitarian demining, non-detection is the worst error).

In the second case, if we have l sensors, each one producing a binary response, the fusion operator is considered as a logical function of these l responses (which constitute the operator's input). The number of possible operators is very high, (2^{2^l}), making it impractical to use methods based on thoroughly counting the possibilities when the value of l goes beyond a few units. However, the number of possibilities is greatly reduced by the monotonicity constraint imposed on the operator [THO 89]. Among the more interesting methods in this field, we can mention those based on entropy for optimizing the fusion operator [DES 99]. This leads to an optimal fusion rule expressed as a weighted sum of local decisions, that can be compared to a threshold which is a function of false alarm and detection probabilities of the various local sensors, *a priori* probabilities and costs. Methods based on entropy can also be used on several levels: for choosing the most relevant sensors, for optimizing local sensors (on each sensor's level) and for optimizing the fusion operator.

6.8. An example of Bayesian fusion in satellite imagery

In this section, we will illustrate Bayesian fusion by a simple example of multi-source classification in satellite imagery, in which fusion is performed on pixels, based on the information of the gray levels. This example was discussed in [CHA 95]. Figure 6.1 shows an example of six images to fuse. These are SPOT images in the XS multi-band spectral mode in green (XS1), red (XS2) and near infrared (XS3), with a sampling increment of 20 meters, registered in a common frame of reference (to allow us to perform fusion on the pixel level).

The classes considered are cities or urban areas (class C_1), rivers (class C_2) and a class C_3 for every other structure (mostly vegetated areas).

Since the main characteristic of the cities in these images is their texture, the three initial images are completed with three texture images obtained by using an algorithm for estimating the parameters of a Gaussian Markovian field [DES 93]. These texture images are also shown in Figure 6.1.

Conditional probabilities are learned using a histogram of the gray levels. These estimates can be smoothed with Parzen windows, for example. Figure 6.2 illustrates the results of the learning process in one of the images. In practice, in order to avoid unjustified hypotheses of independence, the joint probability of the three XS channels conditionally to the classes is estimated, and likewise for the three texture images.

One of the difficulties of the Bayesian method is the *a priori* estimation of probabilities. If they are set according to the proportion of classes found in the images, this leads to a strong decrease in the probability of poorly represented classes (see Figure 6.2, on the right) thus making them very difficult to detect. The method chosen here consists of making these estimations when there is little conflict between the classes and to take uniform *a priori* probabilities when the statistical properties of the classes indicate a strong conflict.

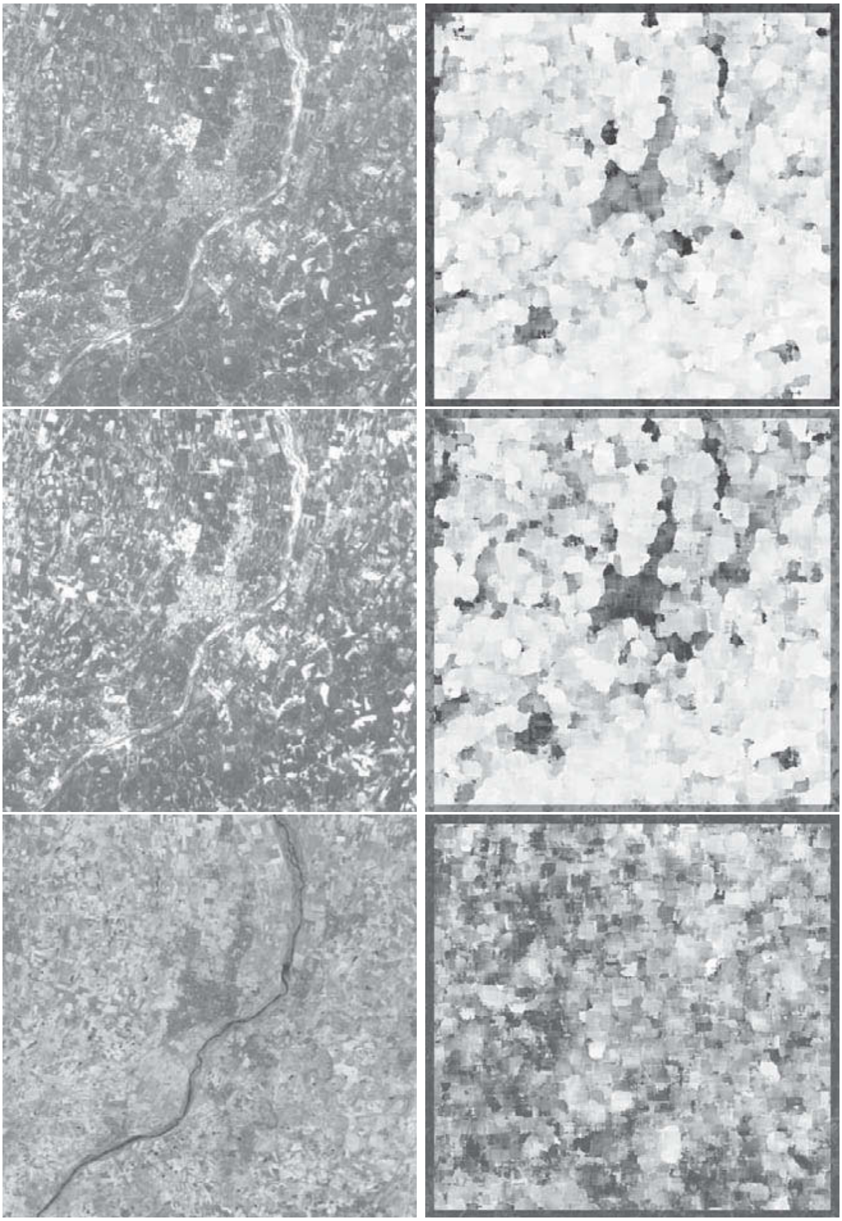


Figure 6.1. *SPOT satellite images in XS multi-band spectral mode with a size of* 512×512 *pixels and a 1-B level of pre-processing in vertical viewing (the area of Vignola, near Modena, in Italy) © Spot Image. From top to bottom, the green (XS1), red (XS2) and near infrared (XS3) channels. Left: original SPOT images. Right: images of temperature parameters of a Gaussian Markovian field on the three SPOT channels (texture images)*

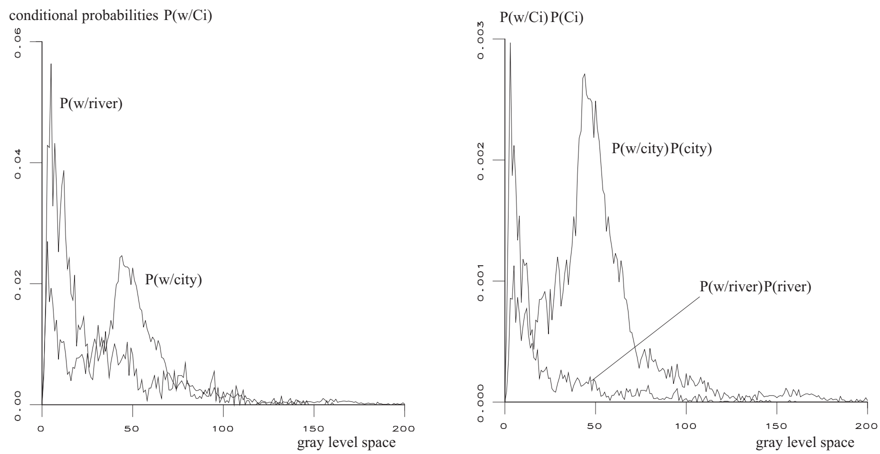


Figure 6.2. *Left: conditional probabilities of the cities and rivers in the texture image on XS2. Right: probabilities conditional to the two classes multiplied by their a priori probability* $P(city) = 11\%$ *and* $P(river) = 2\%$

Figure 6.3 shows the fusion result, for an *a priori* maximum criterion. The river is superimposed in white over the original image, as well as the contours of the urban areas. The rest corresponds to the class C_3.

6.9. Probabilistic fusion methods applied to target motion analysis

After seeing an example of Bayesian fusion in image processing, we are now going to see other probabilistic methods based on different concepts applied to a traditional problem in signal processing, i.e. that of target motion analysis. After a general presentation of target motion analysis, we will examine in detail some aspects that are more relevant to data fusion. In particular, we will focus on showing what the basic techniques of target motion analysis can contribute to a detection system.

6.9.1. *General presentation*

First of all, let us point out the goal of target motion analysis: determining, based on measurements (observations), the trajectory of a moving object. This essentially involves a framework in which a dynamic system is partially observed. A traditional and historical example is the estimation of a planet's trajectory, based on optical observations (angle measurements). Below are the most widespread types of measurements in target motion analysis:

– angles (sonar, ESM, infrared, etc.);

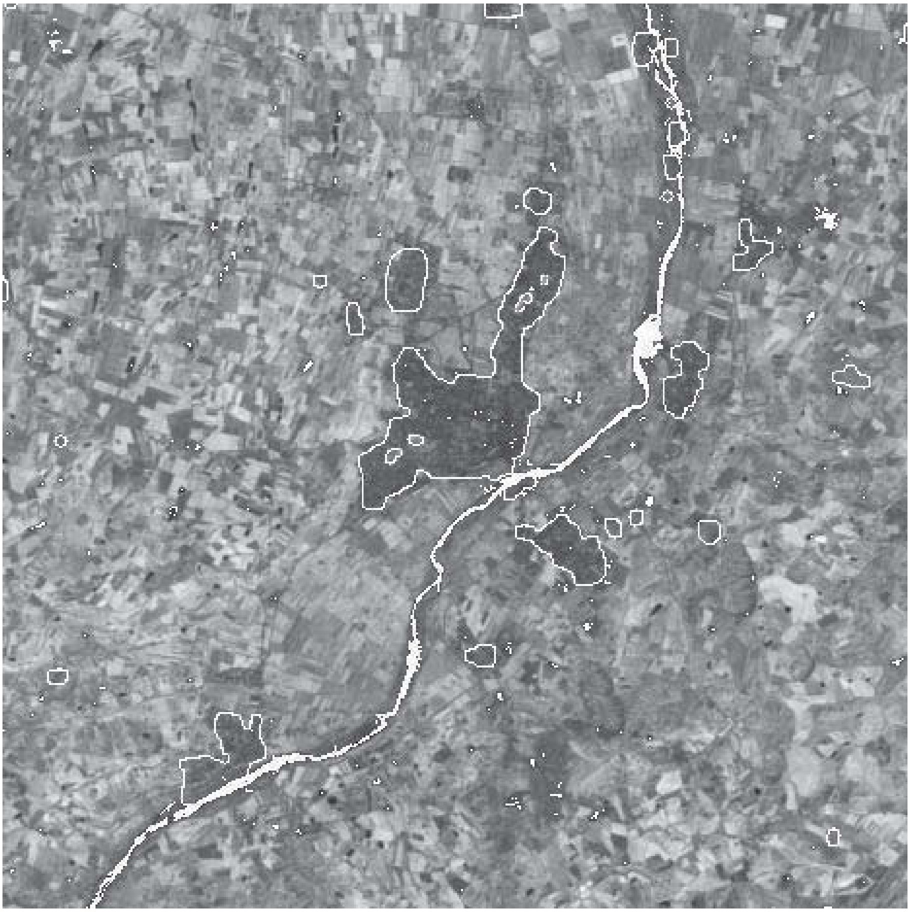


Figure 6.3. *Detection images by Bayesian inference of the classes city, river and C_3 in the scene of Vignola by fusing 6 SPOT satellite images*

– delays (radar, sonar, interferometers, etc.);

– Doppler shifts (differentials);

– intensity variations, image deformations, etc.

We note that these measurements are very diverse. One common trait, however, is that they have a non-linear dependence on parameters defining the moving object's trajectory. In many cases, we will call *system's state* the vector of parameters, instantaneous or not, defining the trajectory. These parameters can be the moving object's position, its speed at each instant or at a reference time.

Problem formulation

Thus, in a deterministic approach, the goal is to infer from the sequence of observations $\{\hat{\beta}_1, \ldots, \hat{\beta}_n\}$ the value of the state estimated at a reference time, i.e. $\hat{\mathbf{X}}_0$,

$$\{\hat{\beta}_1, \ldots, \hat{\beta}_n\} \longrightarrow \hat{\mathbf{X}}_0$$

In the case of tracking, the objective is then to estimate the sequence of states:

$$\{\hat{\beta}_1, \ldots, \hat{\beta}_n\} \longrightarrow \{\hat{\mathbf{X}}_1, \ldots, \hat{\mathbf{X}}_n\}$$

We will now briefly present the elementary model for bearings-only tracking using passive target motion analysis (also known as BOT TMA), by restricting ourselves first to a target in uniform rectilinear motion in the plane. For a general representation, see [NAR 84] and [CHA 92]. Let $\mathbf{X}$ be the state vector related to the target (T), defined by:

$$\mathbf{X} = \mathbf{X}_T - \mathbf{X}_{obs} \triangleq \left[r_x, r_y, v_x, v_y\right]^*,$$

where * indicates the transpose.

The state dynamics equation (time-discrete) then has the following form:

$$\mathbf{X}_k = \Phi(k, k-1)\mathbf{X}_{k-1} + \mathbf{U}_k, \quad [6.11]$$

where:

$$\begin{aligned} \Phi(k, k-1) &= \begin{pmatrix} Id_2 & \alpha Id_2 \\ 0 & Id_2 \end{pmatrix}, \\ Id_2 &\triangleq \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \alpha \triangleq t_k - t_{k-1}. \end{aligned} \quad [6.12]$$

In the formula above, the vector $\mathbf{U}_k$ expresses the effects of the observer's accelerations. The matrix $\Phi\,(k, k-1)$ is the system's transition matrix, also denoted by F from now on. Furthermore, we assume that $\alpha = 1$. The measurement equation is, in this case:

$$\tilde{\beta}_k = \beta_k + w_k = \tan^{-1}\left(\frac{r_{x,k}}{r_{y,k}}\right) + w_k. \quad [6.13]$$

As you can see, the target's trajectory is determined by a state vector $\mathbf{X}$, which is defined at an arbitrary reference time (i.e. $\mathbf{X}_0$). Thus, under the Gaussian hypothesis, the problem is as follows; let $\tilde{\mathbf{B}}$ be the measurement history (the angles), we must now consider the following likelihood functional [NAR 84]:

$$p\big(\tilde{\mathbf{B}} \mid \widehat{\mathbf{X}}\big) = \mathrm{cst}\exp\Big[-\frac{1}{2}\big\|\tilde{\mathbf{B}} - \mathbf{B}\big(\widehat{\mathbf{X}}\big)\big\|^2_{\Sigma^{-1}}\Big],$$

with w_k as a sequence of independent, identically distributed, Gaussian white noise and:

$$\begin{aligned}\tilde{\mathbf{B}} &= \big(\tilde{\beta}_1, \dots, \tilde{\beta}_p\big)^*, \\ \tilde{\beta}_k &= \beta_k + w_k,\ w : \mathcal{N}(0, \Sigma).\end{aligned} \qquad [6.14]$$

Clearly, the likelihood functional has a non-linear dependence on the state $\mathbf{X}$ and there is no general explicit solution to this optimization problem. This is why we often resort to a Gauss-Newton type algorithm [NAR 84], written as follows:

$$\widehat{\mathbf{X}}_{\ell+1} = \widehat{\mathbf{X}}_\ell - \rho_\ell \left[\left(\frac{\partial \widehat{\mathbf{B}}}{\partial \mathbf{X}}\right)^* \Sigma^{-1} \frac{\partial \widehat{\mathbf{B}}}{\partial \mathbf{X}}\right]^{-1} \left(\frac{\partial \widehat{\mathbf{B}}}{\partial \mathbf{X}}\right)^* \Sigma^{-1}\big(\tilde{\mathbf{B}} - \widehat{\mathbf{B}}\big), \qquad [6.15]$$

where $\Sigma = \mathrm{Diag}(\sigma_i^2)$, σ_i^2 is the noise variance of the i^{th} measurement, ℓ is the iteration index, ρ_ℓ the stepsize and, using notations that are not completely accurate, $\widehat{\mathbf{B}} = \mathbf{B}(\widehat{\mathbf{X}}_\ell)$. A simple calculation [NAR 84] helps us calculate the matrix $\partial \widehat{\mathbf{B}}/\partial \mathbf{X}$, i.e.:

$$\frac{\partial \widehat{\mathbf{B}}}{\partial \mathbf{X}} = \begin{pmatrix} \dfrac{\cos\big(\beta_1\big)}{r_1} & -\dfrac{\sin\big(\beta_1\big)}{r_1} & \big(t_1 - t_0\big)\dfrac{\cos\big(\beta_1\big)}{r_1} & -\big(t_1 - t_0\big)\dfrac{\sin\big(\beta_1\big)}{r_1} \\ \vdots & \vdots & \vdots & \vdots \\ \dfrac{\cos\big(\beta_p\big)}{r_p} & -\dfrac{\sin\big(\beta_p\big)}{r_p} & \big(t_p - t_0\big)\dfrac{\cos\big(\beta_p\big)}{r_p} & -\big(t_p - t_0\big)\dfrac{\sin\big(\beta_p\big)}{r_p} \end{pmatrix} \qquad [6.16]$$

where r_i refers to the relative target-receiver distance at the time i.

The phrase *comprehensive target motion analysis methods* is generally used to refer to this type of approach [NAR 84]. They are often presented in competition with Kalman-type methods, even though the objectives are fundamentally different. The methods are simple to implement and have reasonable calculation costs. A few iterations of the algorithm are usually enough. A difficult point can be the choice of the increment ρ. See [BAZ 93] for a state of the art in knowledge on efficient methods for choosing these parameters.

Still in the context of comprehensive methods, another type stands out: *instrumental variable* methods (IVM). The idea is simple: to use the successive estimations of the system's state to infer the instrumental variables by an iterative algorithm. The pseudo-measurements z_k are defined by:

$$z_k = r_{o,x}(k)\cos\hat{\beta}_k - r_{o,y}(k)\sin\hat{\beta}_k \qquad [6.17]$$

This equation can also be written in matrix form:

$$z_k = \tilde{\mathbf{A}}^*(k)\mathbf{X}_0(k) = \tilde{\mathbf{A}}^*(k)\mathbf{X}_T(k) + \eta_k,$$

where $r_{o,x}, r_{o,y}$ represents the observer's co-ordinates and:

$$\left|\begin{array}{l}\tilde{\mathbf{A}}^*(k) = \left(\cos\hat{\beta}_k, -\sin\hat{\beta}_k, 0, 0\right),\\ \eta_k = r_k \sin w_k \simeq r_k w_k.\end{array}\right. \qquad [6.18]$$

We are then led to consider a linear regression problem for which there is an explicit solution. However, the solution obtained this way is usually strongly biased because of the correlation between the columns of the regression matrix and the additive noise. The IVM consists of replacing the usual optimality equation for minimizing the quadratic norm of the error with the following iterative expression:

$$\widehat{\mathbf{X}}_{\ell+1} = \left(\mathcal{A}_p^*\big(\widehat{\mathbf{X}}_\ell\big)\widehat{R}^{-2}\tilde{\mathcal{A}}_p\right)^{-1}\left(\mathcal{A}_p^*\big(\widehat{\mathbf{X}}_\ell\big)\widehat{R}^{-2}\mathbf{Z}_p\right). \qquad [6.19]$$

This is an even simpler method for implementing the Gauss-Newton algorithm. The convergence of these two methods has been debated at length; there are, however, methods that lead to results with a certain degree of generality. More precisely, if we consider the quadratic functional of $\widehat{\mathbf{X}}$:

$$L\big(\widehat{\mathbf{X}}\big) = \left\|\widehat{\mathbf{X}} - \mathbf{X}\right\|^2, \qquad [6.20]$$

Iltis and Anderson [ILT 96] show that this is a Lyapunov functional [BAR 85] for the continuous differential equation:

$$\frac{d}{dt}\widehat{\mathbf{X}} = \mathbf{G}\big(\widehat{\mathbf{X}}\big) \qquad [6.21]$$

where $\mathbf{G}(\widehat{\mathbf{X}})$ is the gradient vector of the likelihood in $\widehat{\mathbf{X}}$. After some simple calculations, we get:

$$\mathbf{G}\big(\widehat{\mathbf{X}}\big) = H^*\big(\widehat{\mathbf{X}}\big)\begin{pmatrix}\beta_1 - \widehat{\beta}_1\\ \vdots \\ \beta_p - \widehat{\beta}_p\end{pmatrix}$$

where $\beta_i \triangleq \beta_i(\mathbf{X}), \widehat{\beta}_i \triangleq \beta_i(\widehat{\mathbf{X}})$ and:

$$H\big(\widehat{\mathbf{X}}\big) = \begin{pmatrix} \dfrac{\cos\widehat{\beta}_1}{\widehat{r}_1} & -\dfrac{\sin\widehat{\beta}_1}{\widehat{r}_1} & \dfrac{\cos\widehat{\beta}_1}{\widehat{r}_1} & -\dfrac{\sin\widehat{\beta}_1}{\widehat{r}_1} \\ \vdots & \vdots & \vdots & \vdots \\ \dfrac{\cos\widehat{\beta}_p}{\widehat{r}_p} & -\dfrac{\sin\widehat{\beta}_p}{\widehat{r}_p} & \dfrac{p\cos\widehat{\beta}_p}{\widehat{r}_p} & -\dfrac{p\sin\widehat{\beta}_p}{\widehat{r}_p} \end{pmatrix} \qquad [6.22]$$

$$\widehat{r}_i = \left(r_{x,i}^2\big(\widehat{\mathbf{X}}\big) + r_{y,i}^2\big(\widehat{\mathbf{X}}\big)\right)^{1/2},\ H^*\big(\widehat{\mathbf{X}}\big) : \text{a } 4\times p \text{ size matrix.}$$

We then have to calculate the line matrix $\big(\widehat{\mathbf{X}} - \mathbf{X}\big)^* H^*\big(\widehat{\mathbf{X}}\big)$ whose k^{th} element, denoted by I_k, has the following form:

$$I_k = \frac{1}{\widehat{r}_k^2}\widehat{\mathbf{W}}_k^*\big(\widehat{\mathbf{X}} - \mathbf{X}\big), \qquad [6.23]$$

where $\widehat{\mathbf{W}}_k \triangleq \left(\widehat{r}_{y,k}, -\widehat{r}_{x,k}, k\,\widehat{r}_{y,k}, -k\,\widehat{r}_{x,k}\right)^*$.

We can then easily prove the following results [LEC 99]:

$$\begin{cases} \widehat{\mathbf{W}}_k^*\widehat{\mathbf{X}} = 0,\ k = 1,\ldots,p, \\ \dfrac{1}{\widehat{r}_k^2}\widehat{\mathbf{W}}_k^*\mathbf{X} = \dfrac{r_k}{\widehat{r}_k}\sin\left(\widehat{\beta}_k - \beta_k\right). \end{cases} \qquad [6.24]$$

We will give a few elements of proof of this property and it will then be easy to imagine extensions. To do this, first, the vectors $\mathbf{X}$ and $\widehat{\mathbf{X}}$ are partitioned in sub-vectors with positions ($\mathbf{R}$) and speeds ($\mathbf{V}$):

$$\mathbf{X} \triangleq \begin{pmatrix} \mathbf{R} \\ \mathbf{V} \end{pmatrix}, \qquad \widehat{\mathbf{X}} \triangleq \begin{pmatrix} \widehat{\mathbf{R}} \\ \widehat{\mathbf{V}} \end{pmatrix},$$

we then have:

$$\begin{aligned} \widehat{\mathbf{W}}_k^*\mathbf{X} &= \left(\widehat{\mathbf{R}}^*, \widehat{\mathbf{V}}^*\right)\begin{pmatrix} J & kJ \\ kJ & k^2J \end{pmatrix}\begin{pmatrix} \mathbf{R} \\ \mathbf{V} \end{pmatrix}, \\ &= \left(\widehat{\mathbf{R}} + k\widehat{\mathbf{V}}\right)^* J(\mathbf{R} + k\mathbf{V}), \end{aligned} \qquad [6.25]$$

where $J = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$.

This calls for the following important comment. If $\mathbf{U}$ and $\mathbf{U}'$ are two vectors $\mathbb{R}^2$, then:

$$\begin{aligned}\mathbf{U}^* J\mathbf{U}' &= u_x u'_y - u'_x u_y, \\ &= \det\left(\mathbf{U}, \mathbf{U}'\right),\end{aligned} \qquad [6.26]$$

and as a result:

$$\widehat{\mathbf{W}}_k^* \mathbf{X} = \det\left(\widehat{\mathbf{R}} + k\widehat{\mathbf{V}}, \mathbf{R} + k\mathbf{V}\right), \qquad [6.27]$$

$$= \frac{r_k}{\widehat{r}_k} \frac{\det\left(\widehat{\mathbf{R}}_k, \mathbf{R}_k\right)}{\left\|\mathbf{R}_k\right\| \left\|\widehat{\mathbf{R}}_k\right\|},$$

$$= \frac{r_k}{\widehat{r}_k} \sin\left(\widehat{\mathbf{R}}_k, \mathbf{R}_k\right) \qquad [6.28]$$

where: $\mathbf{R}_k \triangleq \mathbf{R} + k\mathbf{V}$, $\widehat{\mathbf{R}}_k \triangleq \widehat{\mathbf{R}} + k\widehat{\mathbf{V}}$.

We now simply have to point out that $\sin(\widehat{\mathbf{R}}_k, \mathbf{R}_k) = \sin(\beta_k - \widehat{\beta}_k)$.

This calculation can be applied, not without some difficulties, to the case of a target and an observer that are maneuvering. We then have the following convergence result [LEC 99].

Convergence of the iterative methods

Let us assume that the maneuvering times of the target are known, then the derivative with respect to time $\dot{L}(\hat{\mathbf{X}})$ of the Liapunov function $L(\hat{\mathbf{X}})$ is:

$$\dot{L}\big(\hat{\mathbf{X}}\big) = 2\big(\hat{\mathbf{X}} - \mathbf{X}\big)^* \mathbf{G}_L\big(\hat{\mathbf{X}}\big) = -2\sum_{k=1}^{p} \frac{r_k}{\hat{r}_k}\big(\hat{\beta}_k - \beta_k\big)\sin\big(\hat{\beta}_k - \beta_k\big). \qquad [6.29]$$

As you can see, there is a functional of the gradient vector $\mathbf{G}_L(\hat{\mathbf{X}})$ of the likelihood functional that cannot be equal to zero if all of the $\hat{\beta}_k$ and β_k coincide, i.e. if the model is perfectly estimated. This analysis can be generalized without difficulty to the case of multiple sensors and most importantly to multi-leg cases[2].

This type of calculation can actually easily be applied to the case of a constantly accelerating target and, more generally, to an observation model of the type $\tan(\beta_t) = \mathbf{A}_t^* \mathbf{X} / \mathbf{B}_t^* \mathbf{X}$ and even $f(\beta_t) = \mathbf{A}_t^* \mathbf{X} / \mathbf{B}_t^* \mathbf{X}$ where f is a monotonic and continuously differentiable function. Thus, the non-linearity has a relatively simple structure. This is

2. A leg is a section of the trajectory over which the target has a uniform rectilinear motion.

of great importance whether for the analysis of observability [TRE 96], or of the convergence of the extended Kalman filter [SON 85]. As a result, the analysis of observability becomes much easier if we rewrite equations [6.11] and [6.13] as follows:

$$\left|\begin{array}{l}\mathbf{X}_{k+1} = F\mathbf{X}_k + U \\ 0 \equiv z_k = H_k\mathbf{X}_k\end{array}\right.$$

where:

$$\begin{aligned} &F = \Phi\big(t_{k+1}, t_k\big) = \begin{pmatrix} Id & \alpha Id \\ 0 & Id \end{pmatrix} \big(\alpha \triangleq t_{k+1} - t_k\big) \\ &\mathbf{X}_k \triangleq \mathbf{X}_{t_k} \\ &H_k = \big(\cos\theta_k, -\sin\theta_k, 0, 0\big) \end{aligned} \qquad [6.30]$$

If we now assume that the observer is not maneuvering, $U \equiv 0$. We then have:

$$\left|\begin{array}{l} z_0 = H_0\mathbf{X}_0 \\ z_1 = H_1 F\mathbf{X}_0 \\ \quad\vdots \\ z_k = H_k F^k \mathbf{X}_0 \end{array}\right. \qquad [6.31]$$

with:

$$F^k = \begin{pmatrix} ccId & k\alpha Id \\ 0 & Id \end{pmatrix}$$

so that the observability matrix $\mathcal{O}$ can be written:

$$\mathcal{O} = \begin{pmatrix} cH_0 \\ H_1 F \\ \vdots \\ H_k F^k \end{pmatrix} = \begin{pmatrix} \cos\theta_0 & -\sin\theta_0 & 0 & 0 \\ \cos\theta_1 & -\sin\theta_1 & \alpha\cos\theta_1 & -\alpha\sin\theta_1 \\ \vdots & & & \\ \cos\theta_k & -\sin\theta_k & k\alpha\cos\theta_k & -k\alpha\sin\theta_k \end{pmatrix} \qquad [6.32]$$

Factoring the observability matrix as follows is quite helpful:

$$\mathcal{O} = \Delta_\theta \Delta_r \begin{pmatrix} r_y(0) & -r_x(0) & 0 & 0 \\ r_y(1) & -r_x(1) & \alpha r_y(1) & -\alpha r_x(1) \\ \vdots & & & \\ r_y(k) & -r_x(k) & k\alpha r_y(k) & -k\alpha r_x(k) \end{pmatrix}$$

with:

$$\begin{aligned}\Delta_\theta &\triangleq \operatorname{diag}\big(\cos\theta_0, \cos\theta_1, \ldots, \cos\theta_k\big)\\ \Delta_r &\triangleq \operatorname{diag}\big(r_y^{-1}(0), r_y^{-1}(1), \ldots, r_y^{-1}(k)\big)\end{aligned} \qquad [6.33]$$

This factorization is valid so long as each term $r(k)$ is different from zero, which is hardly restrictive. Let $\mathcal{O}'$ be the factored observability matrix:

$$\mathcal{O}' = \begin{pmatrix} r_x(0) & r_y(0) & 0 & 0 \\ \vdots & \vdots & \alpha r_x(1) & \alpha r_y(1) \\ \underbrace{r_x(k)}_{\mathbf{T}_x} & \underbrace{r_y(k)}_{\mathbf{T}_y} & \underbrace{k\alpha r_x(k)}_{\mathbf{V}_x} & \underbrace{k\alpha r_y(k)}_{\mathbf{V}_y} \end{pmatrix}$$

We then have: $\operatorname{rank}\mathcal{O} = \operatorname{rank}\mathcal{O}'$.

We easily infer from the previous calculations that the vectors $\mathbf{T}_x, \mathbf{T}_y, \mathbf{V}_x$ and $\mathbf{V}_y$ are linear combinations of the three vectors $\mathbf{1}, \mathbf{Z}$ and $\mathbf{Z}^2$, and we have:

$$\begin{cases} \mathbf{T}_x = r_x(0)\mathbf{1} + \alpha v_x \mathbf{Z} \\ \mathbf{V}_x = \alpha r_x(1)\mathbf{Z} + \alpha^2 v_x \mathbf{Z}^2 \\ \mathbf{T}_y = r_y(0)\mathbf{1} + \alpha v_y \mathbf{Z} \\ \mathbf{V}_y = \alpha r_y(1)\mathbf{Z} + \alpha^2 v_y \mathbf{Z}^2 \end{cases}$$

with:

$$\begin{aligned}\mathbf{1} &\triangleq (1, 1, \ldots, 1)^1\\ \mathbf{Z} &\triangleq (0, 1, 2, \ldots, k)^*\\ \mathbf{Z}^2 &\triangleq (0, 0, 2, \ldots, k(k-1))^*\end{aligned}$$

It is then obvious that the rank of the matrix $(\mathcal{O}')$ and therefore of the matrix $\mathcal{O}$ is equal to 3, except if we have $r_x(0)v_y = r_y(0)v_x$, in which case the rank of $\mathcal{O}$ is only 2. This analysis might seem simplistic, however, it makes it possible to handle most problems. Let us consider, for example, the case of a system comprised of two linear sub-antennae located on a same line. Here, the previous calculations lead to:

$$\mathcal{O}' = \left(\frac{\mathcal{O}'_1}{\mathcal{O}'_2}\right), \qquad [6.34]$$

$$= \left(\frac{\mathcal{O}'_1}{\Delta\mathcal{O}'_1}\right) + \beta \begin{pmatrix} \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \Delta\mathbf{1} & \mathbf{O} & \alpha\Delta\mathbf{Z} & \mathbf{O} \end{pmatrix}, \qquad [6.35]$$

where:

$$\begin{cases} \mathcal{O}'_1 = (\mathbf{T}_x, \mathbf{T}_y, \mathbf{V}_x, \mathbf{V}_y) \text{ and: } \mathcal{O}'_2 = \Delta\left(\mathbf{T}'_x, \mathbf{T}'_y, \mathbf{V}'_x, \mathbf{V}'_y\right), \\ \Delta = \operatorname{diag}\left(\dfrac{\cos\theta'_0}{\cos\theta_0}, \ldots, \dfrac{\cos\theta'_k}{\cos\theta_k}\right) = \operatorname{diag}\left(\dfrac{r'_0}{r_0}, \ldots, \dfrac{r'_k}{r_k}\right), \\ \beta \triangleq r'_x(0) - r_x(0). \end{cases} \qquad [6.36]$$

Let us now examine the properties of the matrix $\mathcal{O}'$. Let $\ker \mathcal{O}'_1$ denote the kernel of $\mathcal{O}'_1$ and R its supplementary subspace in $\mathbb{R}^4$, i.e. $\mathbb{R}^4 = \ker \mathcal{O}'_1 \oplus R$ where $\oplus$ represents the direct sum. Then, let $\mathbf{X}$ be any vector in $\mathbb{R}^4$, $\mathbf{X}$ can be decomposed in a unique way as a sum of two vectors of $\ker \mathcal{O}'_1$ and R, meaning that $\mathbf{X} = \mathbf{K} + \mathbf{Y}$ ($\mathbf{K} \in \ker \mathcal{O}'_1$ and $\mathbf{Y} \in R$) and we have the following implications:

– if $\mathbf{Y} \neq \mathbf{0}$, then $\mathcal{O}'\mathbf{X} \neq \mathbf{0}$,

– if $\mathbf{Y} = \mathbf{0}$, then $\mathcal{O}'\mathbf{X} = \beta\Delta\left(x_1\mathbf{1} + \alpha x_3 \mathbf{Z}\right)$.

Therefore, we only have to examine the second hypothesis (namely $\mathbf{Y} = 0$) and we then have:

$$\mathcal{O}'\mathbf{X} = 0 \Longrightarrow x_1 = x_3 = 0 \quad (\mathbf{1} \text{ and } \mathbf{Z} \text{ are linearly independent}),$$

and therefore, since $\mathbf{X} \in \ker \mathcal{O}'_1$, we also have $x_2\mathbf{T}_y + x_4\mathbf{V}_y = 0$ and therefore in the end, $x_2 = x_4 = 0$. As a result, except for the specific case where $\mathbf{T}_y = \mathbf{V}_y = \mathbf{O}$, $\ker \mathcal{O}'$ is simply the zero vector in the multi-platform case. Even if this is a purely algebraic result, we begin to perceive the advantage of fusing the outputs of the platforms.

As for the filtering aspect, particle filtering methods can be used to avoid any linearization. Briefly, its general form is as follows:

– initialization:

$$s_o^n \sim p(\mathbf{X}_0), \quad q_0^n = 1/N; \quad n = 1, \ldots, N;$$

– for $t = 1, \ldots, T$:
  - prediction: $\tilde{s}_t^n \sim f(\mathbf{X}_t | \mathbf{X}_{t-1} = s_{t-1}^n, \hat{\beta}_t)$; $n = 1, \ldots, N$,
  - calculation of the weights:

$$\tilde{q}_t^n = q_{t-1}^n \frac{p\left(\tilde{s}_t^n \mid s_{t-1}^n\right) l_t\left(\hat{\beta}_t; \tilde{s}_t^n\right)}{f\left(\tilde{s}_t^n \mid s_{t-1}^n, \hat{\beta}_t\right)}$$

for $n = 1, \ldots, N$, (normalization step)
  - $\mathbf{E}(\hat{X}_t) = \sum_{n=1}^N q_t^n \tilde{s}_t^n$,

- resampling (the decision to resample is ensured by a test):

$$\begin{cases} s_t^n \sim \displaystyle\sum_{k=1}^{N} q_t^k \delta_{\tilde{s}_t^k}, \\ q_t^n = 1/N, \quad n = 1, \ldots, N. \end{cases} \tag{6.37}$$

Different forms of this filter and application examples in target motion analysis can be found in [HUE 02]. Finally, still on the subject of filtering, models for maneuvering targets hold great importance. Among them, we can mention a standard one, i.e. the Singer model (with correlated noise):

$$\begin{cases} \dot{x}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} a(t), \\ \dot{a}(t) = -\alpha a(t) + w(t) \text{ where } r(\tau) \triangleq \operatorname{cov}\left(w_t, w_{t-\tau}\right) = \sigma^2 e^{-\alpha\tau}. \end{cases} \tag{6.38}$$

More generally, these models can be divided into the following categories:

– maneuvering models decoupled in co-ordinates:

- white noise models for which the command is a white noise,

- Markov models for which the input is a Markov process (includes the Singer model),

- the Semi-Markov Jump Process.

– motion models: 2-D, for example, with a constant gyration rate, 3-D, ballistic;

– measurement models: cartesian, linearized, pseudo-measurements, modified polar, curvilinear.

For the three categories above, we can give the following examples:

– Wiener acceleration model (discrete time):

$$F = \begin{bmatrix} 1 & T & T^2/2 \\ 0 & 0 & T \\ 0 & 0 & 1 \end{bmatrix}, \qquad Q = \begin{bmatrix} T^5/20 & T^4/8 & T^3/6 \\ T^4/8 & T^3/3 & T^2/2 \\ T^3/6 & T^2/2 & T \end{bmatrix}; \tag{6.39}$$

– ARMA acceleration model:

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \beta_1 & \beta_2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\alpha_2 & -\alpha_1 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}; \tag{6.40}$$

– constant gyration rate model (ω is known):

$$\dot{x}(t) = A(\omega)x + w(t), \text{ with: } A(\omega) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -\omega \\ 0 & 0 & 0 & 1 \\ 0 & \omega & 0 & 0 \end{bmatrix}. \qquad [6.41]$$

Finally, in the context of image analysis, the problem of target motion analysis often leads to considering more complex models such as:

– contour models;

– more complex kinematic equations (rigid body, for example);

– object dilation factors;

– more subtle measurement models (luminance, for example).

We are now going to examine some of the implications of data fusion in the field of target motion analysis.

6.9.2. *Multi-platform target motion analysis*

As we have just seen, there are many advantages to considering target motion analysis problems from the *multi-platform* perspective. We will see, in particular, that the performances can be decently estimated. However, in what follows, we will always assume that there is only one target. In other words, multi-target and multi-sensor association problems will not be considered at all in this section. The following results have been obtained [TRE 96]:

$$\mathrm{var}(\hat{r}_x) = \frac{3\,\sigma^2 r^4 \tan^2(\theta)}{(2n+1)(m(2m+1)(m+1)d^2},$$

$$\mathrm{var}(\hat{r}_y) = \frac{3\,\sigma^2 r^4}{(2n+1)(m(2m+1)(m+1)d^2},$$

$$\mathrm{var}(\hat{v}_x) = \frac{45\,\sigma^2 r^4 \sin^2(\theta)}{n(n+1)(2n+1)(2m+1)[5m(m+1)d^2\cos^2(\theta)+(2n-1)(2n+3)v^2\sin^2(\theta-\gamma)]},$$

$$\mathrm{var}(\hat{v}_y) = \frac{45\,\sigma^2 r^4 \cos^2(\theta)}{n(n+1)(2n+1)(2m+1)[5m(m+1)d^2\cos^2(\theta)+(2n-1)(2n+3)v^2\sin^2(\theta-\gamma)]}.$$

The parameters in these equations have the following meaning:

$$\begin{cases} \theta\text{: target's azimuth,} \\ \gamma\text{: target's bearing, with the North as reference,} \\ v\text{: modulus of the target's speed, } r\text{: distance,} \\ m\text{: number of platforms, } n\text{: integration time,} \\ d\text{: inter-platform distance, } \sigma^2\text{: estimated noise variance.} \end{cases} \qquad [6.42]$$

We can now see the influence of the various parameters. It is, however, preferable to give a more physical interpretation of the results given in equation [6.42]. We thus define the quantities $\mathcal{A}_{\text{bas}}$ and $\mathcal{T}_{\text{bas}}$ as follows[3]:

$$\mathcal{A}_{\text{bas}} = 2\, md \big|\cos(\theta)\big|,$$
$$\mathcal{T}_{\text{bas}} = (2n+1)v\big|\sin(\theta-\gamma)\big|,$$

we make the following approximations:

$$m(m+1)d^2\cos^2\theta \simeq \frac{1}{4}{\mathcal{A}_{\text{bas}}}^2$$
$$(2n-1)(2n+3)v^2\sin^2(\theta-\gamma) \simeq {\mathcal{T}_{\text{bas}}}^2,$$

and in the end we get:

$$\text{var}(\hat{r}_x) = \frac{12\,\sigma^2\,r^4\sin^2(\theta)}{(2n+1)(2m+1){\mathcal{A}_{\text{bas}}}^2},$$
$$\text{var}(\hat{r}_y) = \frac{12\,\sigma^2\,r^4\cos^2(\theta)}{(2n+1)(2m+1){\mathcal{T}_{\text{bas}}}^2},$$
$$\text{var}(\hat{v}_x) = \frac{180\,\sigma^2\,r^4\sin^2(\theta)}{n(n+1)(2n+1)(2m+1)\left[5{\mathcal{A}_{\text{bas}}}^2+4{\mathcal{T}_{\text{bas}}}^2\right]},$$
$$\text{var}(\hat{v}_y) = \frac{180\,\sigma^2\,r^4\cos^2(\theta)}{n(n+1)(2n+1)(2m+1)\left[5{\mathcal{A}_{\text{bas}}}^2+4{\mathcal{T}_{\text{bas}}}^2\right]}.$$

With these approximations, we have a good idea of the influence the various parameters have on multi-platform target motion analysis systems. See [LEC 99] for a more comprehensive presentation.

6.9.3. *Target motion analysis by fusion of active and passive measurements*

Again, we will restrict ourselves to the case of a single target, in uniform rectilinear motion.

We have at our disposal passive measurements at every instant $\beta(t) = \tan^{-1}[r_x(t)/r_y(t)]$ and at the transmission times, for example, for the radar or the active sonar, the measurements $r(t) = [r_x^2(t) + r_y^2(t)]^{1/2}$. Of course, these measurements are affected by noise that we will assume to be an independent,

3. $\mathcal{A}_{\text{bas}}$ for *array baseline* and $\mathcal{T}_{\text{bas}}$ for *target baseline*.

identically distributed sequence. Simple calculations lead to the expression of the gradient vectors $\mathbf{M}_k$ ($\mathbf{N}_k$, respectively) of β_k (r_k, respectively) with respect to $\mathbf{X}$:

$$\begin{cases} \mathbf{M}_k = \dfrac{1}{r_k}\,(\cos\beta_k, -\sin\beta_k, k\,\cos\beta_k, -k\,\sin\beta_k)^*, \\[2ex] \mathbf{N}_k = (\sin\beta_k, \cos\beta_k, k\,\sin\beta_k, k\,\cos\beta_k)^*. \end{cases} \qquad [6.43]$$

Assuming that the estimates for the bearing and the distance are independent, the calculation of the Fisher matrix related to the estimate of the state $\mathbf{X}$ is a routine exercise that leads us to:

$$\text{FIM} = \sum_k \left(\frac{1}{\sigma^2_{\beta,k}} \mathbf{M}_k \mathbf{M}_k^* + \frac{\delta_{r,k}}{\sigma^2_{r,k}} \mathbf{N}_k \mathbf{N}_k^* \right). \qquad [6.44]$$

In equation [6.44], $\delta_{r,k}$ is equal to 1 when an active measurement is available and otherwise to 0. We have to consider any number of active measurements, but it can be shown [LEC 00] that we can simply consider the cases: one active measurement out of T, two out of T and finally three out of T. We then have the following results [LEC 00].

PROPOSITION 6.1. *Let* $\det(\text{FIM}_{\tau,T,\beta,r})$ *be the determinant of the matrix* FIM *associated with two active measurements (separated by* τ*) and* T *passive measurements, then:*

$$\det\big(\text{FIM}_{\tau,T}\big) \simeq \frac{\tau^2}{\big(r\sigma_r\sigma_\beta\big)^4} \sum_{0\le t<t'\le T} (t-t')^2 \big[1 + \dot{\beta}^2\big(-tt' + \tau(t+t')\big)\big]^2. \quad [6.45]$$

PROPOSITION 6.2. *We now consider that we have three active measurements (at* 0, τ_2 *and* τ_3*) and* T *passive measurements, then:*

$$\det\big(\text{FIM}_{\tau_2,\tau_3,T}\big) \simeq \frac{T}{r^2\sigma_\beta^2\sigma_r^6} \sum_{0<\tau_2<\tau_3\le T} \big[\tau_2\tau_3\big(\tau_2-\tau_3\big)\dot{\beta}\big]^2. \qquad [6.46]$$

We get the same type of result as in the case of an active measurement [LEC 00]. In fact, we can easily see that the predominant contributions are from the terms of the type *2 active measurements among 4* and we have the following result:

$$\det(\text{FIM}) \propto \sum_{0\le t<t'\le T} (t-t')^2 \big[1 + \dot{\beta}^2\big(-tt' + \tau(t+t')\big)\big]^2. \qquad [6.47]$$

In the case of a maneuvering target (two legs and a maneuvering time t_m that is known), the state vector becomes 6-dimensional and the calculations become noticeably more complicated [LEC 00]. We then obtain:

$$\begin{aligned}\det(\text{FIM}) \approx c \sum_{t_1,\ldots,t_6} & \big[(t_1 - t_2)(t_3 - t_4)(t_5 - t_6)\big]^2 \\ \times \Big[& 2\,t_2 \cos\big(2\beta_0 - 2\big(t_2 - 2t_m + t_2 \sin 2\beta_0\big), \\ & - 2\,t_2\dot{\beta}\Big(t_2 - t_1 + \big(t_1 + t_2\big)\big(\cos 2\beta_0 + \sin 2\beta_0\big), \\ & + \dot{\beta}^2\Big(\big(t_1 - t_2\big)^2\big(t_2 - 2\,t_m\big) - t_2\big(t_1 + t_2\big)^2\big(\cos 2\beta_0 - \sin 2\beta_0\big)\Big), \\ & - \big(t_3 + t_4 - t_5 - t_6\big)^2\big(-t_2 + 2\,t_m + t_2\big(\cos 2\beta_0 - \sin 2\beta_0\big)\big)\Big)\Big]^2.\end{aligned}$$

This formula might seem disconcertingly complex. We notice, however, that the predominant term is simply:

$$2\,t_2 \cos\big(2\beta_0 - 2\big(t_2 - 2t_m + t_2 \sin 2\beta_0\big)\big).$$

Again, we reach the conclusion that the optimal strategy consists of focusing the active measurements on the extremities of the legs. Therefore, as you can see, simple considerations make it possible for us to grasp the fusion of active and passive measurements in target motion analysis.

6.9.4. *Detection of a moving target in a network of sensors*

The problem here is simple. How can we detect a moving target traveling through a network of sensors? In order to better outline the problem, a few specifics are necessary. We will assume in this case that the detection radius of each sensor is relatively limited, so as to make it reasonably impossible for it to detect more than one target. At a given time, a target is only perceivable by a few sensors of the network that are close enough.

In practice, this means that antenna processing methods are of little use. There is no concept of antenna gain and we therefore have to rely first and foremost on time discrimination. As a result, the target motion analysis stage is the core of single sensor detection. We will not discuss any further the method used, or performance calculation in detection. For this, see [DON 00, DON 02].

In general, it is actually impossible to achieve detection on the sensor's level at the output of partial target motion analysis. In practice, this means that the sensor's signal is as follows:

$$\mathbf{z}_t = \mathbf{e}(\mathbf{Y}, t) + \mathbf{n}_t \quad [6.48]$$

where $\mathbf{Y} = [\alpha \triangleq cpa/v, t_{cpa}, \theta]^t$.

This means that, based on the data provided by a single sensor, it is only possible to estimate 3 parameters out of 4. Centralized processing then consists of considering the observations provided by the sensors and to process them together. If we denote by $\mathbf{Z}$ ($\mathbf{Z} = (\mathbf{Z}_1^t, \mathbf{Z}_2^t)^t$) the vector of concatenated measurements, we have:

$$\mathbf{Z} = \mathbf{E}(\mathbf{X}) + \mathbf{N} = \begin{pmatrix} \mathbf{E}\big(\mathbf{Y}_1(\mathbf{X})\big) \\ \mathbf{E}\big(\mathbf{Y}_2(\mathbf{X})\big) \end{pmatrix} + \begin{pmatrix} \mathbf{N}_1 \\ \mathbf{N}_2 \end{pmatrix}, \quad [6.49]$$

where:

$$\begin{cases} \mathbf{X} = \big(\text{cpa}_1, v, \theta, t_{\text{cpa}_1}\big)^t, \\ \mathbf{Y}_1(\mathbf{X}) = \big(\alpha_1 \triangleq \text{cpa}_1/v, \theta, t_{\text{cpa}_1}\big)^t, \\ \mathbf{Y}_2(\mathbf{X}) = \big(\alpha_2 \triangleq \text{cpa}_2/v, \theta, t_{\text{cpa}_2}\big)^t, \\ \alpha_2 = \alpha_1 - (d/v)\sin\theta,\ t_{\text{cpa}_2} = t_{\text{cpa}_1} + (d/v)\cos\theta. \end{cases} \quad [6.50]$$

This tells us that it is possible to achieve both detection and complete target motion analysis. However, it can easily be shown [DON 00] that the computational load becomes prohibitive. This is because the parameter discretization has to be fine enough. Furthermore, because of the context (scale detection of a network), the robustness of such a process is very problematic. One possible solution is to use decentralized processing. Thus, we can consider fusion on the target motion analysis level. The objective is then to estimate the complete parameter vector $\mathbf{X}$ ($\mathbf{X} = (\text{cpa}, v, \theta, t_{\text{cpa}})^t$) from partial vectors estimated on the sensor level.

We then have the following geometric relations:

$$\begin{cases} \dfrac{\text{cpa}_2}{v} = \alpha_2 = \alpha_1 - \dfrac{d}{v}\sin(\theta), \\ t_{\text{cpa}_2} = t_{\text{cpa}_1} + \dfrac{d}{v}\cos(\theta), \end{cases} \quad [6.51]$$

and we assume that the partial estimate densities are governed by the following laws:

$$\begin{pmatrix} \hat{\alpha}_1 \\ \hat{\theta}_1 \\ \hat{t}_{\text{cpa}_1} \end{pmatrix} \longrightarrow \mathcal{N}\left[\begin{pmatrix} \alpha_1 \\ \theta \\ t_{\text{cpa}_1} \end{pmatrix}; \Gamma_1\right]$$

$$\begin{pmatrix} \hat{\alpha}_2 \\ \hat{\theta}_2 \\ \hat{t}_{\text{cpa}_2} \end{pmatrix} \longrightarrow \mathcal{N}\left[\begin{pmatrix} \alpha_2 \\ \theta \\ t_{\text{cpa}_2} \end{pmatrix}; \Gamma_2\right].$$

A first idea consists of considering the partial estimation vectors $\hat{\mathbf{Y}}_1$ and $\hat{\mathbf{Y}}_2$ as observations and using an iterative method for estimating the complete state. We can, however, define an admittedly sub-optimal, but explicit, fusion method as follows:

$$\begin{cases} \hat{\theta} = \dfrac{1}{\sigma_1^2 + \sigma_2^2}\left(\sigma_2^2 \hat{\theta}_1 + \sigma_1^2 \hat{\theta}_2\right), \\ \hat{v} = \dfrac{d \sin(\hat{\theta})}{\hat{\alpha}_1 - \hat{\alpha}_2}, \\ \widehat{\mathrm{cpa}}_1 = \hat{\alpha}_1 \dfrac{d \sin(\hat{\theta})}{\hat{\alpha}_1 - \hat{\alpha}_2}. \end{cases} \qquad [6.52]$$

We can then show using Taylor series expansions [DON 00, DON 02] that this estimator of $\mathbf{X}$ asymptotically follows the law:

$$\begin{pmatrix} \widehat{\mathrm{cpa}}_1 \\ \hat{v} \\ \hat{\theta} \\ \hat{t}_{\mathrm{cpa}_1} \end{pmatrix} \longrightarrow \mathcal{N}\left[\begin{pmatrix} \mathrm{cpa}_1 \\ v \\ \theta \\ t_{\mathrm{cpa}_1} \end{pmatrix}; M_1 \Gamma_1 M_1^t + M_2 \Gamma_2 M_2^t\right], \qquad [6.53]$$

where:

$$M_1 = \begin{pmatrix} v\left(1 - \dfrac{\alpha_1}{\alpha_1 - \alpha_2}\right) & \dfrac{\alpha_1 d \cos(\theta)}{\alpha_1 - \alpha_2} \dfrac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} & 0 \\ -\dfrac{v}{\alpha_1 - \alpha_2} & \dfrac{d \cos(\theta)}{\alpha_1 - \alpha_2} \dfrac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} & 0 \\ 0 & \dfrac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad [6.54]$$

and:

$$M_2 = \begin{pmatrix} v\left(\dfrac{\alpha_1}{\alpha_1 - \alpha_2}\right) & \dfrac{\alpha_1 d \cos(\theta)}{\alpha_1 - \alpha_2} \dfrac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} & 0 \\ \dfrac{v}{\alpha_1 - \alpha_2} & \dfrac{d \cos(\theta)}{\alpha_1 - \alpha_2} \dfrac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} & 0 \\ 0 & \dfrac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} & 0 \\ 0 & 0 & 0 \end{pmatrix}. \qquad [6.55]$$

In the equation above, the matrix Γ_i, $(i = 1, 2)$ is unknown, which is why we have to replace it with its estimate. We can then show that the performances of decentralized estimation are quite close to those of centralized estimation. The advantage of this approach, in this context, is therefore quite obvious.

6.10. Discussion

The widespread progress of probabilistic methods, particularly of Bayesian methods, is the result of the knowledge acquired through numerous experiments that helped guide the modeling and learning phases, rather than of Cox's justification [COX 46] (see Appendix A).

The major advantage of probabilistic methods comes from the fact that they rely on a solid mathematical background, and have been the subject of many studies. As a result, they offer a wide selection of tools that can be used both for modeling (for example, using parametric laws with well-studied properties) and model learning (for parametric or non-parametric laws) (see, for example, [CHA 95, LEE 87, LUO 89]). They also suggest usage rules that are either theoretical (bounds, asymptotic values) or heuristic (hypothesis tests, validity criteria, confidence tables). Finally, probabilistic modeling, supported by the frequentist interpretation, which is widespread in the world of physics and signal processing, is a concept currently shared universally, serving as basis for comparison with other models.

Another advantage of probabilistic and statistical methods, this time from the combination perspective, is again that they rely on solid mathematical background and can be used for updating complex knowledge networks [PEA 86a, PEA 86b]. They allow the introduction of information that can easily be expressed in probability form, such as spatial context in the framework of Markov fields (see Chapter 9) or information quality expressed as the probability for a measurement to be reliable [GRA 00].

However, and despite their solid mathematical background, these methods are also criticized and suffer several drawbacks. We will discuss them all in this section, but we should point out that some of them are disputed by unconditional probabilities.

First of all, even though they lead to a good representation of the uncertain nature of information, they cannot easily be used to represent its imprecision and often cause confusion between these two concepts. Furthermore, during the learning phase, they require that very stringent constraints are met by the measurements (imposed by the basic probability axioms) and by the set of considered classes (comprehensiveness). The constraints can make learning very difficult (how is it possible to characterize areas that are not wheat fields in aerial imaging[4]?), or, if the problem to solve is complex,

4. This problem is an example of the more general problem encountered in shape classification and recognition: generally, the complement of a class is not a class.

can lead in practice to inconsistencies because the user cannot in this case take into account the entire network of probabilistic dependences (this is the case for logic loops [PEA 86a]). Learning probability laws require, in addition to the hypotheses, a significant amount of data. Typically, non-parametric learning of a multi-dimensional law in images or areas of limited size is not always relevant and we often resort to the use of parametric models, which, in turn, require hypotheses about the forms of the laws.

The estimation of *a priori* probabilities is often difficult and has a major importance in the cases where little information is available (very flat distributions of the conditional probabilities). If, in the case of image processing, conditional probabilities can often be estimated by learning based on occurrence frequencies, this usually is not the case for *a priori* probabilities. Evaluating them goes beyond the framework of frequentist probabilities and often requires more subjective concepts. Furthermore, Bayesian combination is constrained, as is modeling, by the probability axioms and its use in practice often requires simpler hypotheses (such as independence) that are rarely verified. Probabilistic and Bayesian theory combines the elements of information in a conjunctive way, using products of conditional probabilities, which in practice often leads to a collapse of the probabilities of events that are obtained from a long chain of inference.

The additivity constraint may be too strong for certain problems. Let us consider the example given by Smets [SME 78], in the field of medical diagnosis. If a symptom s is always present in patients with a pathology A and we observe this symptom s, then the probability for the patient to have A increases. The additivity constraint then imposes that the probability for the patient not to have A must decrease, even though there is no reason for it (Hempel's paradox), if the symptom s can also be observed in other pathologies[5].

Applying Bayesian methods often requires considerable knowledge of the problem and achieving good operating conditions implies additional considerations for

5. Because of the additivity constraint and Bayes' rule (equation [A.8]), if $p(s \mid A) = 1$, then $p(A \mid s) = p(A)/p(s)$ and therefore $p(A \mid s) \geq p(A)$. On the contrary, $p(\bar{A} \mid s) = 1 - p(A \mid s)$ and therefore $p(\bar{A} \mid s) \leq p(\bar{A})$. Smets's argument refuting this inequality is debatable, but it can also be interpreted as follows: the additive probability model may be too simplistic in this case. In particular, the idea of the probability of a pathology A can make sense, whereas it is not certain that the probability of $[\bar{A} \mid s]$ does because $\bar{A}$ does not correspond to a *single* pathology, but instead to an infinite, poorly known, imprecise set and it is difficult to claim that $[\bar{A} \mid s]$ is a well-defined, binary proposition that accurately represents reality. Thus, any model that leads to the conclusion $p(\bar{A} \mid s)$ can easily be disputed. We hope that this interpretation does not misrepresent Smets' ideas.

each problem to tackle[6]. For example, the "Bayesian" diagnosis can be formalized as follows:

$$p(A_i \mid O) = \frac{p(O \mid A_i)p(A_i)}{\sum_j p(O \mid A_j)p(A_j)} = \frac{p(O \mid A_i)p(A_i)}{p(O \mid \bar{A}_i)p(\bar{A}_i) + p(O \mid A_i)p(A_i)}, \quad [6.56]$$

where $p(A_i \mid O)$ refers to the probability for a patient to have the pathology A_i, given a set of observations O (clinical examinations, images, etc.), $p(O \mid A_i)$ refers to the conditional probability of the observations given the pathology and $p(A_i)$ is the *a priori* probability of A_i. The decision is made based on the $p(A_i \mid O)$. The use of this formula requires either that all of the pathologies are known, or to have statistics connecting the observations to $\bar{A}_i$ ("non-pathology A_i"). Both solutions seem unrealistic. Furthermore, it has to be possible for all of the probability distributions involved in the formula to be estimated. The problem then is the limit to the statistical tests connecting symptoms or observations with pathologies and the difficulty of having an estimate of the *a priori* probabilities. These limits are of course general and are not specific to this particular example.

Probabilistic modeling can only deal with singletons that represent the different hypotheses, under the closed world's constraint. We saw in the previous example of medical diagnosis how this hypothesis does not fit reality. Furthermore, singletons cannot be used to represent complex situations. Let us take the case of images affected by the partial volume effect (a common situation in medical imaging). The usual models in other works for representing this phenomenon consist of assigning to a point probabilities of belonging to the types of tissue it is comprised of, which are proportional to the amount of each type of tissue in the volume represented by this point. However, this does not correspond to anything real. This type of probabilistic model implies that we are faced with an uncertainty regarding the class to which the point belongs (we know that the point can belong to several classes but we do not know which one), whereas in fact it belongs to several different classes simultaneously. To us, this seems to be the typical example of probabilistic models that are used but do not properly model the observed phenomenon.

Fuzzy sets or belief function theory (or Dempster-Shafer theory) allow us to better describe the reality of certain problems and to find less disputable interpretations.

6. Frequentist methods lead to the opposite situations, particularly Fisher's theory, which is essentially automatic: "faced with a new situation, the statistician can apply maximum likelihood in an automatic fashion, with little chance (in experienced hands) of going far wrong and considerable chance of providing a nearly optimal inference. In short, he does not have to think a lot about the specific situation in order to solve his problem" [EFR 86]. It is a theory of archetypes, that allows us to obtain reasonable solutions by separating the different problems, in cases where the Bayesian approach, which deals with everything "all at once", would be too complex. This automatic nature is certainly one of the reasons behind the popularity of Fisher's theory, despite the disadvantages of the frequentist methods.

Fuzzy sets, for example, can perfectly account for the partial belonging phenomenon that is in fact observed. Relaxing the probability additivity constraint is not enough to solve the problem: the solution does in fact involve the modeling of a completely different phenomenon. We will discuss this further in the next two chapters.

Another limitation stems from the difficulty of adding, to our reasoning system, knowledge that cannot be simply expressed with probabilities.

Along the same idea, it is difficult to model the absence of knowledge, imprecise knowledge (unlike uncertain knowledge, which is naturally represented by probabilities), or also what we do not know about a phenomenon. The insufficient reason principle is not enough for taking into account what is not known and can lead to contradictions depending on how it is expressed.

The same type of problem arises with the maximum entropy principle. Shafer's well-known example about the probability for the existence of life on planet Sirius is a good illustration [SHA 76][7]. These drawbacks, which are not better solved using a subjective version of probabilities, emerge whenever the objective is to model human reasoning that involves decisions based on data that is at the same time imprecise and uncertain, partial, not completely reliable, conflicting, and constraints and objectives that are not always very precise.

6.11. Bibliography

[BAR 85] BARNETT S., CAMERON R., *Introduction to Mathematical Control Theory*, Oxford University Press (Applied Mathematics Series), 1985.

[BAZ 93] BAZARAA M., SHERALI H., SHETTY C., *Non Linear Programming Theory and Algorithms*, Wiley, New York, 1993.

[CHA 92] CHAN Y., RUDNICKI S., "Bearings-Only and Doppler-Bearing Tracking Using Instrumental Variables", *IEEE Transactions on AES*, vol. 28, no. 4, p. 1076-1083, 1992.

[CHA 95] CHAUVIN S., Evaluation des théories de la décision appliquées à la fusion de capteurs en imagerie satellitaire, PhD Thesis, Ecole Nationale Supérieure des Télécommunications and Nantes University, 1995.

---

7. We will not present here Shafer's original example, which is debatable, but an example given by Dubois that is close. Not knowing whether life exists on Sirius is typically expressed in terms of probability by p(life) = p(no life) = 0.5. If we express the problem in another way, by assuming that there are three possibilities, plant life, animal life or no life, the expression of not knowing leads us to assign a probability of 1/3 to each of the three hypotheses. We then get p(life) = 2/3. Likewise, we can obtain as many different values as there are ways of expressing the problem. Similar examples can be found in image processing, particularly in classification problems.

[COO 88] COOKE R., "Uncertainty in Risk Assessment: A Probabilist's Manifesto", *Reliability in Engineering and Systems Safety*, vol. 23, p. 277-283, 1988.

[COO 91] COOKE R., *Experts in Uncertainty*, Oxford University Press, Oxford, United Kingdom, 1991.

[COX 46] COX R.T., "Probability, Frequency and Reasonable Expectation", *Journal of Physics*, vol. 14, no. 1, p. 115-137, 1946.

[DES 93] DESCOMBES X., Champs markoviens en analyse d'images, PhD Thesis, Télécom-Paris ENST 93E026, 1993.

[DES 99] DESROUSSEAUX C., POMORSKI D., "Optimisation entropique des systèmes de détection distribuée", *Traitement du Signal*, vol. 16, no. 4, 1999.

[DON 00] DONATI R., Détection de signaux électriques océaniques. Application à la surveillance de zones, PhD Thesis, Rennes University 1, 2000.

[DON 02] DONATI R., LECADRE J., "Detection of Oceanic Electric Fields Based on the Generalized Maximum Likelihood Ratio Test", *IEE Proc. on Radar, Sonar and Navigation*, vol. 149, p. 221-230, 2002.

[DUD 73] DUDA R., HART P., *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.

[EFR 86] EFRON B., "Why isn't Everyone a Bayesian ?", *The American Statistician*, vol. 40, no. 1, p. 1-11, 1986.

[FRE 85] FRENCH S., "Group Consensus Probability Distributions: A Critical Survey", in J. BERNARDO *et al.* (ed.) *Bayesian Statistics*, p. 183-201, Elsevier, The Netherlands, 1985.

[GRA 00] GRANDIN J.-F., MARQUES M., "Robust Data Fusion", *Fusion 2000*, Paris, p. MoC3-1-9, 2000.

[HUE 02] HUE C., LECADRE J., PÉREZ P., "Tracking Multiple Objects with Particle Filtering", *IEEE Transactions on AES*, vol. 38, no. 3, p. 791-812, 2002.

[ILT 96] ILTIS R., ANDERSON K., "A Consistent Estimation Criterion for Multisensor Bearings-Only Tracking", *IEEE Transactions on AES*, vol. 32, no. 1, p. 108-121, 1996.

[KUL 59] KULLBACK S., *Information Theory and Statistics*, Wiley, New York, 1959.

[LEC 99] LECADRE J., JAUFFRET C., "On the Convergence of Iterative Methods for Bearings-Only Tracking", *IEEE Transactions on AES*, vol. 35, no. 3, p. 801-817, 1999.

[LEC 00] LECADRE J., "Scheduling Active and Passive Measurements", *Fusion 2000*, Paris, 2000.

[LEE 87] LEE T., RICHARDS J.A., SWAIN P.H., "Probabilistic and Evidential Approaches for Multisource Data Analysis", *IEEE Transactions on Geoscience and Remote Sensing*, vol. GE-25, no. 3, p. 283-293, 1987.

[LUC 72] LUCA A.D., TERMINI S., "A Definition of Non-Probabilistic Entropy in the Setting of Fuzzy Set Theory", *Information and Control*, vol. 20, p. 301-312, 1972.

[LUO 89] LUO R.C., KAY M.G., "Multisensor Integration and Fusion in Intelligent Systems", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, no. 5, p. 901-931, 1989.

[MAI 94] MAÎTRE H., BLOCH I., SIGELLE M., "Spatial entropy: a tool for contextual classification control", *IEEE Int. Conf on Image Processing*, vol. II, Austin, Texas, p. 212-216, November 1994.

[MAI 96] MAÎTRE H., "Entropy, Information and Image", in H. MAÎTRE and J. ZINN-JUSTIN (ed.) *Progress in Picture Processing, Les Houches Session LVIII*, p. 881-1115, Springer Verlag, 1996.

[MAT 70] MATHERON G., La théorie des variables régionalisées et ses applications, Report no. 5, Ecole des Mines de Paris, Centre de Morphologie Mathématique de Fontainebleau, 1970.

[NAR 84] NARDONE S., LINDGREN A., GONG K., "Fundamental Properties and Performance of Conventional Bearings-Only Target Motion Analysis", *IEEE Transactions on Automatic Control*, vol. 29, no. 9, p. 775-787, 1984.

[PEA 86a] PEARL J., "Fusion, Propagation, and Structuring in Belief Networks", *Artificial Intelligence*, vol. 29, p. 241-288, 1986.

[PEA 86b] PEARL J., "On Evidential Reasoning in a Hierarchy of Hypotheses", *Artificial Intelligence*, vol. 28, p. 9-15, 1986.

[SHA 76] SHAFER G., *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[SME 78] SMETS P., "Medical Diagnosis: Fuzzy Sets and Degree of Belief", *Colloque International sur la Théorie et les Applications des Sous-Ensembles Flous*, Marseille, September 1978.

[SON 85] SONG T., SPEYER J., "A Stochastic Analysis of a Modified Gian Extended Kalman Filter with Applications to Estimation with Bearing Only Measurement", *IEEE Transactions on Automatic Control*, vol. 30, no. 10, p. 940-949, 1985.

[THO 89] THOMOPOULOS S.C.A., VISWANATHAN R., BOUGOULIAS D.K., "Optimal Distributged Decision Fusion", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 25, no. 5, p. 761-765, 1989.

[TRE 96] TRÉMOIS O., LECADRE J., "Target Motion Analysis with Multiple Arrays: Performance Analysis", *IEEE Transactions on AES*, vol. 32, no. 3, p. 1030-1046, 1996.

[TUP 00] TUPIN F., SIGELLE M., MAÎTRE H., "Definition of a Spatial Entropy and its Use for Texture Discrimination", *IEEE ICIP'2000*, vol. I, Vancouver, Canada, p. 725-728, 2000.

[VAR 97] VARSHNEY P.K., *Distributed Detection and Data Fusion*, Springer Verlag, New York, 1997.

[VOL 95] VOLDEN E., GIRAUDON G., BERTHOD M., "Modeling Image Redundancy", *IGARSS'95*, vol. 3, Florence, Italy, p. 2148-2150, 1995.

Chapter 7

Belief Function Theory

7.1. General concept and philosophy of the theory

Belief function theory (or Dempster-Shafer theory) dates back to the 1970s but its use in signal and image fusion is relatively recent. Nevertheless, the first applications show some promise and in this chapter we will point out the characteristics of this theory that deserve our attention, both from the perspectives of representing knowledge and its imperfections (imprecision, uncertainty, ambiguity, absence of knowledge, conflict) and of combining it.

Although this theory is inspired by concepts of superior and inferior probabilities, and therefore often considered from a probabilistic point of view, it can be interpreted in a more general way, from a subjective point of view, as a quantitative formal model of degrees of confidence [SME 90a]. One of the main assets of this theory is that it deals with subsets rather than singletons, making it very flexible for modeling many of the situations we come across in signal and image fusion. It also provides us with representations of uncertainty, imprecision, as well as of the absence of knowledge. For this purpose, several functions are used to model the information and manipulate it, instead of simply the probabilities used in the previous chapter. This theory can be used to measure conflicts between sources and to interpret them in terms of the reliability of the sources, of an open world or of contradicting observations. Although several combination modes are possible, conjunctive combination is the most commonly used in the fields we are concerned with here and we will focus mostly on this mode. This means that the essential part of the user's work will be transferred to

---

Chapter written by Isabelle BLOCH.

the phase that consists of modeling and representing the information and knowledge available.

From now on, we will very often illustrate our discussion with the example of multi-source classification.

7.2. Modeling

Belief function theory, like possibility theory, allows us, as we will see in the following chapter, to represent both imprecision and uncertainty using mass functions m, plausibility functions Pls and belief functions Bel [GUA 91, SHA 76, SME 90a]. Mass functions are defined for all of the subsets in the space D, referred to here as the frame of discernment (containing, for example, the classes we are interested in), and not simply the singletons such as probabilities which only measure the probability of belonging to a given class.

Let us assume that $D = \{C_1, C_2, \ldots, C_n\}$ where each C_i refers to a hypothesis that supports a decision (typically a class in a multi-source classification problem). A mass function is defined as a function of 2^D (the sets of subsets of D) into $[0, 1]$. Usually, the condition $m(\emptyset) = 0$ is imposed, as well as a normalization of the form:

$$\sum_{A \subseteq D} m(A) = 1, \tag{7.1}$$

which guarantees the commensurability between several sets of masses.

The constraint $m(\emptyset) = 0$ corresponds to a closed world hypothesis, in which all of the possible situations are in fact represented in D (which implies that we are capable of making a list of them). If this constraint is relaxed and if we accept having a mass that is strictly positive over $\emptyset$, this then corresponds to an open world hypothesis, in which the solutions outside of D can be considered.

A focal element is a subset A of D such that $m(A) > 0$. The collection of focal elements is called the core.

A belief function Bel is a totally increasing function defined from 2^D into $[0, 1]$:

$$\forall A_1 \in 2^D, \ldots A_k \in 2^D,$$
$$\mathrm{Bel}\left(\cup_{i=1\cdots k} A_i\right) \geq \sum_{I \subseteq \{1\cdots k\},\ I \neq \emptyset} (-1)^{|I|+1}\, \mathrm{Bel}\left(\cap_{i \in I} A_i\right), \tag{7.2}$$

where $|I|$ refers to the number of elements in I and such that $\mathrm{Bel}(\emptyset) = 0$, $\mathrm{Bel}(D) = 1$.

Given a mass function m, the Bel function defined by:

$$\forall A \in 2^D, \quad \mathrm{Bel}(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B) \qquad [7.3]$$

is a belief function. Conversely, from a belief function defined as a totally increasing function (inequality [7.2]) such that $\mathrm{Bel}(\emptyset) = 0$, $\mathrm{Bel}(D) = 1$, we can define a mass function by:

$$\forall A \in 2^D, \quad m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \, \mathrm{Bel}(B). \qquad [7.4]$$

This mass function then verifies equation [7.3].

The belief function measures the total confidence placed in the set A. The empty set is excluded from the sum because it would otherwise be found in both the evaluation of A and the evaluation of A^C ($\emptyset \subset A$ and $\emptyset \subset A^C$).

Thus, having a zero mass on a subset A does not mean that this set is impossible, simply that we are not capable of assigning a level precisely to A, since we could have non-zero masses on subsets of A, which would lead us to $\mathrm{Bel}(A) \neq 0$. This comment is very important for modeling because it allows us not to assign confidence values when we are not able to do so (this way, we are not forcing information where none is available).

In the open world hypothesis, we have:

$$\mathrm{Bel}(D) = 1 - m(\emptyset). \qquad [7.5]$$

A plausibility function Pls is also a function of 2^D into $[0, 1]$ defined by:

$$\forall A \in 2^D, \quad \mathrm{Pls}(A) = \sum_{B \cap A \neq \emptyset} m(B) = 1 - \mathrm{Bel}\left(A^C\right). \qquad [7.6]$$

More generally, in order to account for the possibility of dealing with an open world, we have:

$$\mathrm{Pls}(A) = \sum_{B \cap A \neq \emptyset} m(B) = \mathrm{Bel}(D) - \mathrm{Bel}(A^C). \qquad [7.7]$$

Plausibility measures the maximum confidence that can be placed in A. This function has a natural interpretation in the transferable belief model [SME 90a] in which additional information is considered to allow for the transfer of belief to more precise subsets. Plausibility then represents the maximum belief that could potentially be assigned to a subset A if we learned, for example, that the solution is in A (all of the confidence placed in a subset B intersecting A is then transferred to A in order to lower to 0 the confidence in A^C).

We have the following properties:

$$\forall A \in 2^D, \quad \mathrm{Pls}(A) \geq \mathrm{Bel}(A), \qquad [7.8]$$

$$\forall A \in 2^D, \quad \mathrm{Bel}(A) + \mathrm{Bel}(A^C) \leq 1, \qquad [7.9]$$

$$\forall A \in 2^D, \quad \mathrm{Pls}(A) + \mathrm{Pls}(A^C) \geq 1, \qquad [7.10]$$

$$\forall A \in 2^D, \quad \mathrm{Bel}(A) + \mathrm{Bel}(A^C) = 1 \Longleftrightarrow \mathrm{Bel}(A) = \mathrm{Pls}(A). \qquad [7.11]$$

The interval $[\mathrm{Bel}(A), \mathrm{Pls}(A)]$ is referred to as the confidence interval and its length measures the absence of knowledge we have of an event A and its complement.

If we assign masses only to the simple hypotheses ($m(A) = 0$ for $|A| > 1$), then the three functions m, Bel and Pls are equal and are a probability. They are referred to as Bayesian mass functions. In more complex situations, this is not the case and there is no direct equivalence with probabilities. Functions similar to credibility and plausibility functions could be obtained, for example, from probabilities conditional to pessimistic and optimistic behaviors respectively, but their formalization would be much more difficult than what belief function theory has to offer.

Among the distinctive mass functions, there is a category for simple support functions, for which all of the mass is assigned to a non-empty subset A and to a set of discernment D:

$$\begin{aligned} &m(A) = s \\ &m(D) = 1 - s \\ &m(B) = 0 \quad \text{for any } B,\ B \neq A,\ B \neq D, \end{aligned}$$

with $s \in [0, 1]$.

If s is equal to 0, then the entire mass is assigned to D. This function represents the total lack of knowledge, in the sense that no subsets can be distinguished.

The possibility of assigning masses to the composite hypotheses and therefore to work on 2^D rather than on D constitutes one of the advantages of this theory because it allows for very flexible and rich modeling, particularly of ambiguity or hesitation between two classes. Here are a few examples of situations in which fusion based on belief function theory can be used:

– in extreme cases (which can be considered ideal) where we know all of the information regarding the problem at hand;

– when a source only provides information on certain classes: for example, certain PET[1] images provide information on the limits of the brain but not the head;

– when a source is not capable of telling the difference between two hypotheses: belief function theory can then be used to consider the disjunction of these two classes, without adding arbitrary information that would force their separation;

– when attempting to model partial volume effects, typically by representing a pixel or a voxel belonging to several classes;

– when attempting to represent a source's overall reliability: this can be done by assigning a non-zero mass to D;

– in cases where a source's reliability depends on the classes (for example, the anatomical information provided by functional brain images is not very reliable, whereas MRI images are very reliable for anatomical classes);

– in cases where we want to add *a priori* information: even if it is not easy to use probabilities to represent this information, it can still be added if they lead to a way of choosing focal elements (particularly hypothesis disjunctions), of defining or modifying mass functions.

7.3. Estimation of mass functions

Estimation of mass functions is a difficult problem because there is no universal solution. The difficulty gets worse here if we want to assign masses to the composite hypotheses [GAR 86, LOW 91]. In image processing, for example, they can be constructed on three levels: on the highest level (which is often abstract and symbolic), the representation of information is used in a fashion similar to what is done in artificial intelligence and masses are assigned to propositions, and often provided by experts [BAL 92, GOR 85, NEA 92]. Most of the time, this information is not derived directly from data measurements and the corresponding methods are therefore not specific to signal and image processing. On an intermediate level, masses are calculated based on attributes and can rely, for example, on image geometric models [AND 88, CHE 93, CLE 91, CUC 92]. This level is well-suited for model-based shape recognition problems, but it is difficult to use for fusion problems on complex structures without a model. On a low level (the pixel in image processing), many methods are possible and most rely on statistical shape recognition methods.

1. Positron Emission Tomography, used in particular for functional brain imaging.

The simplest way imaginable consists of calculating the masses on the singletons in a source (an image, for example) I_j by:

$$m_j\big(\{C_i\}\big)(x) = M_i^j(x), \quad [7.12]$$

where $M_i^j(x)$ is estimated most often as a probability. The masses on all of the other subsets D are then equal to zero. Clearly, this model is very simplistic and does not put to use the interesting characteristics of belief function theory. However, many methods rely on an initial model like this one, or only use certain composite hypotheses, in a simplifying and often very heuristic method [CLE 91, LEE 87, RAS 90, ZAH 92]. Recent work addressed the problem of estimating belief functions from sample data. For instance, belief functions are estimated from realizations of a random variable, with the constraint that they converge towards the probability distribution of this variable when the sample size goes towards infinity (see [DEN 06]). But other methods can also be considered. In the following sections, we present a few of the models found in other works.

7.3.1. *Modification of probabilistic models*

The simplest and most often used model consists of using the discounting technique [SHA 76]. The new masses m' are calculated based on the initial masses m as follows (the index j representing the source of information, as well as the element x we are reasoning on, are omitted here):

$$m'\big(\{C_i\}\big) = \alpha m\big(\{C_i\}\big), \quad [7.13]$$

$$m'(D) = 1 - \alpha + \alpha m(D), \quad [7.14]$$

where $\alpha \in [0, 1]$ is the discounting coefficient. In the case where the initial masses are learned from singletons only, for example, based on probabilities, then $m(D) = 0$ and $m'(D) = 1 - \alpha$. This technique is often used to weaken a source depending on its reliability and makes it possible to assign to D a mass that will be small if the source is reliable and high if it is not. In extreme cases, the value $\alpha = 0$ is used for a source that is not reliable at all and all of the mass is then assigned to D, which represents the total absence of knowledge. The value $\alpha = 1$ is used for a reliable source in which all of the mass is assigned to the singletons and there is no ambiguity between classes.

This type of model is very simple. Learning the masses of the singletons can benefit from the usual techniques of statistical learning. However, hypothesis disjunctions are not modeled, which strongly reduces the applicability of this model.

Two models based on the probabilistic approach have been suggested by Appriou [APP 93], taking into account disjunctions other than D. These models assume that initial estimations have been conducted of the conditional probabilities $p(f(x) \mid C_i)$ (where $f(x)$ refers to the characteristics of x extracted from the source and on which the fusion is based), which are denoted more simply by $p(x \mid C_i)$. The mass function associated with a source is calculated by combining the mass functions associated with each singleton, defined in the first model by:

$$m^i\big(\{C_i\}\big)(x) = \frac{\alpha_i R p\big(x \mid C_i\big)}{1 + R p\big(x \mid C_i\big)}, \qquad [7.15]$$

$$m^i\big(D \setminus \{C_i\}\big)(x) = \frac{\alpha_i}{1 + R p\big(x \mid C_i\big)}, \qquad [7.16]$$

$$m^i(D)(x) = 1 - \alpha_i, \qquad [7.17]$$

where α_i is a discounting coefficient related to the class C_i, which makes it possible to take into account the source's reliability for this particular class (and not its overall reliability, unlike the previous model) and R is a probability weighting coefficient. If $R = 0$, only the reliability of the source is taken into account, otherwise the data is also taken into account.

In the second model, the masses associated with each singleton are defined by:

$$m^i\big(\{C_i\}\big)(x) = 0, \qquad [7.18]$$

$$m^i\big(D \setminus \{C_i\}\big)(x) = \alpha_i\big(1 - R p\big(x \mid C_i\big)\big), \qquad [7.19]$$

$$m^i(D)(x) = 1 - \alpha_i + \alpha_i R p\big(x \mid C_i\big). \qquad [7.20]$$

This model corresponds to the case where $p(x|C_i)$ gives us information essentially on what C_i is not.

The mass associated with the source is then calculated as $\oplus_i m^i$, where $\oplus$ is Dempster's orthogonal sum (see section 7.4). This model is well-suited for cases where one class is easily learned compared to all of the others, which is common in shape recognition in images, or in the case when each class is determined based on an adequate sensor (for example, a road sensor in an aerial image can be used to define the probability of belonging to the road, as opposed to belonging to all of the other classes, but is not capable of telling these other classes apart).

In [DRO 97], disjunctions are defined based on a significance criterion for the conditional probabilities. If only one probability $p(x|C_i)$ is significant (thus creating the need to define thresholds), then a simple mass model involving the singletons is used. If several probabilities are significant, the disjunctions of the corresponding hypotheses are also taken into account. For example, if three values are significant and are such that $p(x|C_i) > p(x \mid C_j) > p(x \mid C_k)$, the mass function is defined by:

$$m\big(\{C_i\}\big)(x) = p\big(x \mid C_i\big) - p\big(c \mid C_j\big), \qquad [7.21]$$

$$m\big(\{C_i \cup C_j\}\big)(x) = p\big(x \mid C_j\big) - p\big(x \mid C_k\big), \qquad [7.22]$$

$$m\big(\{C_i \cup C_j \cup C_k\}\big)(x) = p\big(x \mid C_k\big), \qquad [7.23]$$

then the masses are normalized. If no probability is significant, the mass is assigned entirely to D.

7.3.2. *Modification of distance models*

An approach using shape recognition is suggested in [DEN 95]. If each class C_i is represented by a prototype (also called a center) x_i, then a mass function associated with each class can be defined, in which C_i and D are the only focal elements:

$$m^i\big(\{C_i\}\big)(x) = \alpha e^{-\gamma d^2(x,x_i)}, \qquad [7.24]$$

$$m^i(D)(x) = 1 - \alpha e^{-\gamma d^2(x,x_i)}. \qquad [7.25]$$

The parameters α and γ allow us to modify the amount of absence knowledge and the types of mass functions. Using the distance $d^2(x, x_i)$, the mass can be set so as to be high when x "is similar" to the class C_i. The m^i are then combined according to Dempster's rule (see section 7.4) in order to have a mass that takes into account the information about all of the classes.

This approach can also be applied to the k closest neighbors. The distance is the one between x and one of its neighbors, and the mass is assigned, according to the previous model, to the class to which this neighbor belongs and to D. The functions calculated for each of the neighbors of x are then combined using Dempster's rule.

7.3.3. *A priori information on composite focal elements (disjunctions)*

In many applications, it is possible to have *a priori* information available that can be used to determine, under some supervision, which focal elements should be taken into account. These methods were used, for example, in [BLO 96, MIL 00, MIL 01, TUP 99]. In [BLO 96], images of the brain are combined to detect pathologies (see section 7.7). Mass functions are automatically estimated based on gray levels [BLO 97b] and the classes that cannot be distinguished in certain images from their gray levels are grouped together in disjunctions. In [TUP 99], the results from sensors in several structures are fused in order to interpret a radar image. The abilities of a sensor to tell the difference or not between various classes of structures is what makes it possible to define focal elements and the class disjunctions that need to be taken into account. In [MIL 00, MIL 01], attributes extracted from images provided by different sensors are combined in order to distinguish mines from harmless objects, in a humanitarian demining program. The measurements to combine can be specific to a class or to the entire frame of discernment. For example, the depth of the objects can be used to assign a mass to harmless objects if it is high, but cannot be used to tell the difference between objects if it is low and the mass is then assigned to D.

This type of approach is very effective if such information is available, but it remains supervised and therefore can only be applied to problems with a reasonable number of elements in D.

7.3.4. *Learning composite focal elements*

Learning methods for focal elements often rely on prior classifications conducted in each source separately. Typically, based on confusion matrices, it is possible to identify classes confused according to a source, whose reunion will constitute a focal element of the mass function assigned to this source.

In a completely non-supervised way, the intersection between the classes detected in a source and those detected in another source can define the singletons and the frame of discernment, with the classes detected in each source then becoming disjunctions [MAS 97].

Dissonance and consonance measures are given in [MEN 96]. The idea consists of modifying an initial mass function, involving only the singletons, by discounting the masses of the singletons depending on their levels of consonance and by creating masses for disjunctions of two classes depending on the level of dissonance between these two classes. This method was applied to the fusion of several classifiers. The consonance of a class is calculated based on the number of elements affected to that class by all of the classifiers and the dissonance based on the number of elements that are classified differently.

In the case of elements that are characterized by a measurement in a one-dimensional space (represented typically by a histogram), the masses on the composite hypotheses can be defined in the areas of overlapping or ambiguity between two neighboring classes. Another method, based on thresholding by hierarchy, is suggested in [ROM 99] where each peak of the histogram corresponds to a singleton. Then the histogram is progressively thresholded at decreasing heights and disjunctions are created when maxima are grouped together. This method can be compared to component trees, which are used, for example, in mathematical morphology with the concept of cup topology [DOK 00], and to confidence intervals and their relations with possibility distributions [DUB 99, MER 05] (see Chapter 8).

7.3.5. *Introducing disjunctions by mathematical morphology*

Without being restricted to one-dimensional representation spaces, the method suggested in [BLO 97a] allows for the calculation of masses for disjunctions, by erosions and dilations of masses first defined for singletons. The properties of these morphological operations (of duality in particular) make it possible to interpret them as beliefs and plausibilities, from which the masses can then be inferred.

7.4. Conjunctive combination

7.4.1. *Dempster's rule*

Let m_j ($j = 1 \cdots l$) be the mass function defined for the source j. The conjunctive combination of the mass functions is conducted according to Dempster's orthogonal rule [SHA 76, SME 90a], defined $\forall A \subseteq D$ by:

$$\big(m_1 \oplus m_2 \oplus \cdots \oplus m_l\big)(A) = \sum_{B_1 \cap \cdots \cap B_l = A} m_1\big(B_1\big)m_2\big(B_2\big) \cdots m_l\big(B_l\big). \quad [7.26]$$

Axiomatic justifications of this rule can be found in [SME 90a]. The differences between these axioms and those of Cox [COX 46] (which are used to justify the probability rules) explain the origins of the differences between the two theories [BLO 95]. These aspects are discussed in Appendix B.

7.4.2. *Conflict and normalization*

In the non-normalized equation [7.26], the mass assigned by combination to the empty set is usually not equal to zero. It is often interpreted as the conflict between the sources. Let us note that this conflict measurement is not absolute, but instead depends on the modeling (particularly of the distribution of masses among the different subsets of D). There are two essential sources of conflict: either the sources are not reliable, or they provide information about different phenomena. In the first case, it is acceptable to combine the sources and a solution for taking the conflict into account is to weaken the sources according to their reliability. We will discuss this further later on. In the second case, combination makes no sense. Methods for regrouping sources according to the phenomena they observe have been suggested, with the objective of combining only the sources within each group. These groups are calculated in such a way as to minimize the conflict within each group [MIL 01, SCH 93].

In an open world hypothesis, a non-zero mass on the empty set can also represent a solution that was not predicted in D. Under a closed world hypothesis, where everything that is possible is represented in D, this interpretation is not acceptable, which leads us to normalizing the result of the combination in the following form[2]:

$$\big(m_1 \oplus \cdots \oplus m_l\big)(A) = \frac{\sum_{B_1 \cap \cdots \cap B_l = A} m_1\big(B_1\big) \cdots m_l\big(B_l\big)}{1 - \sum_{B_1 \cap \cdots \cap B_l = \emptyset} m_1\big(B_1\big) \cdots m_l\big(B_l\big)}, \quad [7.27]$$

2. This normalized form is Dempster's rule in its strict sense [SHA 76]; the non-normalized rule was suggested later [SME 90a] but seems preferable today for most applications.

and

$$\big(m_1 \oplus m_2 \oplus \cdots \oplus m_l\big)(\emptyset) = 0$$

if the denominator of equation [7.27] is not equal to zero, i.e. if:

$$k = \sum_{B_1 \cap \cdots \cap B_l = \emptyset} m_1\big(B_1\big) m_2\big(B_2\big) \cdots m_l\big(B_l\big) < 1. \qquad [7.28]$$

Therefore, this quantity (which measures the conflict between sources) is directly taken into account in the combination in the form of normalization factor. It represents the mass that would be assigned to the empty set if we did not have this normalization (equation [7.26]). It is important to take into account this value in order to appreciate the quality of the combination: it may not make much sense in case of a strong conflict and lead to decisions that could be disputed.

Let us consider a simple example in which $D = \{C_1, C_2, C_3\}$ and two mass functions with only the singletons as their focal elements and the following values:

	C_1	C_2	C_3
m_1	0.9	0.0	0.1
m_2	0.0	0.9	0.1

Their non-normalized and normalized fusions lead to:

	C_1	C_2	C_3	$\emptyset$
$m_1 \oplus m_2$ non-normalized	0.0	0.0	0.01	0.99
$m_1 \oplus m_2$ normalized	0.0	0.0	1.0	0.0

All of the mass is then concentrated in C_3, which is the only class in which both sources agree, but only to say that the solution is hardly plausible. The normalization masked the conflict. The non-normalized form is often preferable in case of conflict. Here, it allows us to assign the essential part of the mass to the empty set and the origin of the conflict can be attributed to the open world hypothesis, a low reliability of at least one of the two sources, or to the fact that one source sees a class C_1 object whereas the second source see another class C_2 object.

Methods other than normalization have been suggested for eliminating the mass assigned to the empty set. For example, this mass is assigned to D in [YAG 87], in other words assigned to ignorance. In [DUB 88], a more subtle method is suggested: for example, if the focal elements A_1 and A_2 of two sources are in conflict ($A_1 \cap A_2 = \emptyset$), then the product $m_1(A_1)m_2(A_2)$ is assigned to $m(A_1 \cup A_2)$. This means we are assuming that at least one of the two sources is reliable without specifying which and the disjunctive form of the result is the most cautious attitude.

7.4.3. *Properties*

Let us now examine the properties of the combination rule. It is commutative and associative. The mass function defined by:

$$m_0(D) = 1 \text{ and } \forall A \subseteq D,\ A \neq D,\ m_0(A) = 0 \qquad [7.29]$$

is the neutral element for the combination. This mass represents a completely uninformative source, which is unable to distinguish any element of D. In fact, this is what our intuition tells us, that the mass function plays no part in the combination. The definition of this mass function replaces the indifference principle used in probabilities (equal distribution of the probabilities over all of the elements) and better represents the absence of information.

The law $\oplus$ is not idempotent. Consider again the previous example, but this time with the following mass functions:

	C_1	C_2	C_3
m_1	0.7	0.2	0.1
m_2	0.7	0.2	0.1

Their non-normalized and normalized fusions lead us to:

	C_1	C_2	C_3	$\emptyset$
$m_1 \oplus m_2$ non-normalized	0.49	0.04	0.01	0.46
$m_1 \oplus m_2$ normalized	0.91	0.07	0.02	0.0

This example illustrates the non-idempotence of the combination rule. The strongest values are reinforced and the smaller ones weakened. It is also important to note that the conflict between two identical mass functions is not equal to zero, and that it increases as the mass becomes more spread out over the singletons.

At first, this combination rule was thought to be applicable only under the source independent hypothesis. It has been shown [QUI 89, QUI 91] that the rule can still be applied without this hypothesis, by relying on the analogy with random closed sets. In less technical and more philosophical terms, independence in the framework of belief functions should not be understood in a statistical sense, but instead in a more "cognitive" sense [SME 93]. This is referred to as cognitive independence. Imagine, for example, that we wish to combine the opinions of experts. They are likely not to be statistically independent (if they are experts in the same field), but we can expect them to be cognitively independent, i.e. each one makes up his own opinion without consulting the others. This is the type of independence Dempster's rule applies to, which results in the non-idempotence of the rule, causing the reinforcement of identical mass functions. Under the dependence hypothesis, we would want an idempotent rule instead. We will continue with these considerations with fuzzy set theory in Chapter 8.

When the functions m, Bel and Pls are probabilities (i.e. when only the focal elements are singletons), Dempster's combination law is consistent with the traditional probability laws. In this light, probabilities are shown as the limit of belief theory, when there is no ambiguity or imprecision and only the uncertainty of the data needs to be taken into account.

Dempster's rule has a conjunctive behavior, since it provides focal elements that are the intersections of the focal elements of the initial mass functions. Therefore, it reinforces focusing and decreases the length of the confidence intervals $[\mathrm{Bel}, \mathrm{Pls}]$.

In practice, the combination calculation is conducted by laying down the intersection table of the focal elements. For example, if m_1 pertains to $C_1 \cup C_2$ (typical of a source no longer capable to tell two classes apart) and C_3, and m_2 to C_1 and $C_2 \cup C_3$, the focal elements of $m_1 \oplus m_2$ are given by the following intersection table:

	$C_1 \cup C_2$	C_3
C_1	C_1	$\emptyset$
$C_2 \cup C_3$	C_2	C_3

In this case, the focal elements are simply the singletons and the empty set. This example illustrates how conjunctive combination reduces the imprecision and solves (or generally reduces) the ambiguity of each source.

In particular, the intervals $[\mathrm{Bel}, \mathrm{Pls}]$ are reduced after the combination, since the mass functions are more focused than the initial masses (they pertain to smaller sets).

On the other hand, the mass on the empty set (which measures conflict) increases during fusion.

Let us now assume that the model partly includes the absence of knowledge and that a non-zero mass is assigned to D in both sources. We then get the following intersection table:

	$C_1 \cup C_2$	C_3	D
C_1	C_1	$\emptyset$	C_1
$C_2 \cup C_3$	C_2	C_3	$C_2 \cup C_3$
D	$C_1 \cup C_2$	C_3	D

This time, the ambiguity is only partially reduced and non-zero mass remains on the imprecise elements (class disjunction). The conflict, on the other hand, is diminished, which is a property that is true in general of the combination of weakened masses by reinforcing D.

7.4.4. *Discounting*

Discounting allows us to model the reliability of the sources with the help of a coefficient α ($\alpha \in [0, 1]$) used for increasing the mass on D. The idea is to strengthen the absence of knowledge as the source becomes less reliable. A mass function m is then transformed into a mass function m' according to the following formulae:

$$m'(A) = \alpha m(A)\ \forall A,\ A \neq D,$$

$$m'(D) = 1 - \alpha\big(1 - m(D)\big) = 1 - \alpha + \alpha m(D).$$

The effect of discounting is to increase the intervals $[\mathrm{Bel}, \mathrm{Pls}]$ and, during the combination, to reduce conflict.

7.4.5. *Conditioning*

We now consider the specific case of a source that provides a certain element of information on a subset B of D. This information is modeled as follows:

$$m_B(B) = 1 \text{ and } \forall A \subseteq D,\ A \neq B,\ m_B(A) = 0. \qquad [7.30]$$

All of the sources have to be “conditioned” by m_B, in order to account for the fact that the truth can only be in B. Conditioning is done simply by combining a mass function m with m_B:

$$\forall A \subseteq D,\ m \oplus m_B(A) = \sum_{A=B\cap C} m(C), \tag{7.31}$$

which can also be written:

$$\forall A \subseteq D,\ A \not\subseteq B,\ m \oplus m_B(A) = 0, \tag{7.32}$$

$$\forall A \subseteq D,\ A \subseteq B,\ m \oplus m_B(A) = \sum_{X\subseteq B^C} m(A \cup X). \tag{7.33}$$

Conditioning is performed in accordance with the transferable belief model [SME 90a]: knowledge of B leads us to transferring all of the mass on the subsets included in B. Thus, the belief initially assigned to a subset $A = A_1 \cup A_2$ (with $A_1 \subseteq B$ and $A_2 \subseteq B^C$) represented the fact that the truth could be anywhere in A. Knowledge of B can now be used to specify the information and to reduce A to A_1. In a way, the diffuse belief in A is now concentrated in the only part that is included in B.

Conditioning performed according to the conjunctive rule is the equivalent, in the framework of belief functions, of conditional probabilities, which also corresponds to a conjunction. This is because we have:

$$P(X \mid B) = \frac{P(X \cap B)}{P(B)}.$$

7.4.6. *Separable mass functions*

We now consider simple support mass functions. If m_1 and m_2 are simple support functions with the same support A, with weights s_1 and s_2, then the combination yields a function with the same support and a weight $s_1 + s_2 - s_1 s_2$. Such functions are never cause for conflict.

If both functions have different supports A_1 and A_2, then the combination leads to:

$$\begin{gathered}
\big(m_1 \oplus m_2\big)\big(A_1 \cap A_2\big) = s_1 s_2 \\
\big(m_1 \oplus m_2\big)\big(A_1\big) = s_1\big(1 - s_2\big) \\
\big(m_1 \oplus m_2\big)\big(A_2\big) = s_2\big(1 - s_1\big) \\
\big(m_1 \oplus m_2\big)(D) = \big(1 - s_1\big)\big(1 - s_2\big) \\
\big(m_1 \oplus m_2\big)(B) = 0\ \forall B,\ B \neq A_1,\ B \neq A_2,\ B \neq A_1 \cap A_2,\ B \neq D.
\end{gathered}$$

Particularly, if $A_1 \cap A_2 = \emptyset$, then both functions cause conflict, since a non-zero mass is assigned to the empty set. When $s_1 s_2 \neq 1$, the resulting function is referred to as a separable mass function.

7.4.7. *Complexity*

In the general case, as shown by formula [7.26], the combination has an exponential complexity. In practice, it is rare to have to take into account all of the subsets of D and the complexity is often more reasonable. A linear complexity is obtained if the masses are modeled according to Barnett's structure [BAR 81], i.e. if the focal elements of each source are only the singletons and their complements (separable functions). This structure is suited for shape recognition problems in which each source is a detector that can be used to distinguish one class from all the others. But it is not general and cannot be applied to sources which require focal elements that can be any disjunctions.

7.5. Other combination modes

Other combination modes, such as disjunctive or compromise modes, are possible by replacing the intersection in formula [7.26] with another set operation. For example, disjunctive fusion is obtained by taking the union [SME 93]:

$$\big(m_1 \oplus_{\cup} \cdots \oplus_{\cup} m_l\big)(A) = \sum_{B_1 \cup \cdots \cup B_l = A} m_1\big(B_1\big) \cdots m_l\big(B_l\big). \qquad [7.34]$$

Let us note that this combination cannot lead to conflict. It widens the focal elements therefore providing less precise information from each of the sources. This fusion can be useful if we are unable to model beforehand the reliabilities, ambiguities and imprecisions of the sources. For example, if a source is focused in A and another one in B with $A \cap B = \emptyset$, one way of not solving the conflict is to conclude that the truth is in $A \cup B$, thus allowing a disjunctive fusion.

However, in most image fusion applications, the goal is to obtain a combined mass function that is more focused than the initial masses. This is why conjunctive fusion is the preferred method, since it implies that the imprecisions, reliabilities, ambiguities of each source are taken into account in the modeling stage. It then constitutes the most crucial phase and requires the most attention.

7.6. Decision

Once the combined mass functions have been combined, the belief and plausibility functions are inferred from equations [7.3] and [7.6]. The last step is the decision phase, where a subset of D has to be chosen to maximize a certain criterion. From here on, m, Bel and Pls refer to the mass, belief and plausibility functions obtained after the combination.

In belief function theory, several decision rules are possible and are most of the time applied to the choice of a singleton C_i.

The maximum plausibility:

$$x \in C_i \text{ if } \mathrm{Pls}\big(C_i\big)(x) = \max\big\{\mathrm{Pls}\big(C_k\big)(x),\ 1 \leq k \leq n\big\}, \qquad [7.35]$$

this rule being optimal in the sense laid down by probabilistic criteria for mass functions derived from probabilities [APP 91].

The maximum credibility:

$$x \in C_i \text{ if } \mathrm{Bel}\big(C_i\big)(x) = \max\big\{\mathrm{Bel}\big(C_k\big)(x),\ 1 \leq k \leq n\big\}, \qquad [7.36]$$

which is equivalent to the maximum plausibility criterion in the case where the result of the combination only involves singletons.

The maximum credibility without confidence interval overlap (without the risk of an error):

$$x \in C_i \text{ if } \mathrm{Bel}\big(C_i\big)(x) \geq \max\big\{\mathrm{Pls}\big(C_k\big)(x),\ 1 \leq k \leq n,\ k \neq i\big\}, \qquad [7.37]$$

this last condition being particularly strict and possibly leading to no decision being made.

The maximum credibility with discarding [MAS 97]:

$$x \in C_i \text{ if } \mathrm{Bel}\big(C_i\big)(x) = \max\big\{\mathrm{Bel}\big(C_k\big)(x),\ 1 \leq k \leq n\big\} \qquad [7.38]$$

and

$$\mathrm{Bel}\big(C_i\big)(x) \geq \mathrm{Bel}\big(C_i^C\big),$$

which expresses the fact that the decision has to be unambiguous enough since the condition will be met if the mass is very focused on C_i.

The maximum pignistic probability defined by [SME 90b]:

$$\forall C_j \in D,\ BetP\big(C_j\big) = \sum_{C_i \in A} \frac{m(A)}{|A|\big(1 - m(\emptyset)\big)}, \qquad [7.39]$$

where $|A|$ refers to the number of elements in A thus making it possible to switch to a probabilistic framework, which is often desired for making the decision (or the bet) or for associating this decision with other probabilistic criteria, for example, in the framework of Markov fields for spatial regularization criteria [TUP 99].

Mixed rules have also been suggested, in which plausibility is used for certain classes and belief for others. This makes it possible to favor the detection of classes for which plausibility is considered [MIL 01].

The decision can also be made in favor of a disjunction. In this case, it is imprecise but does allow us to take into account class mixture or ambiguities remaining after the fusion. This type of decision is interesting, for example, when taking into account the partial volume effect and the voxels affected by it will thus be classified as mixture voxels, rather than voxels from pure classes, as our intuition would tell us [BLO 96]. The decision also allows us to indicate the elements for which fusion is not enough to clear up the ambiguities and therefore to suggest the acquisition of new information, as well as the use of active fusion [GAN 96, PIN 95].

Finally, decision rules with costs have been suggested [DEN 95]. For any function f of D in $\mathbb{R}$, the lower and upper expectations of f relative to a belief function Bel, in Dempster's sense, are defined by:

$$E_*(f) = \sum_{A \subseteq D} m(A) \min_{C_i \in A} f\big(C_i\big), \quad [7.40]$$

$$E^*(f) = \sum_{A \subseteq D} m(A) \max_{C_i \in A} f\big(C_i\big). \quad [7.41]$$

Decision rules with costs are then obtained by choosing for f a function that expresses the cost of an action when the element to which the decision pertains belongs to the class C_i. This cost function can also be introduced with a traditional probabilistic decision rule with costs, using pignistic probability. Thus, the decision can be optimistic if the lower expectation is minimized, pessimistic if the upper expectation is minimized, or intermediate if the pignistic probability is used.

7.7. Application example in medical imaging

The application we have chosen here, to give the reader an idea of the potential of belief function theory, is the classification of MRI images presenting a pathology known as adrenoleukodystrophy (ALD), which are acquired with two echo times [BLO 96]. For doctors, obtaining significant measurements requires a segmentation of both the pathological areas and ventricles, which are visible on different images. The initial images are represented in Figure 7.1. This figure shows a good discrimination between the brain, the ventricles (V) and the cerebrospinal fluid (CSF) on the first image, but white matter (WM) cannot be distinguished from gray matter (GM), or WM from CSF. On the other hand, the ALD area is clearly visible on the second image (in white). This image presents small differences between WM and GM, but the ventricles have almost the same gray levels as the GM and their contours are indistinct.

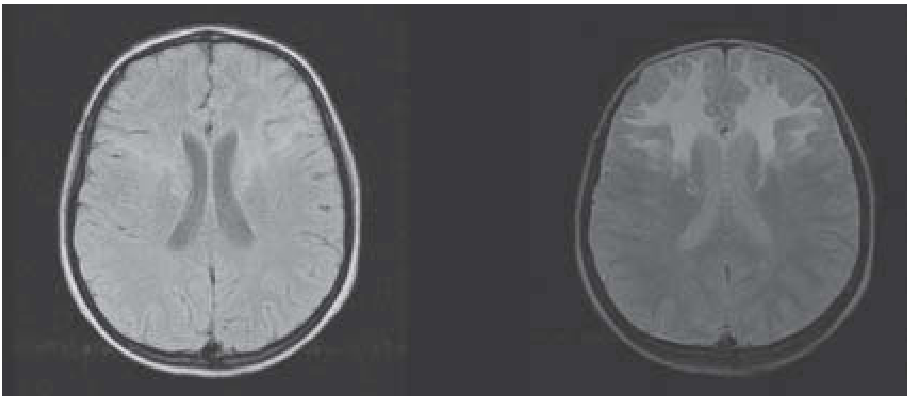


Figure 7.1. *Example of an MRI section of the brain acquired with two echo times (Saint-Vincent de Paul hospital, radiology service, Professor Catherine Adamsbaum). The pathological area corresponds to the whiter areas in the upper part of the image*

These two images constitute a typical example to illustrate the fusion by belief function theory. We will only present here the results obtained for three classes ($C_1 =$ WM + GM, $C_2 =$ V + CSF and $C_3 =$ ALD).

The definition of the focal elements is supervised using a reasoning method that takes into account the knowledge available and the characteristics of the image with respect to the classes we are focusing on. For the example described here, the focal elements of the mass function m_1 assigned to the first image are C_2, $C_1 \cup C_3$, since C_1 and C_3 are not well discriminated on this image. Zero mass functions are assigned to the other composite hypotheses, since the corresponding classes cannot be confused. On the second image, it is on the other hand difficult to separate the brain from the ventricles and therefore the focal elements of m_2 are C_3 and $C_1 \cup C_2$. We will discuss later the introduction of overall absence of knowledge and of a mass explicitly representing the partial volume. The mass functions are chosen with a simple trapezoidal shape, whose parameters are automatically determined on the histograms [BLO 97b]. This is a crude model, but it has proven to be sufficient for this application. The functions are then normalized so as to satisfy the normalization constraint $\sum_{A \subset D} m(A) = 1$. With this model, the classification is performed only based on the gray levels and the fusion is performed on the pixel level, therefore without spatial information.

Conjunctive combination by Dempster's rule only provides focal elements which are the singletons C_1, C_2, C_3. The conflict is not equal to zero in this case.

The last phase is the decision making. Always making a decision in favor of a simple hypothesis forces us in fact to always making a clear decision, which is not adapted to all of the actual situations in medical imaging, where pixels can belong to a union of classes but also to none strictly. However, because $\mathrm{Bel}(A) \geq \mathrm{Bel}(C_i)$ for any $C_i \in A$ and because $\mathrm{Bel}(D) \geq \mathrm{Bel}(A)$, a certain number of precautions have to be taken in order to decide in favor of a composite hypothesis. We can, for example, imagine making a decision in favor of a composite hypothesis if the arguments involving the simple hypotheses are not strong enough.

Thus, in Figure 7.2, the decision was made according to the maximum belief over all of the hypotheses except D. Let us note that, in this simple case, the maximum credibility is equivalent to the maximum plausibility, since $m_1 \oplus m_2$ is a Bayesian mass function. With this rule, the decision is made in favor of a simple hypothesis in the points where other masses are equal to zero and in favor of a composite hypothesis otherwise. This way, we obtain interesting results since the partial volume points are detected as a composite hypothesis, whereas the areas without ambiguity are well segmented. Figure 7.3 shows the results obtained by making a decision in favor of a simple hypothesis with the maximum belief.

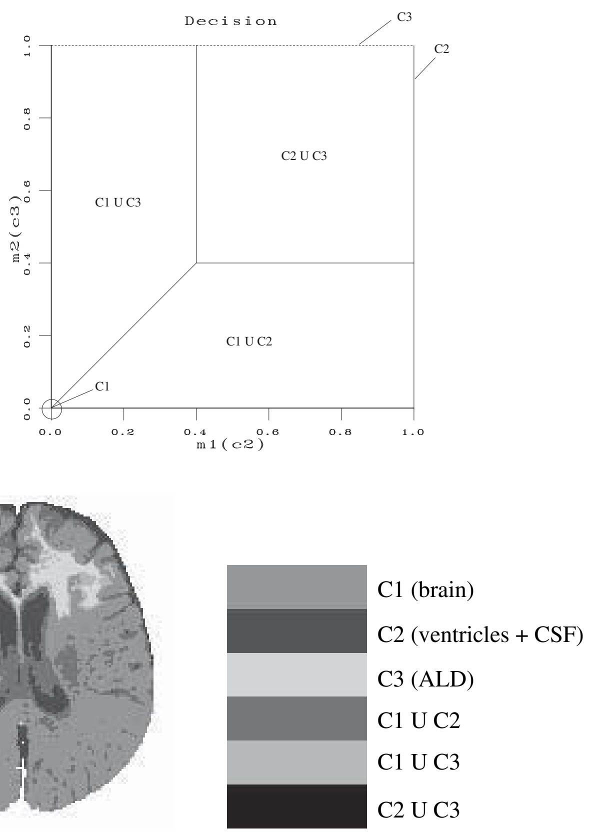


Figure 7.2. *The different decision areas depending on the values of $m_1(C_2)$ and $m_2(C_3)$ and the decision image, by taking the maximum belief over all of the hypotheses except D*

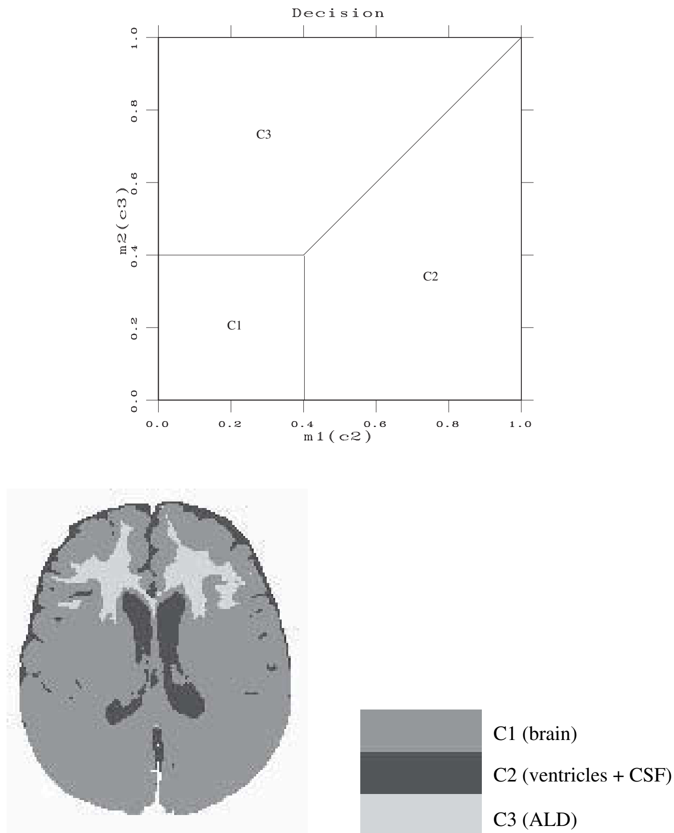


Figure 7.3. *The different decision zones depending on the values of $m_1(C_2)$ and $m_2(C_3)$ and the decision image, by taking the maximum belief over the simple hypotheses*

Studying the influence of the weighting of $m_1(C_1 \cup C_3)$ and $m_2(C_1 \cup C_2)$, relative to $m_1(C_2)$ and $m_2(C_3)$, has shown that the conflict calculation does not present a systematic evolution towards an increase or a decrease of conflict. For example, if $m_2(C_3)$ is small, the conflict is reduced and the weight attributed to $m_1(C_2)$ increases (in this case, the two sources are in better agreement over C_2, i.e. the ventricle class). The calculation of the decision areas (as in Figures 7.2 and 7.3) shows that if the weights of $m_2(C_3)$ and $m_1(C_2)$ increase, the decision area in favor of C_1 decreases (when the decisions are made based on simple hypotheses). This can also be observed in decision images: Figure 7.4 shows a close-up of the decision images in which the ventricles and the CSF are better detected if the weight for $m_2(C_3)$ and $m_1(C_2)$ increases. Differences are also apparent on the small ALD branches which are better detected as well. It is very important to note that the areas that are different are classified as favoring the $C_1 \cup C_2$ hypothesis (i.e. brain or ventricles and CSF) when the decision is taken on all the hypotheses except D (in favor of the hypothesis $C_1 \cup C_3$ for the ALD branches, respectively). With this decision rule, there is practically no difference in the decision images, showing the robustness with respect to weighting. This supports the idea that composite hypotheses should not be so often ignored.

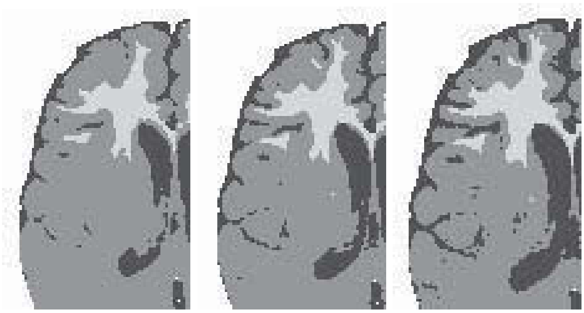


Figure 7.4. *Decision by maximum belief based on simple hypotheses, for an increasing weight on $m_2(C_3)$ and $m_1(C_2)$*

Let us now examine the influence of overall absence of knowledge, modeled by a mass on D added by discounting, according to reliability coefficients for each source. This time, the result of the combination $m_1 \oplus m_2$ is no longer a Bayesian mass function and the decision by maximum belief is no longer equivalent to the maximum plausibility. In fact, we do observe small differences in the decision images. We have observed only very slight differences when the decision is made based on simple hypotheses. However, decisions on all of the hypotheses except D are always made in favor of a composite hypothesis, as expected. In our application, there is no particular argument for suggesting that one image is more or less reliable than another (overall because this is not true on a class per class basis). This is why it is not very useful to assign a mass to $m(D)$ and this is confirmed by the results. On the other hand, we have strong arguments to support partial absence of knowledge, depending on the image, which leads to ambiguities between classes and this was included by masses for the composite hypotheses. Once again, this justifies the method used for assigning masses, which relies on how the problem is modeled, rather than the traditional method based on probabilities and overall reliability factors.

The images used in this example were acquired with rather thick sections, thus causing a strong partial volume effect, particularly between white matter and the ALD (see Figure 7.1). We will now include this knowledge explicitly as a mass on $C_1 \cup C_3$ in the second image (trapezoidal function inferred from the histogram). This reduces conflict. The combination $m_1 \oplus m_2$ is no longer a Bayesian mass function and the decision areas are modified, as shown in Figure 7.5.

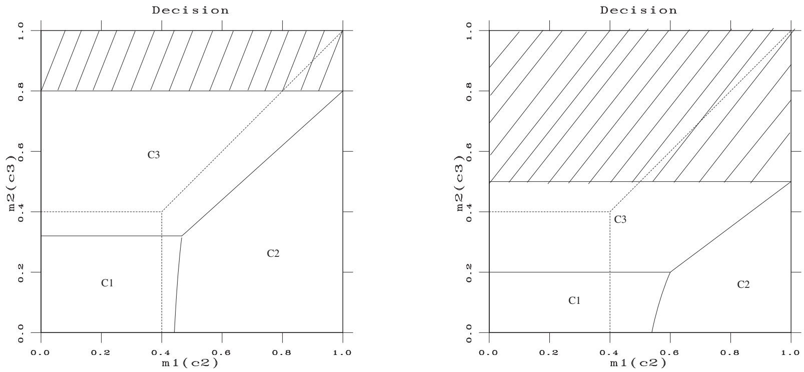


Figure 7.5. *Decision areas depending on the values of $m_1(C_2)$ and $m_2(C_3)$ by including an increasing mass on $C_1 \cup C_3$ for the second image ($m_2(C_3)$ which has to be smaller than $1 - m_2(C_1 \cup C_3)$). The dotted lines represent the previous limits (see Figure 7.3)*

The decision images in Figure 7.6 respectively show the results obtained for all of the hypotheses except D and for all of the simple hypotheses only, first with $m_2(C_1 \cup C_3) = 0$ and then with an increasing weight assigned to $m_2(C_1 \cup C_3)$. This figure shows that the decision for all of the hypotheses includes all of the partial volume areas between WM and the ALD in $C_1 \cup C_3$, and does not change if the weights of $m_2(C_1 \cup C_3)$ increase, which is another indication of the robustness with respect to weighting. On the contrary, the decision images for the simple hypotheses only show an increasing number of partial volume points that are included in the ALD. This modeling makes it possible to imitate how a doctor would make his decision, based on his objective. In the image farther to the left, where the partial volume is not taken into account, the area classified as ALD presents no ambiguity (and corresponds to "pure" ALD, without mixture), whereas on the image farther to the right, all of the partial volume is included in the ALD (this corresponds to the actual segmentation manually obtained by doctors) and the classified areas of the brain contain no ambiguous parts.

We have tried here to illustrate a few of the characteristics of belief function theory that can be used in image fusion for classification, segmentation or recognition and that constitute advantages compared to the traditional probabilistic and Bayesian methods. They reflect the high flexibility of possible models, taking into account at the same time uncertainty and imprecision, partial or overall absence of knowledge, the reliabilities of the sources, the ability of each source to provide reliable or unreliable information on each class, *a priori* information it may be impossible to represent using probabilities, etc. The application presented here is a good illustration of these various advantages. First of all, a model that is well suited to the problem is possible,

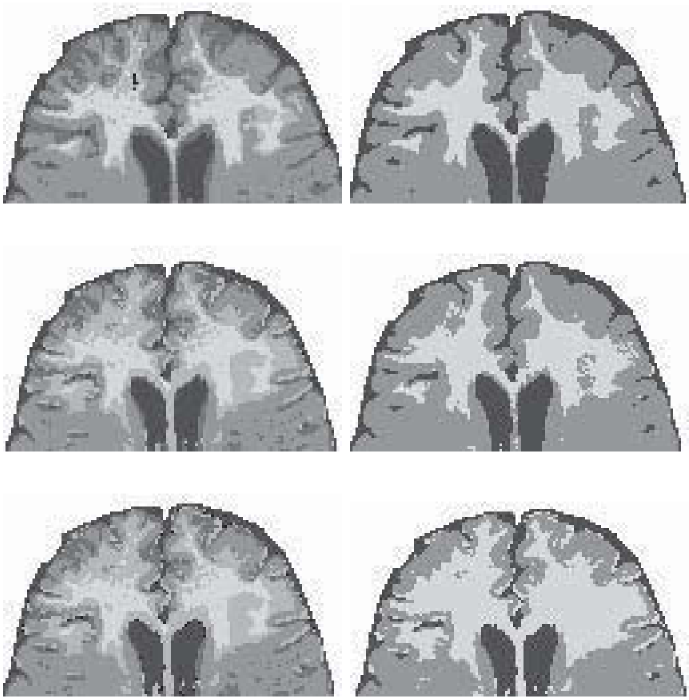


Figure 7.6. *Decision by maximum belief for all of the hypotheses except D (first column) and for all of the simple hypotheses (second column), without any mass on $C_1 \cup C_2$ (top) and by including a mass on $C_1 \cup C_2$ (representing the partial volume effect between the brain the ALD) in the second image, with an increasing weight (middle and bottom)*

particularly by assigning masses to the composite hypotheses, expressing, for example, the fact that a source does not make it possible to correctly differentiate between two classes or even modeling the partial volume effect. The probabilities are not well adapted to the modeling of partial volume at the limit between two classes. One solution is suggested here, by assigning a mass directly to the union of these classes, which again leads to a satisfactory interpretation. Even a rather crude definition of the mass functions and of their relative weights turned out to be sufficient and robust. Finally, the decision was made according to two rules: a traditional rule where the decision is always made in favor of a simple hypothesis and a second rule where it is also possible to decide in favor of a composite hypothesis. This latter rule is closer to what happens in reality, by highlighting the partial volume areas, and by adapting itself to the doctor's reasoning mode.

7.8. Bibliography

[AND 88] ANDRESS K.M., KAK A.C., "Evidence Accumulation and Flow Control in a Hierarchical Spatial Reasoning System", *AI Magazine*, p. 75-94, 1988.

[APP 91] APPRIOU A., "Probabilités et incertitude en fusion de données multi-senseurs", *Revue Scientifique et Technique de la Défense*, no. 11, p. 27-40, 1991.

[APP 93] APPRIOU A., "Formulation et traitement de l'incertain en analyse multi-senseurs", *Quatorzième Colloque GRETSI*, Juan-les-Pins, p. 951-954, 1993.

[BAL 92] BALDWIN J.F., "Inference for Information Systems Containing Probabilistic and Fuzzy Uncertainties", in L. ZADEH and J. KACPRZYK (ed.) *Fuzzy Logic and the Management of Uncertainty*, p. 353-375, J. Wiley, New York, 1992.

[BAR 81] BARNETT J.A., "Computational Methods for a Mathematical Theory of Evidence", *Proc. of 7th IJCAI*, Vancouver, p. 868-875, 1981.

[BLO 95] BLOCH I., "Fondements des probabilités et des croyances: une discussion des travaux de Cox et Smets", *15th GRETSI Symposium*, Juan-les-Pins, France, p. 909-912, September 1995.

[BLO 96] BLOCH I., "Some Aspects of Dempster-Shafer Evidence Theory for Classification of Multi-Modality Medical Images Taking Partial Volume Effect into Account", *Pattern Recognition Letters*, vol. 17, no. 8, p. 905-919, 1996.

[BLO 97a] BLOCH I., "Using Fuzzy Mathematical Morphology in the Dempster-Shafer Framework for Image Fusion under Imprecision", *IFSA'97*, Prague, p. 209-214, 1997.

[BLO 97b] BLOCH I., AURDAL L., BIJNO D., MÜLLER J., "Estimation of Class Membership Functions for Grey-Level Based Image Fusion", *IEEE ICIP'97*, vol. III, Santa Barbara, California, p. 268-271, October 1997.

[CHE 93] CHEN S.Y., LIN W.C., CHEN C.T., "Evidential Reasoning based on Dempster-Shafer Theory and its Application to Medical Image Analysis", *SPIE*, vol. 2032, p. 35-46, 1993.

[CLE 91] VAN CLEYNENBREUGEL J., OSINGA S.A., FIERENS F., SUETENS P., OOSTERLINCK A., "Road Extraction from Multi-temporal Satellite Images by an Evidential Reasoning Approach", *Pattern Recognition Letters*, vol. 12, p. 371-380, 1991.

[COX 46] COX R.T., "Probability, Frequency and Reasonable Expectation", *Journal of Physics*, vol. 14, no. 1, p. 115-137, 1946.

[CUC 92] CUCKA P., ROSENFELD A., Evidence-based Pattern Matching Relaxation, Technical Report no. CAR-TR-623, Center of Automation Research, University of Maryland, May 1992.

[DEN 95] DENŒUX T., "A k-nearest Neighbor Classification Rule based on Dempster-Shafer Theory", *IEEE Transactions on Systems, Man and Cybernetics*, vol. 25, no. 5, p. 804-813, 1995.

[DEN 06] DENŒUX T., "Construction of belief functions from sample data using multimodal confidence regions", *International Journal of Approximate Reasoning*, vol. 42, p. 228-252, 2006.

[DOK 00] DOKLÁDAL P., Grey-Scale Image Segmentation: A Topological Approach, PhD Thesis, Marne la Vallée University, 2000.

[DRO 97] DROMIGNY-BADIN A., ROSSATO S., ZHU Y.M., "Fusion de données radioscopiques et ultrasonores *via* la théorie de l'évidence", *Traitement du Signal*, vol. 14, no. 5, p. 147-160, 1997.

[DUB 88] DUBOIS D., PRADE H., "Representation and Combination of Uncertainty with Belief Functions and Possibility Measures", *Computational Intelligence*, vol. 4, p. 244-264, 1988.

[DUB 99] DUBOIS D., PRADE H., YAGER R., "Merging Fuzzy Information", in J. BEZDEK, D. DUBOIS, H. PRADE (ed.) *Handbook of Fuzzy Sets Series, Approximate Reasoning and Information Systems*, Chapter 6, Kluwer, 1999.

[GAN 96] GANSTER H., PINZ A., "Active Fusion using Dempster-Shafer Theory of Evidence", in A. PINZ (ed.) *Proceedings of 20th ÖAGM / AAPR Workshop*, Schriftenreihe der OCG, Oldenbourg, 1996.

[GAR 86] GARVEY T.D., "Evidential Reasoning for Land-Use Classification", *Analytical Methods in Remote Sensing for Geographic Information Systems, International Association of Pattern Recognition, Technical Committee 7 Workshop*, Paris, 1986.

[GOR 85] GORDON J., SHORTLIFFE E.H., "A Method for Managing Evidential Reasoning in a Hierarchical Hypothesis Space", *Artificial Intelligence*, vol. 26, p. 323-357, 1985.

[GUA 91] GUAN J., BELL D.A., *Evidence Theory and its Applications*, North-Holland, Amsterdam, 1991.

[LEE 87] LEE T., RICHARDS J.A., SWAIN P.H., "Probabilistic and Evidential Approaches for Multisource Data Analysis", *IEEE Transactions on Geoscience and Remote Sensing*, vol. GE-25, no. 3, p. 283-293, 1987.

[LOW 91] LOWRANCE J.D., STRAT T.M., WESLEY L.P., GARVEY T.D., RUSPINI E.H., WILKINS D.E., The Theory, Implementation and Practice of Evidential Reasoning, SRI project 5701 final report, SRI, Palo Alto, June 1991.

[MAS 97] MASCLE S., BLOCH I., VIDAL-MADJAR D., "Application of Dempster-Shafer Evidence Theory to Unsupervised Classification in Multisource Remote Sensing", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, p. 1018-1031, 1997.

[MEN 96] MÉNARD M., ZAHZAH E.H., SHAHIN A., "Mass Function Assessment: Case of Multiple Hypotheses for the Evidential Approach", *Europto Conf. on Image and Signal Processing for Remote Sensing*, Taormina, Italy, 1996.

[MER 05] MERCIER D., CRON G., DENOEUX T., MASSON M., "Fusion of Multi-level Decision Systems Using the Transferable Belief Model", *8th International Conference on Information Fusion, Fusion 2005*, Philadelphia, p. 655-658, 2005.

[MIL 00] MILISAVLJEVIĆ N., BLOCH I., ACHEROY M., "Characterization of Mine Detection Sensors in Terms of Belief Functions and their Fusion, First Results", *International Conference on Informatin Fusion - Fusion 2000*, vol. II, Paris, p. ThC3 15-22, 2000.

[MIL 01] MILISAVLJEVIĆ N., BLOCH I., “A Two-Level Approach for Modeling and Fusion of Humanitarian Mine Detection Sensors within the Belief Function Framework”, *Applied Stochastic Models and Data Analysis*, vol. 2, Compiègne, p. 743-748, 2001.

[NEA 92] NEAPOLITAN R.E., “A Survey of Uncertain and Approximate Inference”, in L. ZADEH and J. KAPRZYK (ed.) *Fuzzy Logic for the Management of Uncertainty*, p. 55-82, J. Wiley, New York, 1992.

[PIN 95] PINZ A., PRANTL M., “Active Fusion for Remote Sensing Image Understanding”, *European Symposium on Satellite Remote Sensing, Paris*, vol. 2579, EOS/SPIE, p. 67-77, 1995.

[QUI 89] QUINIO P., Representation and Accumulation of Uncertain Informations: A Theoretical Comparison of Probabilistic and some Non-Probabilistic Formalisms, Report, Ito Lab., Tohoku University, 1989.

[QUI 91] QUINIO P., MATSUYAMA T., “Random Closed Sets: A Unified Approach to the Representation of Imprecision and Uncertainty”, in R. KRUSE and P. SIEGEL (ed.) *Symbolic and Quantitative Approaches to Uncertainty, ECSQARU*, Marseille, Springer Verlag, p. 282-286, 1991.

[RAS 90] RASOULIAN H., THOMPSON W.E., KAZDA L.F., PARRA-LOERA R., “Application of the Mathematical Theory of Evidence to the Image Cueing and Image Segmentation Problem”, *SPIE Signal and Image Processing Systems Performance Evaluation*, vol. 1310, p. 199-206, 1990.

[ROM 99] ROMBAUT M., “Fusion de données images segmentées à l’aide du formalisme de Dempster-Shafer”, *GRETSI’99*, Vannes, France, p. 655-658, 1999.

[SCH 93] SCHUBERT J., “On Nonspecific Evidence”, *International Journal of Intelligent Systems*, vol. 8, p. 711-725, 1993.

[SHA 76] SHAFER G., *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[SME 90a] SMETS P., “The Combination of Evidence in the Transferable Belief Model”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, p. 447-458, 1990.

[SME 90b] SMETS P., “Constructing the Pignistic Probability Function in a Context of Uncertainty”, *Uncertainty in Artificial Intelligence*, vol. 5, p. 29-39, 1990.

[SME 93] SMETS P., “Belief Functions: The Disjunctive Rule of Combination and the Generalized Bayesian Theorem”, *International Journal of Approximate Reasoning*, vol. 9, p. 1-35, 1993.

[TUP 99] TUPIN F., BLOCH I., MAÎTRE H., “A First Step Towards Automatic Interpretation of SAR Images using Evidential Fusion of Several Structure Detectors”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, p. 1327-1343, 1999.

[YAG 87] YAGER R.R., “On the Dempster-Shafer Framework and New Combination Rules”, *Information Science*, vol. 41, p. 93-137, 1987.

[ZAH 92] ZAHZAH E., Contribution à la représentation des connaissances et à leur utilisation pour l’interprétation automatique des images satellites, PhD Thesis, Paul Sabatier University, Toulouse, 1992.

This page intentionally left blank

Chapter 8

Fuzzy Sets and Possibility Theory

8.1. Introduction and general concepts

As we have seen in the first chapters of this book, imprecisions and uncertainties are inherent to the data handled in the application fields that concern us.

The advantages of fuzzy sets and possibility theory for information processing, particularly in image and vision [KRI 92], fall into the four following categories:

– the ability of fuzzy sets to represent spatial information in images as well as its imprecision, on several levels (local, regional, or global) and in different forms (numerical, symbolic, quantitative, qualitative);

– the possibility of representing very heterogenous information, directly extracted from images or obtained from outside knowledge, such as expert or generic knowledge in a field or about a problem;

– the possibility of generalizing to fuzzy sets operations for manipulating spatial information;

– the various possible semantics;

– the flexibility of the combination operators, which makes it possible to fuse elements of information that are different in nature, in very different situations.

We will particularly insist on this last point.

In this chapter, we will first of all present the basic elements of fuzzy set and possibility theory. Their use in the more specific context of fusion will be discussed

Chapter written by Isabelle BLOCH.

later. This theory was introduced by Zadeh and the first article on the subject dates back to 1965 [ZAD 65]. See [DUB 80, KAU 75, ZIM 91] which contain most of the theory.

8.2. Definitions of the fundamental concepts of fuzzy sets

8.2.1. *Fuzzy sets*

Let $\mathcal{S}$ be the universe or the space of reference. This is a classical (or crisp) set. Its elements will be denoted by x, y, etc. In image processing, $\mathcal{S}$ will typically be the space in which the image is defined ($\mathbb{Z}^n$ or $\mathbb{R}^n$, with $n = 2, 3$, etc.). The elements of $\mathcal{S}$ are then the points of the image (pixels, voxels). The universe can also be a set of values taken from characteristics of the image such as the scale of gray levels. The elements are then values (gray levels). The set $\mathcal{S}$ can also be a set of primitives or objects extracted from the images (segments, areas, objects, etc.) in a representation on a higher level of image content.

A subset X of $\mathcal{S}$ is defined by its characteristic function μ_X, such that:

$$\mu_X(x) = \begin{cases} 1 & \text{if } x \in X \\ 0 & \text{if } x \notin X \end{cases} \qquad [8.1]$$

The characteristic function μ_X is a binary function, specifying for each point of $\mathcal{S}$ whether it belongs to X.

Fuzzy set theory deals with gradual membership. A fuzzy subset of $\mathcal{S}$ is defined by its membership function μ of $\mathcal{S}$ in $[0, 1]$[1]. For any x of $\mathcal{S}$, $\mu(x)$ is the value in $[0, 1]$ that represents the degree to which x is a member of the fuzzy subset (often referred to simply as "fuzzy set").

Different notations are often used to refer to a fuzzy set. The set $\{(x, \mu(x)), x \in X\}$, completely defines the fuzzy set and is sometimes denoted by $\int_{\mathcal{S}} \mu(x)/x$, or in the discrete finite case $\sum_{i=1}^{N} \mu(x_i)/x_i$ where N indicates the number of elements in $\mathcal{S}$.

Since knowing the set of all of the pairs $(x, \mu(x))$ is completely equivalent to having the definition of the membership function μ, from now on we will simplify the notations and use the functional notation μ (a function of $\mathcal{S}$ into $[0, 1]$) to refer to both the fuzzy set and its membership function. We denote by $\mathcal{F}$ the set of fuzzy sets defined on $\mathcal{S}$.

1. The interval $[0, 1]$ is the most commonly used, but any interval or any other set (a lattice, typically) can be used.

The support of a fuzzy set μ is the set of points with strictly positive membership to μ (it is a binary set):

$$\operatorname{Supp}(\mu) = \{x \in \mathcal{S},\ \mu(x) > 0\}. \qquad [8.2]$$

The core of a fuzzy set μ is the set of points that are completely included in μ (it is also a binary set):

$$\operatorname{Core}(\mu) = \{x \in \mathcal{S},\ \mu(x) = 1\}. \qquad [8.3]$$

A fuzzy set μ is said to be normalized if at least one point is completely included in the set μ (which is equivalent to $\operatorname{Core}(\mu) \neq \emptyset$):

$$\exists x \in \mathcal{S},\ \mu(x) = 1. \qquad [8.4]$$

A fuzzy set μ is described as unimodal if there is only one point x such that $\mu(x) = 1$. A less restrictive definition allows the core to be a compact set, not necessarily reduced to a point.

8.2.2. *Set operations: Zadeh's original definitions*

Since fuzzy sets were introduced to generalize the concept of sets, the first operations defined were set operations. In this section, we will present Zadeh's original definitions [ZAD 65]. More general classes of operations will be presented in section 8.5.

The equality of two fuzzy sets is defined by the equality of their membership functions:

$$\mu = \nu \Longleftrightarrow \forall x \in \mathcal{S},\ \mu(x) = \nu(x). \qquad [8.5]$$

The inclusion of a fuzzy set in another is defined by an inequality between the two membership functions:

$$\mu \subseteq \nu \Longleftrightarrow \forall x \in \mathcal{S},\ \mu(x) \leq \nu(x). \qquad [8.6]$$

The equality of μ and ν is of course equivalent to having the inclusion in both directions.

Let us note that these concepts lead to a binary result. It is also possible to define a degree of inclusion between two fuzzy subsets, but we will not discuss this here.

The intersection (the union, respectively) of two fuzzy subsets is defined by the point-to-point minimum (maximum, respectively) between the membership functions:

$$\forall x \in \mathcal{S},\ (\mu \cap \nu)(x) = \min\big[\mu(x), \nu(x)\big], \qquad [8.7]$$

$$\forall x \in \mathcal{S},\ (\mu \cup \nu)(x) = \max\big[\mu(x), \nu(x)\big]. \qquad [8.8]$$

The complement of a fuzzy set is defined by:

$$\forall x \in \mathcal{S},\ \mu^C(x) = 1 - \mu(x). \qquad [8.9]$$

Here are the major properties of these operations:

– they are all consistent with set operations: in the specific case where the membership functions only have 0 and 1 as values (sets are then binary), these definitions amount to the traditional binary definitions (this is an important property and the least we would expect from the fuzzy extension of a binary operation);

– $\mu = \nu \Leftrightarrow \mu \subseteq \nu$ and $\nu \subseteq \mu$;

– fuzzy complementation is involutive: $(\mu^C)^C = \mu$;

– intersection and union are commutative and associative;

– intersection and union are idempotent and mutually distributive;

– intersection and union are dual with respect to complementation: $(\mu \cap \nu)^C = \mu^C \cup \nu^C$;

– if we consider that the empty set $\emptyset$ is a fuzzy set with an identically zero membership function, then we have $\mu \cap \emptyset = \emptyset$ and $\mu \cup \emptyset = \mu$, for any fuzzy set μ defined in $\mathcal{S}$;

– if we consider that the universe is a fuzzy set with a membership function equal to 1, then we have $\mu \cap \mathcal{S} = \mu$ and $\mu \cup \mathcal{S} = \mathcal{S}$, for any fuzzy set μ defined in $\mathcal{S}$.

These properties are the same as the corresponding binary operations. However, some properties that are true in the binary case are lost in the fuzzy case, such as the law of the excluded middle ($X \cup X^C = \mathcal{S}$) and the law of non-contradiction ($X \cap X^C = \emptyset$). This is because we have, in general:

$$\mu \cup \mu^C \neq \mathcal{S}, \qquad [8.10]$$

$$\mu \cap \mu^C \neq \emptyset. \qquad [8.11]$$

8.2.3. α-*cuts*

The α-cut of a fuzzy set μ is the binary set defined by:

$$\mu_\alpha = \big\{x \in \mathcal{S},\ \mu(x) \geq \alpha\big\}. \quad [8.12]$$

Strict (or strong) α-cuts are defined by:

$$\mu_{\alpha_S} = \big\{x \in \mathcal{S},\ \mu(x) > \alpha\big\}.$$

A fuzzy set can be interpreted as its α-cuts stacked on top of each other. It can be reconstructed from them using several formulae, the most common of which are:

$$\mu(x) = \int_0^1 \mu_\alpha(x) d\alpha,$$

$$\mu(x) = \sup_{\alpha \in]0,1]} \min\big(\alpha, \mu_\alpha(x)\big),$$

$$\mu(x) = \sup_{\alpha \in]0,1]} \big(\alpha \mu_\alpha(x)\big).$$

Most of the operations we have defined so far commute with α-cuts. More precisely, we have the following relations:

$$\forall(\mu,\nu) \in \mathcal{F}^2,\ \mu = \nu \Longleftrightarrow \forall\alpha \in]0,1],\ \mu_\alpha = \nu_\alpha,$$

$$\forall(\mu,\nu) \in \mathcal{F}^2,\ \mu \subseteq \nu \Longleftrightarrow \forall\alpha \in]0,1],\ \mu_\alpha \subseteq \nu_\alpha,$$

$$\forall(\mu,\nu) \in \mathcal{F}^2,\ \forall\alpha \in [0,1],\ (\mu \cap \nu)_\alpha = \mu_\alpha \cap \nu_\alpha,$$

$$\forall(\mu,\nu) \in \mathcal{F}^2,\ \forall\alpha \in [0,1],\ (\mu \cup \nu)_\alpha = \mu_\alpha \cup \nu_\alpha,$$

$$\forall\mu \in \mathcal{F},\ \forall\alpha \in [0,1],\ \big(\mu^C\big)_\alpha = \big(\mu_{1-\alpha_S}\big)^C.$$

Choosing an α-cut in a fuzzy set is equivalent to thresholding the membership function in order to select the points with a level of membership of at least α. This operation can be interpreted as a "defuzzification" and is used in the decision phases after the fusion.

8.2.4. *Cardinality*

In this section, we will restrict ourselves to fuzzy sets defined over a finite domain, or that have a finite support (this will always be the case in image applications).

The cardinality of a fuzzy set μ is defined by:

$$|\mu| = \sum_{x \in \mathcal{S}} \mu(x),$$

or, if only the support of μ is finite:

$$|\mu| = \sum_{x \in \mathrm{Supp}(\mu)} \mu(x).$$

This definition is consistent with the traditional concept of the cardinality of a binary set. In the case of a fuzzy set, each point counts as an amount equal to its membership level. The cardinality is also referred to as the power of the fuzzy set (for example, [LUC 72]).

This definition can be extended to the case where $\mathcal{S}$ is not finite but is measurable. Let M be a measure $\mathcal{S}$ (such that $\int_{\mathcal{S}} dM(x) = 1$). The cardinality of μ is then defined by:

$$|\mu| = \int_{\mathcal{S}} \mu(x) dM(x).$$

8.2.5. *Fuzzy number*

In this section, we will assume that $\mathcal{S} = \mathbb{R}$.

A fuzzy quantity is a fuzzy set μ in $\mathbb{R}$. A fuzzy interval is a convex fuzzy quantity (all of its α-cuts are intervals). The upper semi-continuity of μ is equivalent to the fact that the α-cuts are closed intervals.

A fuzzy number is an upper semi-continuous (u.s.c.) interval with a compact and unimodal support. An example of a fuzzy number representing "roughly 10" is shown in Figure 8.1.

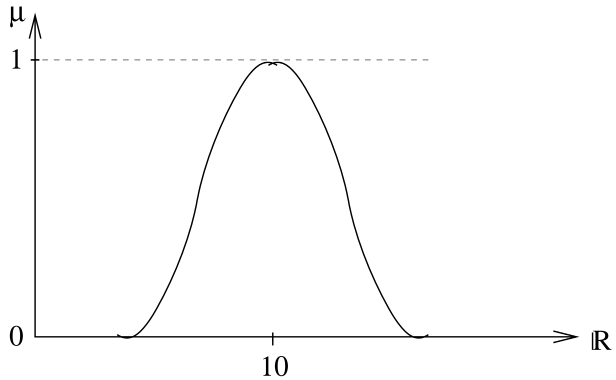


Figure 8.1. *Fuzzy number representing "roughly 10"*

We can also find less stringent definitions, particularly if we accept an interval of modal values, i.e. if there are four real numbers a, b, c, d, with $a \leq b \leq c \leq d$ such that $\mu(x) = 0$ outside the interval $[a, d]$, μ is non-decreasing on $[a, b]$, non-increasing on $[c, d]$ and equal to 1 on $[b, c]$ [GOE 83, GOE 86].

A fuzzy number can be interpreted as a flexible representation of an imprecise quantity, which is a more general representation than the traditional interval.

We now turn again to the concept of the cardinality of a fuzzy set, which is defined above as a number. If the set is not well defined, we can expect for any measure of this set to be imprecise as well, particularly its cardinality, which should therefore be defined as a fuzzy number [DUB 80]:

$$|\mu|_f(n) = \sup \big\{\alpha \in [0,1],\ |\mu_\alpha| = n\big\}.$$

The quantity $|\mu|_f(n)$ represents the degree to which the cardinality of μ is equal to n.

A very common class of fuzzy number is comprised of the L-R fuzzy numbers. They are defined by a parametric representation of their membership function:

$$\forall x \in \mathbb{R},\ \mu(x) = \begin{cases} L\left(\dfrac{m-x}{\alpha}\right) & \text{if } x \leq m \\ R\left(\dfrac{x-m}{\beta}\right) & \text{if } x \geq m \end{cases}$$

where α and β are strictly positive numbers referred to as left and right spreads, m is a number referred to as the mean value, and L and R are functions with the following properties:

– $\forall x \in \mathbb{R},\ L(x) = L(-x)$;

– $L(0) = 1$;

– L is non-increasing on $[0, +\infty[$.

The function R has similar properties.

One of the main advantages of these fuzzy numbers is their compact representation, which allows simple calculations.

In fusion, fuzzy numbers are often used for representing knowledge about measurements or observations, or flexible constraints applied to the values they can be equal to, for example: “the gray level of this structure is roughly equal to 10”. Elements of knowledge like this one can then be fused with the data or with other elements of knowledge.

8.3. Fuzzy measures

The definitions and a few examples of fuzzy measures are presented here. For a more detailed presentation, see [DUB 80, SUG 74].

8.3.1. *Fuzzy measure of a crisp set*

A fuzzy measure is a function f from $\mathcal{C}$, which is the set of subsets of $\mathcal{S}$ (hence defined on crisp sets), into $[0,1]$ that satisfies the following conditions:

– $f(\emptyset) = 0$;

– $f(\mathcal{S}) = 1$;

– monotonicity: $\forall (A,B) \in \mathcal{C}^2$, $A \subseteq B \Rightarrow f(A) \leq f(B)$;

– continuity:

$$\forall i \in \mathbb{N},\ \forall A_i \in \mathcal{C},\ A_1 \subseteq A_2 \cdots \subseteq A_n \cdots \text{ or } A_1 \supseteq A_2 \cdots \supseteq A_n \cdots$$
$$\Longrightarrow \lim_{i\to\infty} f\big(A_i\big) = f\Big(\lim_{i\to\infty} A_i\Big).$$

Some notable properties of fuzzy measures are:

$$\forall (A,B) \in \mathcal{C}^2,\ f(A \cup B) \geq \max\big[f(A), f(B)\big], \qquad [8.13]$$

$$\forall (A,B) \in \mathcal{C}^2,\ f(A \cap B) \leq \min\big[f(A), f(B)\big]. \qquad [8.14]$$

This definition assumes no additivity constraint. They could simply be called non-additive measures, since the link with fuzzy set theory which was presented earlier is relatively weak.

8.3.2. *Examples of fuzzy measures*

Several families of fuzzy measures can be found in other works, the most common of which are:

– probability measures;

– fuzzy λ-measures, obtained by relaxing the additivity constraint for probability measures:

$$\forall (A,B) \in \mathcal{C}^2,\ A \cap B = \emptyset \Longrightarrow f(A \cup B) = f(A) + f(B) + \lambda f(A) f(B) \qquad [8.15]$$

with $\lambda > -1$;

– belief and plausibility function in belief function theory [SHA 76] (see Chapter 7);

– possibility measures [ZAD 78], which will be introduced in section 8.4.

The links between the various fuzzy measures can be found in [BAN 78] or in [DUB 80].

8.3.3. *Fuzzy integrals*

Fuzzy integrals [GRA 92, SUG 74] are the counterpart of Lebesgue integrals when the integration is performed with respect to a fuzzy measurement. Fuzzy integrals can be divided into two types.

The Sugeno integral of a measurable function f, defined from $\mathcal{S}$ into $[0,1]$, with respect to a fuzzy measure μ is defined by:

$$S_\mu(f) = \int f \circ \mu = \sup_{\alpha \in [0,1]} \min \big[\alpha, \mu\big(\{x \in \mathcal{S}, f(x) > \alpha\}\big)\big].$$

In the finite case ($|\mathcal{S}| = N$), this expression is equivalent to:

$$S_\mu(f) = \int f \circ \mu = \max_{i=1}^{N} \min \big[f\big(x_{p(i)}\big), \mu\big(A_i\big)\big],$$

where p is a permutation of $\{1, 2 \ldots N\}$ such that:

$$0 \leq f\big(x_{p(1)}\big) \leq \cdots \leq f\big(x_{p(N)}\big)$$

and where $A_i = \{x_{p(1)}, \ldots, x_{p(N)}\}$.

The Choquet integral of a measurable function f, defined from $\mathcal{S}$ into $\mathbb{R}^+$, with respect to a fuzzy measurement μ is defined by:

$$C_\mu(f) = \int f d\mu = \int_0^{+\infty} \mu\big(\{x, f(x) > \alpha\}\big) d\alpha.$$

In the finite case, we get:

$$C_\mu(f) = \int f d\mu = \sum_{i=1}^{N} \big[f\big(x_{p(i)}\big) - f\big(x_{p(i-1)}\big)\big] \mu\big(A_i\big),$$

with $f(x_p(0)) = 0$.

The properties of these integrals are presented in detail in [GRA 92, MUR 89, SUG 74]. The major ones in the finite case are the following:

– for the measure $\mu_{\min}$ defined by $\forall A \subset \mathcal{S},\ A \neq \mathcal{S},\ \mu_{\min}(A) = 0$ and $\mu_{\min}(\mathcal{S}) = 1$, the integrals $S_{\mu_{\min}}(f)$ and $C_{\mu_{\min}}(f)$ are equal to the minimum of the values taken by f;

– for the measure $\mu_{\max}$ defined by $\forall A \subseteq \mathcal{S},\ A \neq \emptyset,\ \mu_{\max}(A) = 1$ and $\mu_{\max}(\emptyset) = 0$, the integrals $S_{\mu_{\max}}(f)$ and $C_{\mu_{\max}}(f)$ are equal to the maximum of the values taken by f;

– for any two measurable functions f and f' and for any fuzzy measure μ, we have the following monotonicity property:

$$\big(\forall x \in \mathcal{S},\ f(x) \leq f'(x)\big) \Longrightarrow \begin{cases} S_\mu(f) \leq S_\mu(f') \\ C_\mu(f) \leq C_\mu(f') \end{cases}$$

which is also true in the infinite case;

– for any measurable function f and fuzzy measures μ and μ', we have the following monotonicity property:

$$\big(\forall A \subseteq \mathcal{S},\ \mu(A) \leq \mu'(A)\big) \Longrightarrow \begin{cases} S_\mu(f) \leq S_{\mu'}(f) \\ C_\mu(f) \leq C_{\mu'}(f) \end{cases}$$

which is also true in the infinite case;

– the following inequalities are inferred from the previous properties for any measurable function f and fuzzy measure μ:

$$\min_{i=1}^{N} f\big(x_i\big) \leq S_\mu(f) \leq \max_{i=1}^{N} f\big(x_i\big);$$

$$\min_{i=1}^{N} f\big(x_i\big) \leq C_\mu(f) \leq \max_{i=1}^{N} f\big(x_i\big);$$

– for any additive measure (or σ-additive in the infinite case), the Choquet integral coincides with the Lebesgue integral; in this regard, fuzzy integrals can be considered as an extension of Lebesgue integrals;

– for any fuzzy measurement μ, the Sugeno and Choquet integrals satisfy the following continuity property: for any sequence of measurable functions f_n in $\mathcal{S}$ such that:

$$\lim_{n \to +\infty} f_n = f$$

we have:

$$\lim_{n \to +\infty} S_\mu\big(f_n\big) = S_\mu(f),$$

$$\lim_{n \to +\infty} C_\mu\big(f_n\big) = C_\mu(f).$$

Fuzzy integrals are applied in particular in multi-criteria aggregation, but also in data fusion (as part of the mean operators) and in shape recognition.

8.3.4. *Fuzzy set measures*

Up until now, measures have been applied to crisp sets. If we now consider fuzzy sets, we need measures that provide quantitative evaluations of such sets. These measures are referred to as fuzzy sets [BOU 96] or evaluation measures [DUB 92b]. There is no real consensus over the definition of such measures. Here is the least strict of them, given in [BOU 96].

A fuzzy set measure is a function M from $\mathcal{F}$ in $\mathbb{R}^+$ such that:

1) $M(\emptyset) = 0$;

2) $\forall(\mu, \nu) \in \mathcal{F}^2,\ \mu \subseteq \nu \Rightarrow M(\mu) \leq M(\nu)$.

Other conditions can be added depending on the applications, for example:

– M has its values in $[0, 1]$;

– $M(\mathcal{S}) = 1$;

– $M(\mu) = 0 \Leftrightarrow \mu = \emptyset$;

– $M(\mu) = 1 \Leftrightarrow \mu = \mathcal{S}$.

Simple examples of fuzzy set measures are fuzzy cardinality, the cardinality of the support of μ, the supremum of μ, etc. Other examples are given in the following section: measures of fuzziness.

8.3.5. *Measures of fuzziness*

An important question regarding the evaluation of a fuzzy set involves the degree of fuzziness of the set. De Luca and Termini [LUC 72] suggested defining a degree of fuzziness as a function f of $\mathcal{F}$ into $\mathbb{R}^+$ such that:

1) $\forall \mu \in \mathcal{F},\ f(\mu) = 0 \Leftrightarrow \mu \in \mathcal{C}$ (crisp sets are non-fuzzy and are the only ones to satisfy this property);

2) $f(\mu)$ is maximum if and only if $\forall x \in \mathcal{S},\ \mu(x) = 0.5$;

3) $\forall(\mu, \nu) \in \mathcal{F}^2,\ f(\mu) \geq f(\nu)$ if ν is more contrasted than μ (closer to a binary set), i.e.:

$$\forall x \in \mathcal{S}, \begin{cases} \nu(x) \geq \mu(x) & \text{if } \mu(x) \geq 0.5 \\ \nu(x) \geq \mu(x) & \text{if } \mu(x) \leq 0.5 \end{cases}$$

4) $\forall \mu \in \mathcal{F},\ f(\mu) = f(\mu^C)$ (a fuzzy set and its complement are both as fuzzy).

De Luca and Termini also defined the entropy of a fuzzy set [LUC 72], as a degree of fuzziness, in the finite case:

$$E(\mu) = H(\mu) + H\left(\mu^C\right), \qquad [8.16]$$

where $H(\mu)$ is defined in a similar fashion to Shannon's entropy:

$$H(\mu) = -K \sum_{i=1}^{N} \mu(x_i) \log \mu(x_i). \tag{8.17}$$

It is easy to see that E satisfies all of the axioms of the degree of fuzziness. Furthermore, we have:

$$H\big(\max(\mu,\nu)\big) + H\big(\min(\mu,\nu)\big) = H(\mu) + H(\nu), \tag{8.18}$$

and:

$$E\big(\max(\mu,\nu)\big) + E\big(\min(\mu,\nu)\big) = E(\mu) + E(\nu). \tag{8.19}$$

Many other measures of fuzziness have been suggested, with similar properties. Here are the most important ones.

The Hamming distance to the closest binary set, which is nothing more than the 0.5-cut, is given by [KAU 75]:

$$f(\mu) = \sum_{i=1}^{N} \big|\mu(x_i) - \mu_{1/2}(x_i)\big|.$$

The Hamming or quadratic distance between μ and its complement [YAG 79] or more generally:

$$f(\mu) = \left[\sum_{i=1}^{N} \left|\mu(x_i) - \mu^C(x_i)\right|^p\right]^{1/p} = \left[\sum_{i=1}^{N} \left|2\mu(x_i) - 1\right|^p\right]^{1/p}.$$

The measure suggested by Kosko [KOS 90] compares the intersection of μ and μ^C with their union according to the formula:

$$\frac{\big|\min\left(\mu,\mu^C\right)\big|}{\big|\max\left(\mu,\mu^C\right)\big|}.$$

Generalized entropy is defined based on a generating function, either in additive or multiplicative form [BEZ 92]:

– the additive form is given by:

$$f(\mu) = \sum_{i=1}^{N} g\big[\mu(x_i)\big] + g\big[1 - \mu(x_i)\big] \qquad [8.20]$$

where g is a function of $[0,1]$ in $\mathbb{R}^+$ such that: $\forall t \in [0,1],\ g''(t) < 0$. Examples of generating functions are $g(t) = te^{1-t}$, $g(t) = at - bt^2$ (with $0 < b < a$), $g(t) = -t \log t$ (this last form gives the fuzzy entropy of [LUC 72]).

– the multiplicative form is given by:

$$f(\mu) = \sum_{i=1}^{N} g\big[\mu(x_i)\big] g\big[1 - \mu(x_i)\big] \qquad [8.21]$$

where g is a function of $[0,1]$ in $\mathbb{R}^+$ such that: $\forall t \in [0,1],\ g'(t) > 0$ and $g''(t) < 0$. Two examples of generating functions are $g(t) = te^{1-t}$, $g(t) = t^\alpha$.

In fusion problems, these measures of fuzziness can be used for learning membership functions. It is also possible to infer comparison measures from fuzzy set measures [BOU 96] which are used, for example, in order to compare an element of information to a model or a constraint and then are combined in a fusion or multicriteria aggregation process.

8.4. Elements of possibility theory

Possibility theory, which is derived from fuzzy set theory, was introduced by Zadeh in [ZAD 78] and later developed by several researches, particularly Dubois and Prade in France [DUB 80, DUB 88].

8.4.1. *Necessity and possibility*

A possibility measure is a function Π of $\mathcal{C}$ (whose argument is therefore a crisp subset of $\mathcal{S}$) in $[0,1]$ such that:

– $\Pi(\emptyset) = 0$;

– $\Pi(\mathcal{S}) = 1$;

– $\forall I \subset \mathbb{N},\ \forall A_i \subseteq \mathcal{S}(i \in I),\ \Pi(\cup_{i\in I} A_i) = \sup_{i\in I} \Pi(A_i)$.

In the finite case, a possibility measure is a fuzzy measure. It corresponds to the limit of equation [8.13], which is inferred from the monotonicity of a fuzzy measure.

By duality, a measure of necessity is defined as a function N of $\mathcal{C}$ into $[0,1]$ such that:

$$\forall A \subseteq \mathcal{S},\ N(A) = 1 - \Pi\big(A^C\big). \qquad [8.22]$$

This duality means that if an event is necessary, its opposite is impossible.

A measure of necessity verifies the following properties:

– $N(\emptyset) = 0$;

– $N(\mathcal{S}) = 1$;

– $\forall I \subset \mathbb{N},\ \forall A_i \subseteq \mathcal{S}(i \in I),\ N(\cap_{i\in I} A_i) = \inf_{i\in I} N(A_i)$.

Conversely, any measure satisfying these properties is, by duality, a possibility measure.

Possibility and necessity measures also have the following properties:

– $\forall A \subseteq \mathcal{S},\ \max(\Pi(A), \Pi(A^C)) = 1$, which expresses the fact that one of the two sets A and A^C is completely possible;

– $\forall A \subseteq \mathcal{S},\ \min(N(A), N(A^C)) = 0$, which expresses the fact that two opposite events cannot be simultaneously necessary;

– $\forall A \subseteq \mathcal{S},\ \Pi(A) \geq N(A)$: an event has to be possible for it to be necessary;

– $\forall A \subseteq \mathcal{S},\ N(A) > 0 \Rightarrow \Pi(A) = 1$ (since $N(A) > 0 \Rightarrow \Pi(A^C) < 1$ and $\max(\Pi(A), \Pi(A^C)) = 1$);

– $\forall A \subseteq \mathcal{S},\ \Pi(A) < 1 \Rightarrow N(A) = 0$;

– $\forall A \subseteq \mathcal{S},\ N(A) + N(A^C) \leq 1$;

– $\forall A \subseteq \mathcal{S},\ \Pi(A) + \Pi(A^C) \geq 1$.

The last two properties reflect non-additivity. Knowing $\Pi(A)$ is not enough to completely determine $\Pi(A^C)$, unlike with probability measures. The uncertainty related to an event is expressed by two numbers instead of one as before.

8.4.2. *Possibility distribution*

A possibility distribution is a function π of $\mathcal{S}$ in $[0, 1]$ with the following normalization condition:

$$\sup_{x\in\mathcal{S}} \pi(x) = 1. \qquad [8.23]$$

This condition corresponds to a closed world hypothesis, in which at least one element of $\mathcal{S}$ is completely possible. This condition can be relaxed in an open world hypothesis.

In the finite case, a possibility distribution makes it possible to define a possibility measure using the formula:

$$\forall A \in \mathcal{C},\ \Pi(A) = \sup\big\{\pi(x),\ x \in A\big\}. \qquad [8.24]$$

Conversely, a possibility measure leads to a possibility distribution:

$$\forall x \in \mathcal{S},\ \pi(x) = \Pi(\{x\}). \tag{8.25}$$

By duality, a necessity measure is defined from a possibility distribution as:

$$\forall A \in \mathcal{C},\ N(A) = 1 - \sup\big\{\pi(x),\ x \notin A\big\} = \inf\big\{1 - \pi(x),\ x \in A^C\big\}. \tag{8.26}$$

In the non-normalized case, we no longer have $\Pi(\mathcal{S}) = 1$. Likewise, the properties $N(A) > 0 \Rightarrow \Pi(A) = 1$ and $\Pi(A) < 1 \Rightarrow N(A) = 0$ are no longer true.

These definitions have a simple interpretation if we consider the problem of how to represent the value of a variable, in which $\mathcal{S}$ represents the variation range of this variable. A possibility distribution on $\mathcal{S}$ describes the degrees to which the variable can have each possible value. It is actually the fuzzy set of the possible values for this variable. The degree of membership of each value to this set corresponds to the degree of possibility for the variable to have this value. Therefore, a possibility distribution can represent the imprecision related to the variable's exact value. Typically, a fuzzy number is a possibility distribution that describes the possible values that this number can have.

Let us consider, for example, a classification problem in image processing. Here is a list of examples (not a comprehensive one) of possibility distributions:

– let $\mathcal{S}$ be the set of classes. A possibility distribution on $\mathcal{S}$, defined for each object to classify (point, area, etc.), can represent the degrees to which each object can belong to each of these classes;

– let $\mathcal{S}$ be a characteristic space (for example, a scale of gray levels). A possibility distribution on $\mathcal{S}$ can be defined for each class and represent, for each gray level, the possibility for that class to appear in the image with that gray level;

– let $\mathcal{S}$ be the image space. A possibility distribution on $\mathcal{S}$ can be defined for each class and give for each point of the image its degree of possibility of belonging to that class.

In the definition given here, we have always considered the possibility and the necessity of a crisp subset of $\mathcal{S}$. Now, consider a fuzzy set μ of $\mathcal{S}$ ($\mu \in \mathcal{F}$). The concept of possibility must then be extended [ZAD 78]:

$$\Pi(\mu) = \sup_{x \in \mathcal{S}} \min\big(\mu(x), \pi(x)\big). \tag{8.27}$$

This corresponds to the following interpretation: given a possibility distribution π on $\mathcal{S}$, associated with a variable X taking its values in $\mathcal{S}$, we can assess to which extent "X is μ". In this way, the possibility of μ combines the degree to which the variable X has the value x and the degree of membership of x to the fuzzy set.

8.4.3. *Semantics*

The membership functions and possibility distributions can have different semantics. Here are the major ones:

– a semantics of degree of similarity (the concept of distance);

– a semantics of degree of plausibility for an object for which only an imprecise description is available to actually be the one we are trying to identify;

– a semantics of degree of preference (a fuzzy class is then the set of "right" choices), this interpretation being closer to the concept of a utility function.

These three types of semantics are used in signal and image processing as well as in fusion.

8.4.4. *Similarities with the probabilistic, statistical and belief interpretations*

Rather than opposing the various formalisms, it is interesting to emphasize the cases where the interpretations overlap. Several of these similarities are given in [DUB 99], and we will sum them up here.

A possibility distribution π, representing, for example, knowledge about the possible values of a variable x, can also be interpreted as a family of subsets $\{A_1, \dots A_n\}$, each one included in the next with $A_i \subset A_{i+1}$, to which levels of confidence λ_i are attributed, which are defined by:

$$\lambda_i = N\big(A_i\big) = 1 - \Pi\big(A_i^C\big). \qquad [8.28]$$

Since the necessity N is monotonic, we have $\lambda_1 \leq \dots \leq \lambda_n$. This implies that the set of values of the possibility distribution is finite. Let $\alpha_1 = 1,\ \alpha_2 \geq \dots \geq \alpha_n$ be these values and let $\alpha_{n+1} = 0$. Then the (A_i, λ_i) are given by:

$$A_i = \big\{s \in \mathcal{S},\ \pi(s) \geq \alpha_i\big\}, \quad \lambda_i = 1 - \alpha_{i+1}. \qquad [8.29]$$

Conversely, the least specific possibility distribution (i.e. the most possible) associated with $\{(A_1, \lambda_1), \dots (A_n, \lambda_n)\}$ that verifies $\lambda_i = N(A_i)$ is given by:

$$\pi(s) = \min_i \max\big(1 - \lambda_i, A_i(s)\big). \qquad [8.30]$$

If we define $p_i = \alpha_i - \alpha_{i+1}$, we have $\sum_i p_i = 1$ and:

$$\pi(s) = \sum_{i, s \in A_i} p_i. \qquad [8.31]$$

Therefore, the (A_i, p_i) form random closed sets, each one included in the next, or even focal sets in belief function theory. The corresponding mass functions are thus described as consonant. Let us note that p_i is not the probability for x (the variable whose possibility distribution is π) to belong to A_i, but rather the probability for A_i to actually represent the knowledge available regarding x. Once again, we are dealing with the two concepts of imprecision (through the size of A_i) and uncertainty (the value of p_i).

The value of λ_i can also be interpreted as a lower bound of the probability for the actual value of x to be in A_i. The distribution π is then equivalent to a family $\mathcal{P}$ of probabilities:

$$\mathcal{P} = \left\{P,\ P\left(A_i\right) \geq \lambda_i,\ i = 1 \ldots n\right\}. \quad [8.32]$$

Possibility is then interpreted as the upper probability:

$$\Pi(B) = P^*(B) = \sup\left\{P(B),\ P \in \mathcal{P}\right\}, \quad [8.33]$$

and necessity as the lower probability:

$$N(B) = P_*(B) = \inf\left\{P(B),\ P \in \mathcal{P}\right\}. \quad [8.34]$$

Interpretations in terms of likelihood functions involve probabilities of the form $P(s_m \mid s)$ where s_m refers to the measured value of x and s to its actual value. The distribution π can then be identified with $P(s_m \mid s)$. For any subset A, we have:

$$\min_{s \in A} P\left(s_m \mid s\right) \leq P\left(s_m \mid A\right) \leq \max_{s \in A} P\left(s_m \mid s\right) \quad [8.35]$$

which then makes it possible to interpret $\Pi(A)$ as the upper bound of $P(s_m \mid A)$. However, we cannot get very far with this since the information available is usually lower than $P(s_m \mid s)$.

8.5. Combination operators

Following Zadeh's original work [ZAD 65], many operators were suggested in the fuzzy community to combine membership functions or possibility distributions[2]. These operators are also called connectives, or combination or aggregation operators.

2. Let us note that because a possibility distribution and a membership function have similar mathematical expressions and because there is a connection between the two, the operators can apply to either of them. However, they have different origins, meanings and semantics, which is a point that should not be overlooked.

The major classes of operators are described in [BLO 96b, DUB 85, DUB 88, DUB 99, YAG 91]. Among the main operators, we can mention in particular T-norms, T-conorms [MEN 42, SCH 83], means [GRA 95, YAG 88], symmetric sums and operators that take into account conflict measures or also the reliability of sources [DEV 93, DUB 92a]. This sections contains the major definitions. The interpretations in terms of set operations and information fusion will be discussed further in section 8.10.

Since most operators work point by point (i.e. by combining the membership or plausibility degrees in the same point of $\mathcal{S}$), it is sufficient to define them for the possible values of the membership functions or possibility distributions. Therefore, the operators are defined as functions of $[0,1]$ or $[0,1] \times [0,1]$ in $[0,1]$. In what follows, the letters x, y, etc. will refer to the values we wish to combine, i.e. the values in $[0,1]$ representing the degrees of membership or possibility.

8.5.1. *Fuzzy complementation*

A fuzzy complementation is a function c of $[0,1]$ in $[0,1]$ such that:

– $c(0) = 1$;

– $c(1) = 0$;

– c is involutive: $\forall x \in [0,1],\ c(c(x)) = x$;

– c is strictly decreasing.

The simplest example is that given in section 8.2:

$$\forall x \in [0,1],\ c(x) = 1 - x. \qquad [8.36]$$

Since it is difficult to directly construct involutive functions, it is useful to characterize them using a simpler, more general form. Thus, continuous complementations have the following general form:

$$\forall x \in [0,1],\ c(x) = \varphi^{-1}\big[1 - \varphi(x)\big], \qquad [8.37]$$

with $\varphi : [0,1] \rightarrow [0,1]$, such that:

– $\varphi(0) = 0$;

– $\varphi(1) = 1$;

– φ is strictly increasing.

There are several functions φ that verify these properties and it is easy to come up with one. The simplest example is:

$$\forall x \in [0,1],\ \varphi(x) = x^n, \qquad [8.38]$$

which enables us to construct the following complementation:

$$\forall x \in [0,1],\ c(x) = \left(1 - x^n\right)^{1/n}. \qquad [8.39]$$

For $n = 1$, we find the most common form used, i.e. $c(x) = 1 - x$. The higher n becomes (for $n > 1$), or the lower it becomes (for $n < 1$), the more binary the resulting form. In the first case, most of the values, except for those close to 1, have a complement that is close to 1 and in the second case, most of the values, except for those close to 0, have a complement close to 0.

If, for a real number a in $]0,1]$, φ has the following form:

$$\forall x \in [0,1],\ \varphi(x) = \frac{ax}{(1-a)x+1}, \qquad [8.40]$$

then the corresponding complementation is:

$$\forall x \in [0,1],\ c(x) = \frac{1-x}{1+a^2x}. \qquad [8.41]$$

Here is another example, depending on four parameters a, b, c and n such that $0 \leq a < b < c \leq 1$:

$$\forall x \in [0,1],\ c(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq a \\ 1 - \dfrac{1}{2}\left[\dfrac{x-a}{b-a}\right]^n & \text{if } a \leq x \leq b \\ \dfrac{1}{2}\left[\dfrac{c-x}{c-b}\right]^n & \text{if } b \leq x \leq c \\ 0 & \text{if } c \leq x \leq 1 \end{cases} \qquad [8.42]$$

A few examples are illustrated in Figure 8.2.

8.5.2. *Triangular norms and conorms*

In the context of stochastic geometry [MEN 42, SCH 83], a triangular norm, or t-norm, is a function $t : [0,1] \times [0,1] \rightarrow [0,1]$ such that:

– t is commutative: $\forall (x,y) \in [0,1]^2,\ t(x,y) = t(y,x)$;

– t is associative: $\forall (x,y,z) \in [0,1]^3,\ t[t(x,y),z] = t[x,t(y,z)]$;

– 1 is a neutral element: $\forall x \in [0,1],\ t(x,1) = t(1,x) = x$;

– t is increasing with respect to the two variables:

$$\forall (x,x',y,y') \in [0,1]^4,\ (x \leq x' \text{ and } y \leq y') \Longrightarrow t(x,y) \leq t(x',y').$$

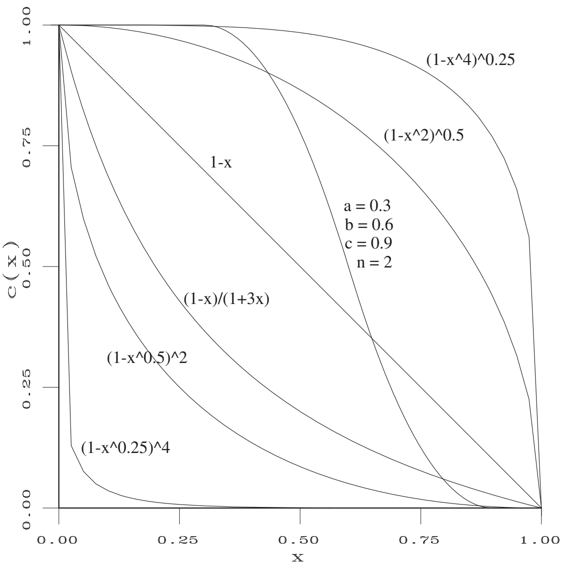


Figure 8.2. *A few examples of fuzzy complementation*

Additionally, we have: $t(0,1) = t(0,0) = t(1,0) = 0$, $t(1,1) = 1$ and 0 is a zero element ($\forall x \in [0,1],\ t(x,0) = 0$).

Continuity is often added to this list of properties.

The operators $\min(x,y)$, xy, $\max(0, x+y-1)$ are examples of t-norms, which are by far the most commonly used.

T-norms generalize to fuzzy sets the concept of intersection as well as the logical "and".

The following result is easy to prove. For any t-norm t, we have:

$$\forall (x,y) \in [0,1]^2,\ t(x,y) \leq \min(x,y). \qquad [8.43]$$

This shows that the "min" is the highest t-norm and that any t-norm has a conjunctive behavior.

On the other hand, any t-norm is always higher than t_0, which is the smallest t-norm, defined by:

$$\forall (x,y) \in [0,1]^2,\ t_0(x,y) = \begin{cases} x & \text{if } y = 1 \\ y & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases} \qquad [8.44]$$

Furthermore, the t-norms mentioned above verify:

$$\forall(x, y) \in [0, 1]^2,\ t_0(x, y) \leq \max(0, x + y - 1) \leq xy \leq \min(x, y). \quad [8.45]$$

Let us note, however, that there is no complete order for all of the t-norms.

Parametric forms allow some variations between certain of these operators. For example, the t-norm defined in [YAG 80] by:

$$\forall(x, y) \in [0, 1]^2,\ t(x, y) = 1 - \min\left[1, \left[(1 - x)^p + (1 - y)^p\right]^{1/p}\right] \quad [8.46]$$

varies from the Lukasiewicz t-norm $\max(0, x + y - 1)$ for $p = 1$ to the min for $p = +\infty$.

Examples of t-norms are shown in Figure 8.3.

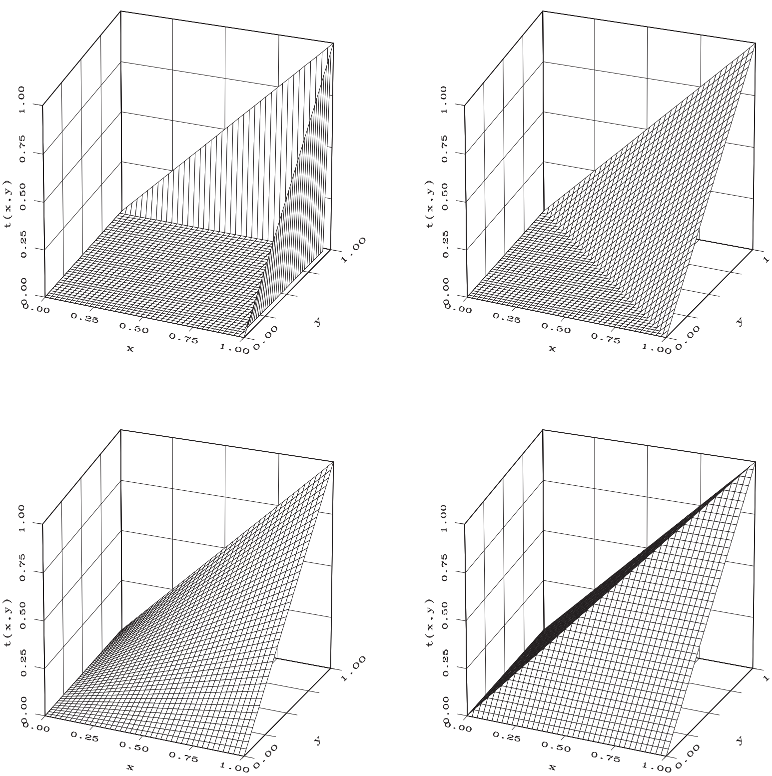


Figure 8.3. *Four examples of t-norms. First line: t_0 (lowest t-norm) Lukasiewicz t-norm. Second line: product and minimum (highest t-norm)*

Based on a t-norm t and a complementation c, another operator T, referred to as the t-conorm, can be defined by duality:

$$\forall (x,y) \in [0,1]^2,\ T(x,y) = c\big[t\big(c(x), c(u)\big)\big]. \qquad [8.47]$$

Therefore, a t-conorm is a function $T : [0,1] \times [0,1] \to [0,1]$ such that:

– T is commutative: $\forall (x,y) \in [0,1]^2,\ T(x,y) = T(y,x)$;

– T is associative: $\forall (x,y,z) \in [0,1]^3,\ T[T(x,y), z] = T[x, T(y,z)]$;

– 0 is an identity element: $\forall x \in [0,1],\ T(x,0) = T(0,x) = x$;

– T is increasing with respect to the two variables:

$$\forall (x,x',y,y') \in [0,1]^4,\ (x \leq x' \text{ and } y \leq y') \Longrightarrow T(x,y) \leq T(x',y').$$

Furthermore, we have: $T(0,1) = T(1,1) = T(1,0) = 1$, $T(0,0) = 0$ and 1 is a zero element ($\forall x \in [0,1],\ T(x,1) = 1$).

The most common examples of t-conorms are: $\max(x,y)$, $x + y - xy$, $\min(1, x+y)$.

T-conorms generalize to fuzzy sets the concept of union or of the logical "or".

For any t-conorm, we have:

$$\forall (x,y) \in [0,1]^2,\ T(x,y) \geq \max(x,y). \qquad [8.48]$$

This shows that the "max" is the smallest t-conorm and that any t-conorm has a disjunctive behavior.

On the other hand, any t-conorm is smaller than T_0, which is the highest t-conorm, defined by:

$$\forall (x,y) \in [0,1]^2,\ T_0(x,y) = \begin{cases} x & \text{if } y = 0 \\ y & \text{if } x = 0 \\ 1 & \text{otherwise} \end{cases} \qquad [8.49]$$

Furthermore, we have the following inequalities between the most common t-conorms[3]:

$$\forall (x,y) \in [0,1]^2,\ T_0(x,y) \geq \min(1, x+y) \geq x + y - xy \geq \max(x,y). \qquad [8.50]$$

3. Again, there is no complete order for all of the t-conorms.

Examples of t-conorms are shown in Figure 8.4.

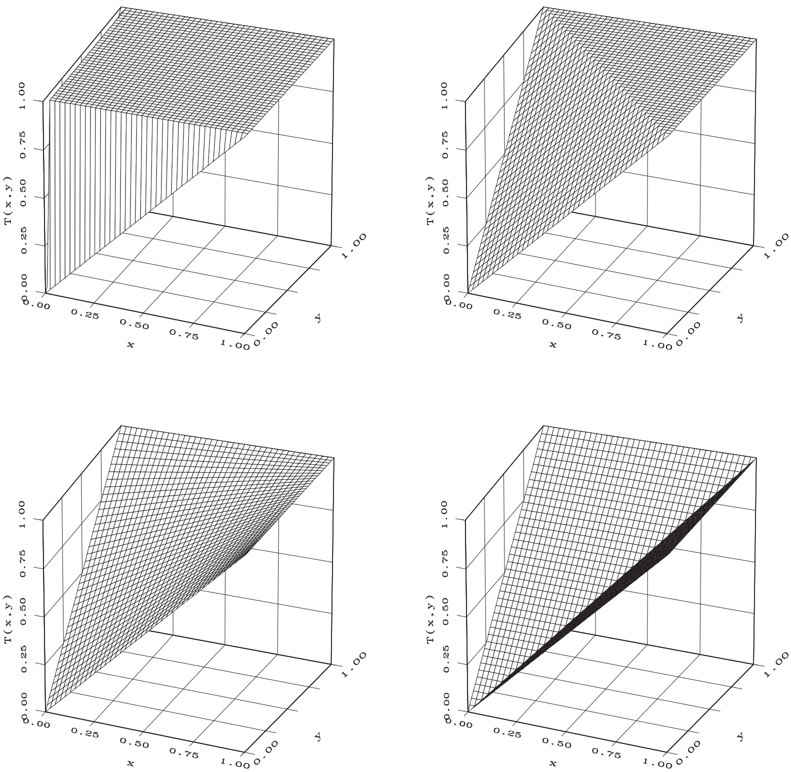


Figure 8.4. *Four examples of t-conorms. First line: T_0 (highest t-conorm) and Lukasiewicz t-conorm. Second line: algebraic sum and maximum (smallest t-conorm)*

Here are a few other useful properties of these operators:

– any t-norm or t-conorm is distributive over the "min" and the "max", and therefore we have equalities of the type:

$$\forall (x, y, z) \in [0, 1]^3,\ t\big[x, \min(y, z)\big] = \min\big[t(x, y), t(x, z)\big]; \qquad [8.51]$$

– the only mutually distributive t-norms and t-conorms are the "min" and the "max";

– the only idempotent t-norm is the "min", and the only idempotent t-conorm is the "max";

– based on any t-norm t and any continuous, strictly increasing function h from $[0,1]$ in $[0,1]$ such that $h(0) = 0$ and $h(1) = 1$, we can define another t-norm t' using the formula [SCH 63]:

$$\forall (x,y) \in [0,1]^2,\ t'(x,y) = h^{-1}\big[t\big(h(x), h(y)\big)\big]. \tag{8.52}$$

This gives us a way of generating families of t-norms based on an example.

There are generic forms for t-norms and t-conorms with specific properties [DUB 85]. We will now discuss the two most useful groups: archimedean and nilpotent t-conorms.

A strictly monotonic, archimedean t-norm t verifies:

$$\forall x \in [0,1],\ t(x,x) < x, \tag{8.53}$$

$$\forall (x,y,y') \in [0,1]^3,\ y < y' \Longrightarrow t(x,y) < t(x,y'). \tag{8.54}$$

Likewise, a strictly monotonic archimedean t-conorm T verifies the two following properties:

$$\forall x \in [0,1],\ T(x,x) > x, \tag{8.55}$$

$$\forall (x,y,y') \in [0,1]^3,\ y < y' \Longrightarrow T(x,y) < T(x,y'). \tag{8.56}$$

Any strictly monotonic, archimedean t-norm t can be expressed in the following form:

$$\forall (x,y) \in [0,1]^2,\ t(x,y) = f^{-1}\big[f(x) + f(y)\big], \tag{8.57}$$

where f, referred to as the generating function, is a continuous and decreasing bijection of $[0,1]$ into $[0,+\infty]$ such that $f(0) = +\infty$ and $f(1) = 0$.

The associated t-conorms have the following form:

$$\forall (x,y) \in [0,1]^2,\ T(x,y) = \varphi^{-1}\big[\varphi(x) + \varphi(y)\big], \tag{8.58}$$

where the generating function φ is a continuous and increasing bijection of $[0,1]$ into $[0,+\infty]$ such that $\varphi(0) = 0$ and $\varphi(1) = +\infty$.

Such t-norms and t-conorms never satisfy the non-contradiction law and the excluded middle law. These laws are expressed as:

$$\forall x \in [0,1],\ t\big[x, c(x)\big] = 0, \tag{8.59}$$

$$\forall x \in [0,1],\ T\big[x, c(x)\big] = 1. \tag{8.60}$$

These two equations are usually not verified by strictly monotonic, archimedean t-norms and t-conorms.

Any strictly monotonic, archimedean t-norm (or t-conorm) can be defined based on multiplicative generators, by equivalence with additive generators [CHE 89]:

$$\forall (x,y) \in [0,1]^2,\ t(x,y) = h^{-1}\big[h(x)h(y)\big], \tag{8.61}$$

where h is a strictly increasing function of $[0,1]$ into $[0,1]$ such that $h(0) = 0$ and $h(1) = 1$. The equivalence with the additive form is obtained simply by defining:

$$h = e^{-f} \tag{8.62}$$

where f is an additive generating function.

The most common t-norms and t-conorms in this class are the product and the algebraic sum:

$$\forall (x,y) \in [0,1]^2,\ t(x,y) = xy, \quad T(x,y) = x + y - xy. \tag{8.63}$$

The only rational t-norms of this class are the Hamacher t-norms defined by [HAM 78]:

$$\forall (x,y) \in [0,1]^2,\ \frac{xy}{\gamma + (1-\gamma)(x + y - xy)}, \tag{8.64}$$

where γ is a positive parameter (for $\gamma = 1$ we get the product again). They are illustrated in Figure 8.5.

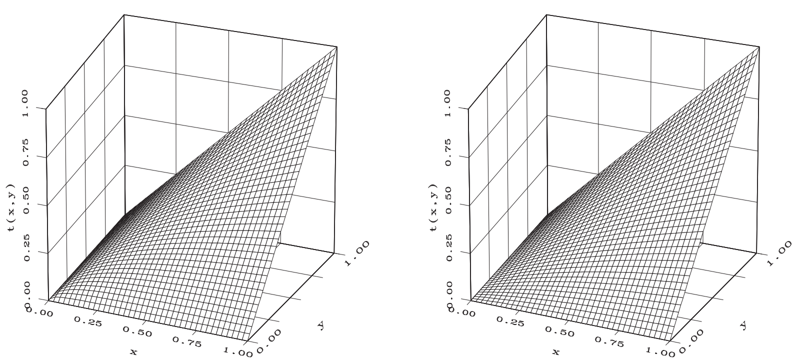


Figure 8.5. *Two examples of Hamacher t-norms, for $\gamma = 0$ (left) and $\gamma = 0.4$ (right)*

Another parametric family of this class is comprised of the Frank functions, defined by [FRA 79]:

$$\forall (x,y) \in [0,1]^2,\ t(x,y) = \log_s \left[1 + \frac{\left(s^x - 1\right)\left(s^y - 1\right)}{s-1} \right], \qquad [8.65]$$

where s is a strictly positive parameter. These t-norms and their dual t-conorms satisfy the following notable relation (and t-norms and t-conorms alone satisfy this relation):

$$\forall (x,y) \in [0,1]^2,\ t(x,y) + T(x,y) = x + y. \qquad [8.66]$$

Examples of Frank t-norms are shown in Figure 8.6. If s is small and tends to 0, the t-norm tends to the minimum. If s tends to $+\infty$, the t-norm tends to the Lukasiewicz t-norm. If $s = 1$, we end up with the product again.

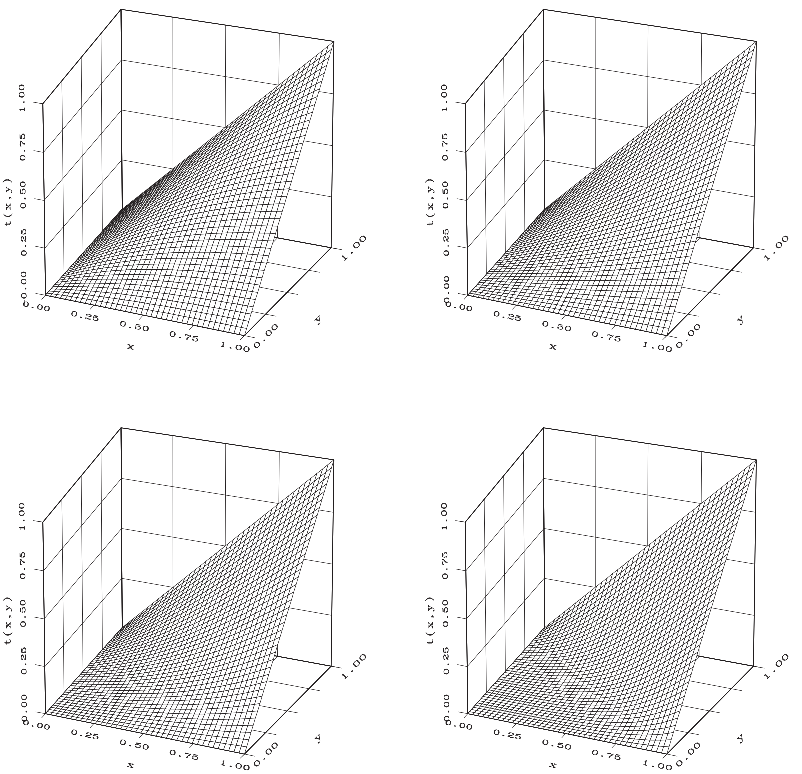


Figure 8.6. *Four examples of Frank t-norms. First line: $s = 0.1$ and $s = 2$. Second line: $s = 10$ and $s = 1,000$*

The second useful family of t-norms and t-conorms is comprised of the nilpotent operators, which have the following general form:

$$\forall(x,y) \in [0,1]^2,\ t(x,y) = f^*\big[f(x) + f(y)\big], \tag{8.67}$$

where f is a decreasing bijection of $[0,1]$ into $[0,1]$, such that $f(0) = 1$, $f(1) = 0$ and $f^*(x) = f^{-1}(x)$ if $x \in [0,1]$, with $f^*(x) = 0$ if $x \geq 1$. The general form of nilpotent t-conorms is inferred from it by duality.

These operators satisfy the excluded middle law and the non-contradiction law.

The most common t-norms and t-conorms in this class are the Lukasiewicz operators:

$$\forall(x,y) \in [0,1]^2,\ t(x,y) = \max(0, x+y-1),\ T(x,y) = \min(1, x+y).$$

Examples of generating functions f have been suggested by Schweizer and Sklar [SCH 63], as well as by Yager [YAG 80].

Operators combining t-norms and t-conorms have been suggested, for example, by [ZIM 80]:

$$\forall(x,y) \in [0,1]^2,\ C_\gamma(x,y) = t(x,y)^{1-\gamma} T(x,y)^\gamma, \tag{8.68}$$

where γ is a parameter in $[0,1]$.

8.5.3. *Mean operators*

A mean operator is a function $m : [0,1] \times [0,1] \to [0,1]$ such that:

– the result of the combination is always included between the min and the max: $\forall(x,y) \in [0,1]^2,\ \min(x,y) \leq m(x,y) \leq \max(x,y)$, but $m \neq \min$ and $m \neq \max$;

– m is commutative;

– m is increasing with respect to the two variables:

$$\forall(x,x',y,y') \in [0,1]^4,\ (x \leq x' \text{ and } y \leq y') \Longrightarrow m(x,y) \leq m(x',y').$$

The first property entails that m is always an idempotent operation:

$$\forall x \in [0,1],\ m(x,x) = x.$$

These operators are generally not associative; the medians are the only ones that are, where $m(x, y)$ is the median value of x, y and a parameter α in $[0, 1]$:

$$m(x,y) = med(x,y,\alpha) = \begin{cases} x & \text{if } y \le x \le \alpha \text{ or } \alpha \le x \le y \\ y & \text{if } x \le y \le \alpha \text{ or } \alpha \le y \le x \\ \alpha & \text{if } y \le \alpha \le x \text{ or } x \le \alpha \le y \end{cases} \qquad [8.69]$$

Figure 8.7 illustrates a few median operators.

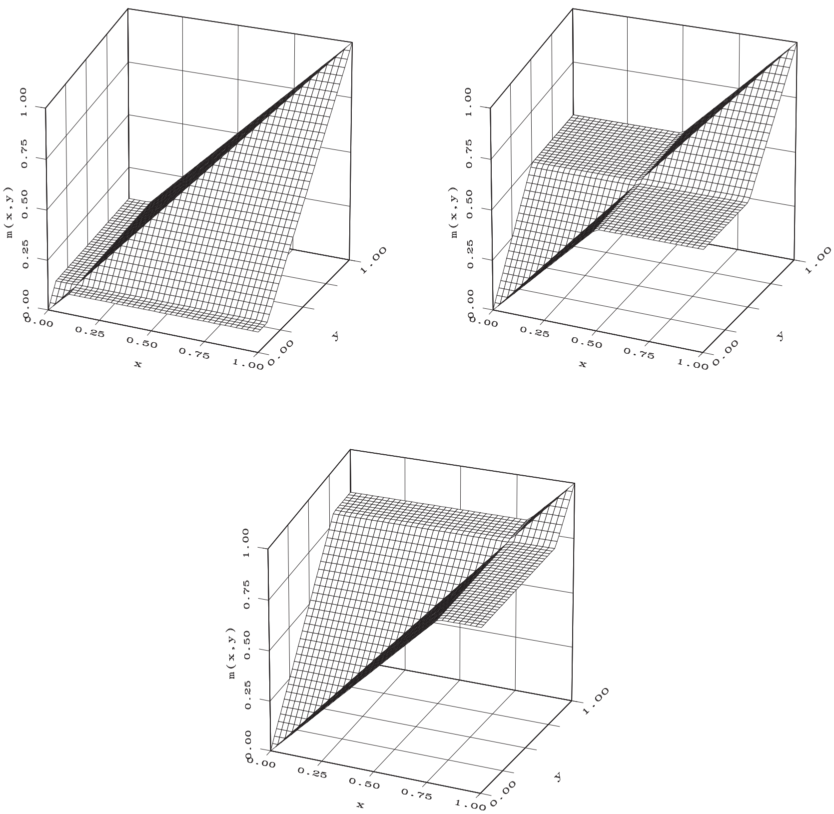


Figure 8.7. *Three examples of medians, for α equal to 0.1, 0.5 and 0.8, respectively*

A weaker property than associativity is bisymmetry:

$$\forall (x, y, z, t) \in [0, 1]^4,\ m\big[m(x, y), m(z, t)\big] = m\big[m(x, z), m(y, t)\big]. \qquad [8.70]$$

The means that satisfy this property and are continuous and strictly increasing have the following general form:

$$\forall (x, y) \in [0, 1]^2,\ m(x, y) = k^{-1}\left[\frac{k(x) + k(y)}{2}\right], \qquad [8.71]$$

where k is a continuous, strictly increasing function of $[0, 1]$ into $[0, 1]$. The function k can be interpreted as a change of scale or dynamics of the values to combine. These values, once they have been transformed by k, are then combined using a simple arithmetic mean, and the result is then changed back to the initial scale.

The most common means are obtained from functions k of the type:

$$\forall x \in [0, 1],\ k(x) = x^\alpha,$$

with $\alpha \in \mathbb{R}$. The arithmetic mean $(x+y)/2$ is obtained for $\alpha = 1$, the quadratic mean $\sqrt{(x^2 + y^2)/2}$ for $\alpha = 2$, the harmonic mean $2xy/(x+y)$ for $\alpha = -1$, the geometric mean $\sqrt{xy}$ for $\alpha = 0$. At the limit when α tends to $-\infty$ or $+\infty$, m tends to the min or the max. Table 8.1 sums up these results.

α	$m(x, y)$	comment
$-\infty$	$\min(x, y)$	limit value
-1	$\frac{2xy}{x+y}$	harmonic mean
0	$(xy)^{-1/2}$	geometric mean
$+1$	$\frac{x+y}{2}$	arithmetic mean
$+2$	$\sqrt{\frac{x^2+y^2}{2}}$	quadratic mean
$+\infty$	$\max(x, y)$	limit value

Table 8.1. *Examples of continuous, strictly increasing and bisymmetric means. For the harmonic mean, we adopt the convention that* $m(0, 0) = 0$

Figure 8.8 shows a few examples of means.

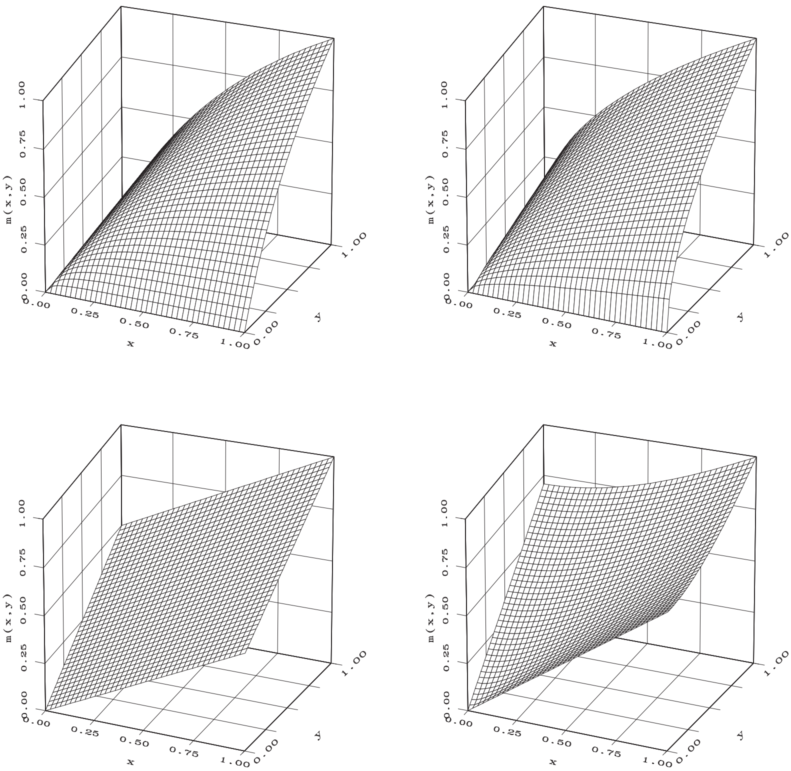


Figure 8.8. *Four examples of mean operators. First line: harmonic and geometric means. Second line: arithmetic and quadratic means*

Within the class of mean operators, we also have weighted means, OWAs (Ordered Weighted Average) [YAG 88] and the fuzzy integrals discussed above [GRA 95].

In OWAs, the weights are defined by the ranks of the values to combine. Let $a_1, a_2, \ldots a_n$ be these values. They are arranged in a sequence $a_{j_1}, a_{j_2}, \ldots a_{j_n}$ such that:

$$a_{j_1} \leq a_{j_2} \leq \cdots \leq a_{j_n}.$$

Then, for a set of weights w_i that verifies:

$$\sum_{i=1}^{n} w_i = 1, \quad \forall i,\ 1 \leq i \leq n,\ w_i \in [0,1],$$

the OWA operator is defined by the expression:

$$\text{OWA}\left(a_1, a_2, \ldots, a_n\right) = \sum_{i=1}^{n} w_i a_{j_i}. \qquad [8.72]$$

We can also consider fuzzy integrals to be included in this class of operators [GRA 95], since Choquet and Sugeno integrals are idempotent, continuous, increasing and included between the minimum and the maximum. It includes the specific case of order statistics and therefore the minimum, the maximum and the median. The Choquet integrals defined with respect to an additive measure μ are equivalent to a weighted arithmetic mean, in which the weights w_i assigned to the values x_i are equal to $\mu(\{x_i\})$.

OWAs can also be interpreted as a particular class of Choquet integrals, where the fuzzy measure is defined by:

$$\forall A,\ |A| = i, \quad \mu(A) = \sum_{j=0}^{i-1} w_{n-j}.$$

Conversely, any commutative Choquet integral is such that $\mu(A)$ only depends on $|A|$ and is equal to an OWA whose weights are given by:

$$w_1 = 1 - \sum_{i=2}^{n} w_i,$$

$$\forall i \geq 2,\ w_i = \mu\big(A_{n-i+1}\big) - \mu\big(A_{n-i}\big),$$

where A_i refers to any subset such that $|A_i| = i$.

A more detailed study of the properties of these operators can be found in [GRA 92, GRA 95].

8.5.4. *Symmetric sums*

Symmetric sums are defined by an auto-duality property, which corresponds to the invariance of the result of the operation by inverting the scale of values to combine. More specifically, a symmetric sum is a function $\sigma : [0, 1] \times [0, 1] \to [0, 1]$ such that:

– $\sigma(0, 0) = 0$;

– σ is commutative;

– σ is increasing with respect to the two variables;

– σ is continuous;

– σ is self-dual: $\forall (x, y) \in [0, 1]^2$, $\sigma(x, y) = 1 - \sigma(1 - x, 1 - y)$.

Let us note that auto-duality differs from the duality mentioned between t-norms and t-conorms. For those operators, inverting the scale changes the type of operator. Here, the scale of values can be inverted without changing the way they are combined. This property was used in particular for combining expert opinions. It could also be expressed with other complementations.

From these basic properties, we can infer that:

– $\sigma(1,1) = 1$,

– $\forall x \in]0,1[, \sigma(x, 1-x) = 1/2$,

– the only symmetric sum that is both associative and a mean is the median with the parameter $1/2$.

The general form of symmetric sums is given by:

$$\sigma(x,y) = \frac{g(x,y)}{g(x,y) + g(1-x, 1-y)}, \qquad [8.73]$$

where g is a continuous, positive, increasing function of $[0,1] \times [0,1]$ into $[0,1]$, such that $g(0,0) = 0$. Typically, a continuous t-norm or t-conorm can be chosen as g.

If $\forall x \in [0,1]$, $g(0,x) = 0$, then $\sigma(0,1)$ is not defined; otherwise $\sigma(0,1) = 1/2$.

The general form of strictly increasing, associative, symmetric sums is given by:

$$\forall (x,y) \in [0,1]^2,\ \sigma(x,y) = \psi^{-1}\big[\psi(x) + \psi(y)\big], \qquad [8.74]$$

where ψ is a strictly monotonic function such that $\psi(0)$ and $\psi(1)$ are not bounded and $\forall x \in [0,1], \psi(1-x) + \psi(x) = 0$. From this, we infer that 0 and 1 are identity elements and that $1/2$ is a zero element.

Table 8.2 shows a few typical examples of symmetric sums. They are obtained by using various t-norms and t-conorms as generating function g.

$g(x,y)$	$\sigma(x,y)$	property
xy	$\sigma_0(x,y) = \frac{xy}{1-x-y+2xy}$	associative
$x+y-xy$	$\sigma_+(x,y) = \frac{x+y-xy}{1+x+y-2xy}$	non-associative
$\min(x,y)$	$\sigma_{\min}(x,y) = \frac{\min(x,y)}{1-\lvert x-y\rvert}$	mean
$\max(x,y)$	$\sigma_{\max}(x,y) = \frac{\max(x,y)}{1+\lvert x-y\rvert}$	mean

Table 8.2. *Examples of symmetric sums, defined based on t-norms and t-conorms. For σ_0, we adopt the convention that $\sigma_0(0,1) = \sigma_0(1,0) = 0$ and for $\sigma_{\min}$, we assume that $\sigma_{\min}(0,1) = \sigma_{\min}(1,0) = 0$*

These operations can be arranged in two different orders depending on the values to combine:

$$x + y \leq 1 \Longrightarrow \sigma_0(x, y) \leq \sigma_{\min}(x, y) \leq \sigma_{\max}(x, y) \leq \sigma_+(x, y), \qquad [8.75]$$

$$x + y \geq 1 \Longrightarrow \sigma_0(x, y) \geq \sigma_{\min}(x, y) \geq \sigma_{\max}(x, y) \geq \sigma_+(x, y). \qquad [8.76]$$

The four examples in Table 8.2 are illustrated in Figure 8.9.

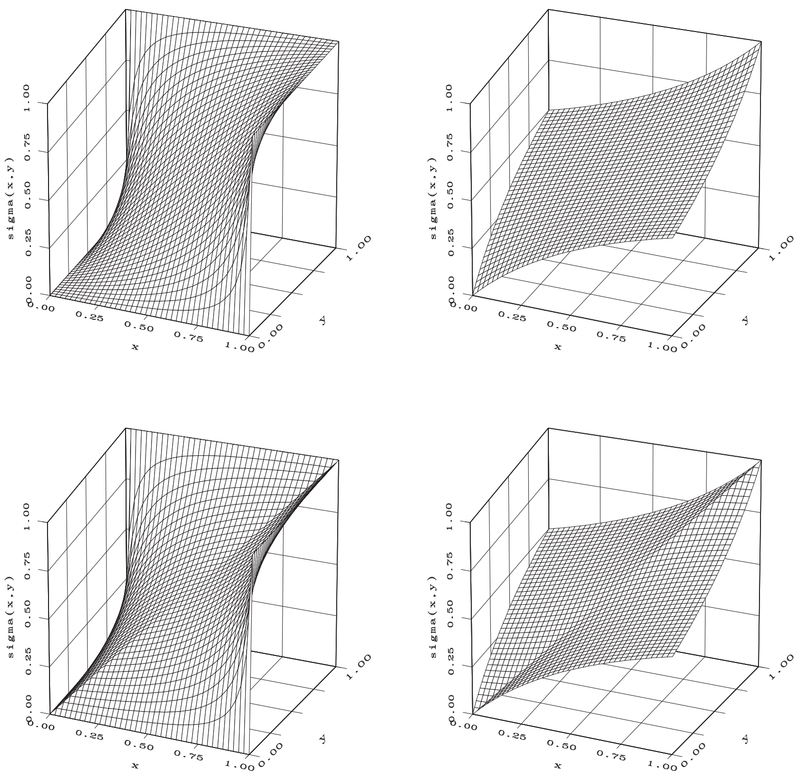


Figure 8.9. *Four examples of symmetric sums. First line: σ_0 and σ_+. Second line: $\sigma_{\min}$ and $\sigma_{\max}$*

8.5.5. *Adaptive operators*

There are many operators found in other works that we will not discuss here. We will simply mention adaptive operators according to conflict between possibility distributions [DUB 92a], which behave like a min when the distributions are consonant and like a max when they produce a strong conflict. Let π_1 and π_2 be two possibility distributions that we wish to combine into an overall distribution π'. They may represent, for example, the imprecision on a variable estimated in two different ways, for which we want an overall estimation.

Let us consider, for example, the case of the conjunctive combination of two possibility distributions π_1 and π_2 defined on D. This type of combination is well suited for the case where the distributions overlap at least partially, i.e. when certain classes are presented as possible by the two sources. If this is not the case, the sources are in conflict and a possible measure of conflict is:

$$h\big(\pi_1, \pi_2\big) = 1 - \max_{c \in D} \min\big(\pi_1(c), \pi_2(c)\big), \tag{8.77}$$

which represents 1 minus the height of the intersection between the two distributions (calculated by a min in this equation). The combination can be normalized by this height, but this would hide the conflict: a possibility of 1 is always assigned to the classes presented as the most possible by both sources, even if that possibility is low (this problem is similar to that mentioned in section 7.4 about the conjunctive combination of belief functions). In terms of conflict, the interpretation of this quantity matches our intuition of triangular or trapezoidal possibility distributions (and more generally of monomodal possibility distributions), but it is not well suited for forms in which a single point can generate a strong conflict value, even if the two distributions are different in that point only.

In the extreme case of completely conflicting distributions, conjunctive combination leads to an identically zero distributions. A disjunctive combination is then the preferred method, making it possible to keep all of the data if it is presented as possible by at least one of the two sources. The underlying hypothesis is that at least one of the sources is reliable.

Here are a few examples of the possible formulae for π':

$$\pi'(s) = \max\left[\frac{t\big[\pi_1(s), \pi_2(s)\big]}{h\big(\pi_1, \pi_2\big)}, 1 - h\big(\pi_1, \pi_2\big)\right], \tag{8.78}$$

$$\pi'(s) = \min\left[1, \frac{t\big[\pi_1(s), \pi_2(s)\big]}{h\big(\pi_1, \pi_2\big)} + 1 - h\big(\pi_1, \pi_2\big)\right], \tag{8.79}$$

$$\pi'(s) = t\big[\pi_1(s), \pi_2(s)\big] + 1 - h\big(\pi_1, \pi_2\big), \tag{8.80}$$

$$\pi'(s) = \max\left[\frac{\min\big(\pi_1, \pi_2\big)}{h}, \min\big[\max\big(\pi_1, \pi_2\big), 1 - h\big]\right]. \tag{8.81}$$

The first two forms combine normalized conjunction with a constant distribution of conflict value, whereas the latter allows us to switch from a strictly conjunctive combination, when the conflict is equal to zero, to a strictly disjunctive combination, when the conflict is equal to 1 (see the example in Figure 8.10). However, this operator is not associative. Let us note that the min could be replaced with another t-norm.

The last formula (equation [8.81]) is illustrated in Figure 8.10 for two possibility distributions with an increasing conflict.

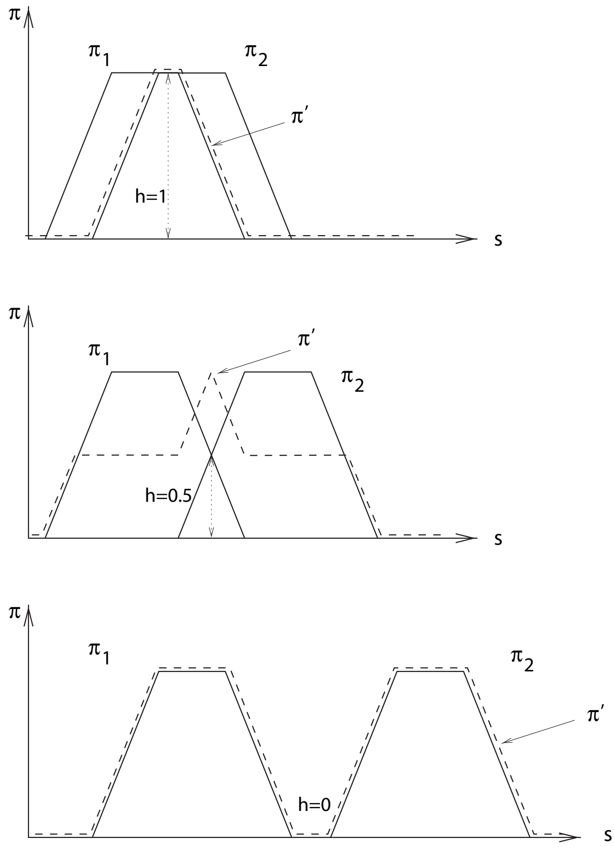


Figure 8.10. *Example of adaptive operators varying from the min to the max when the conflict between two distributions increases*

When the sources are unequally reliable and information is available regarding this reliability, the level of conflict between two sources indicates to what extent the information provided by the less reliable source can be taken into account. If, for example, π_1 is more reliable than π_2, we can consider that if they are in agreement with each other, π_2 can provide information and make the fusion more accurate by conjunction. If, on the other hand, the two sources are in conflict, it is preferable not to take into account π_2. The following operator models this behavior:

$$\min\left[\pi_1, \max\left[\pi_2, h\left(\pi_1, \pi_2\right)\right]\right]. \qquad [8.82]$$

This only assumes knowing an order of reliability of the sources.

If, additionally, we have access to numerical values of reliability (a much more stringent constraint than the previous hypothesis), we can then transform the possibility distributions into distributions with equivalent reliabilities. Let w_j be the reliability coefficient of π_j; if the source is completely reliable, this coefficient is equal to 1 and it is equal to 0 if the source is not reliable at all. The transformation of π_j works according to the formula:

$$\max\left(\pi_j, 1 - w_j\right) \qquad [8.83]$$

which amounts to conducting a disjunction between π_j and a constant distribution of value $1 - w_j$. Thus, if the source is completely reliable, the corresponding distribution is not modified, whereas if it had not been reliable at all, the distribution would have become constant and equal to 1, which represents absence of knowledge (any element is completely possible). Once the distributions have been transformed, they can be combined conjunctively.

Other operators of this kind can be found in [DUB 99], but we will not discuss them in detail here.

These operators can also be used conditionally to the classes, to take into account the specificities of the sources for each class. Two sources may, for example, be in conflict over a class but not over the others, a source may be reliable for certain classes and not others, etc. Although these ideas are still not often applied in fuzzy fusion of signals and images, the theoretical framework allows it.

8.6. Linguistic variables

It often happens to have numerical representations that are not suited for the description of a situation. For example, if a variable has a wide variation range, it can be difficult to assign a precise value to each specific situation and the preferred method will consist of using more qualitative terms, taken from natural language, to generally crudely define subsets that are typical of interesting situations. For example, to describe the size of an object, it can be easier and more appropriate to only use a few terms with flexible boundaries, such a small, medium, large. This corresponds to a certain granularity of information. According to [ZAD 96], the concept of "granule" is the starting point for "computing with words" theories. Zadeh defined a granule as "a fuzzy set of points having the form of a clump of elements drawn together by similarity" [ZAD 96]. A word then becomes a label for a granule. In order to perform calculations with such representations, specific tools have to be developed. The field of fuzzy reasoning particularly benefits from these tools, as well as the field of knowledge fusion or approximate or crude data.

These types of representations are referred to as linguistic variables. They are variable whose values are words, phrases or sentences [ZAD 75]. Their advantage lies essentially in the fact that linguistic characterizations can be less specific than numerical ones and therefore require less information to be used and handled in reasoning systems.

8.6.1. *Definition*

A linguistic variable is defined by a quintuple $(x, T(x), \mathcal{S}, G, M)$ where x is the variable's name, $T(x)$ the set of values of x (referred to as terms), $\mathcal{S}$ is the domain or the universe in which the values of the variable are defined, G is a syntactic rule which makes it possible to generate the name X of each value of x and M is a semantic rule, since $M(X)$ is the fuzzy set defined in $\mathcal{S}$ that represents the meaning of X [DUB 80, ZAD 75, ZIM 91].

This definition represents a symbolic-numerical conversion and establishes ties between language and numerical scales.

8.6.2. *An example of a linguistic variable*

Let us consider the example of an object's size. In numerical terms, this size can be expressed using a value that varies inside a domain $\mathcal{S}$ (typically, $\mathcal{S}$ is a subset of $\mathbb{R}^+$). In linguistic terms, size can be expressed by using terms such as very small, small, medium, large, very large, etc. The semantics of these terms are defined by fuzzy sets in $\mathcal{S}$. Figure 8.11 illustrates the concept of the linguistic variable "size".

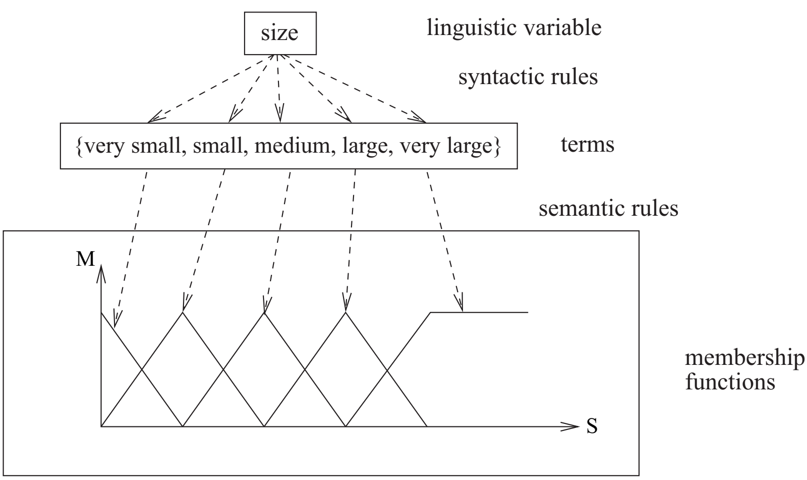


Figure 8.11. *Illustration of the linguistic variable "size", its terms and the associated fuzzy sets. The arrows drawn from the linguistic variable to the term set represent syntactic rules. The second set of arrows represents the semantic rules and translates the terms into membership functions*

8.6.3. *Modifiers*

The meaning of a term of a linguistic variable can be modulated by operators known as modifiers. If A is a fuzzy set, then the modifier h allows us to construct a composite term $h(A)$ that is a fuzzy set in the same universe $\mathcal{S}$. Here are the most common operators:

– normalization:

$$\mu_{\text{norm}(A)}(u) = \frac{\mu_A(u)}{\sup_{v \in \mathcal{S}} \mu_A(v)},$$

where μ_A refers to the membership function to A and u is an arbitrary value in $\mathcal{S}$;

– concentration:

$$\mu_{\text{con}(A)}(u) = \big[\mu_A(u)\big]^2;$$

– dilation[4]:

$$\mu_{\text{dil}(A)}(u) = \big[\mu_A(u)\big]^{0.5};$$

– contrast enhancement:

$$\mu_{int(A)}(u) = \begin{cases} 2\big[\mu_A(u)\big]^2 & \text{if } \mu_A(u) \in [0, 0, 5] \\ 1 - 2\big[1 - \mu_A(u)\big]^2 & \text{otherwise.} \end{cases}$$

Typical modifiers defined using these operators are [DUB 80]:

– very $A = \text{con}(A)$,

– more or less $A = \text{dil}(A)$,

– plus $A = A^{1.25}$,

– slightly A = int[norm(plus A and not(very A))] where "and" and "not" are defined by a t-norm and a complementation, respectively.

8.7. Fuzzy and possibilistic logic

The development of fuzzy logic is directly related to the specificities of human reasoning: more flexible than traditional propositional logic, it tolerates imprecision and can be used to make inferences even in the presence of imperfect data and knowledge. It is capable of dealing with gradual predicates, originating either from the use

4. Note that this is not a dilation in the morphological sense.

of continuous frames of reference, or from concepts of typicality (a situation may generally be typical of known situations and this has to be accounted for when reasoning by analogy).

When reasoning with propositions, the uncertainty (in a wide sense) corresponds to the inability to state whether a proposition is true or false, either because the information is incomplete, vague, imprecise, or because the information is contradictory or fluctuating.

In the first case, a possibilistic model makes it possible to take into account this type of uncertainty, whereas in the second case, a probabilistic model would be well suited.

There is another important distinction between the degree of certainty and the degree of truth. In fuzzy logic, propositions are assigned a degree of truth, whereas in possibilistic logic, they are usually assigned degrees of uncertainty.

8.7.1. *Fuzzy logic*

In fuzzy logic [DUB 80, DUB 91], reasoning is based on elementary fuzzy propositions of the type:

$$X \text{ is } P \qquad [8.84]$$

where X is a variable with possible values in the reference space $\mathcal{S}$ and P is a fuzzy subset of $\mathcal{S}$, with the membership function μ_P.

The degrees of truth of such propositions are defined as values in $[0,1]$ based on μ_P.

Logical connectives are defined in a very simple way, by using the same operators as their set equivalents. For example, the degree of truth of a conjunction such as:

$$X \text{ is } A \quad \text{and} \quad Y \text{ is } B$$

is defined based on a t-norm t by:

$$\mu_{A\wedge B}(x,y) = t\big[\mu_A(x), \mu_B(y)\big].$$

Likewise, a disjunction such as:

$$X \text{ is } A \quad \text{or} \quad Y \text{ is } B$$

has a degree of truth based on a t-conorm T:

$$\mu_{A\vee B}(x,y) = T\big[\mu_A(x), \mu_B(y)\big],$$

and a negation has a degree of truth defined by a fuzzy complementation c:

$$\mu_{\neg A}(x) = c\big[\mu_A(x)\big].$$

In the case of variables with values in a product space, i.e. X with values in $\mathcal{S}$ and Y with values in $\mathcal{V}$, conjunction is interpreted as a cartesian product. The degree of truth of:

$$X \text{ is } A \quad \text{and} \quad Y \text{ is } B$$

is then written:

$$\mu_{A\times B}(x,y) = t\big[\mu_A(x), \mu_B(y)\big].$$

Now let us consider the implication. In classical logic, we have:

$$A \Longrightarrow (B) \Longleftrightarrow (B \text{ or non-}A), \qquad [8.85]$$

and therefore the implication is expressed based on a disjunction and a negation. By using the same equivalence in the fuzzy case, a fuzzy implication is defined based on a t-conorm (disjunction) and a complementation (negation). Let A and B be non-fuzzy sets. The degree to which A implies B is defined by:

$$\text{Imp}(A,B) = T\big[c(A), B\big] \qquad [8.86]$$

where T is a t-conorm and c is a complementation.

In the case where A and B are fuzzy, we have:

$$\text{Imp}(A,B) = \inf_x T\big[c\big(\mu_A(x)\big), \mu_B(x)\big]. \qquad [8.87]$$

The following table sums up the major fuzzy implications used in other works for fuzzy reasoning:

$T(x,y) = \max(x,y)$	$\max(1-a, b)$	Kleene-Diene
$T(x,y) = \min(1, x+y)$	$\min(1, 1-a+b)$	Lukasiewicz
$T(x,y) = x+y-xy$	$1-a+ab$	Reichenbach

In any case, we get the same classical implication table, given below, in the extreme cases of true (1) and false (0) propositions, in other words when using binary degrees of truth.

A	B	$A \Rightarrow B$
0	0	1
0	1	1
1	0	0
1	1	1

With these few definitions, we can now define the fuzzy equivalents of the major reasoning modes: modus ponens, modus tollens, syllogism, contraposition. Consider the example of modus ponens. In classical logic, it is written:

$$\big(A \wedge (A \Longrightarrow B)\big) \Longrightarrow B. \qquad [8.88]$$

Its fuzzy equivalent is defined as follows:

– let us consider the rule:

$$\text{if } X \text{ is } A \quad \text{then} \quad Y \text{ is } B;$$

– and the knowledge:

$$X \text{ is } A'$$

where A' is an approximation of A;

– we then come to the conclusion:

$$Y \text{ is } B'$$

where B' is an approximation of B, with the degree:

$$\mu_{B'}(y) = \sup_x t\big[\mu_{A \Rightarrow B}(x, y), \mu_{A'}(x)\big]. \qquad [8.89]$$

We can now model and handle fuzzy rule systems. For example, let us consider the rule:

$$\text{IF} \quad (x \text{ is } A \quad \text{AND} \quad y \text{ is } B) \quad \text{THEN} \quad z \text{ is } C$$

and α as the degree of truth of x is A, β as the degree of truth of y is B, γ as the degree of truth of z is C. The rule's degree of truth (or of satisfaction) is obtained by combining the fuzzy connectors defined above:

$$\operatorname{Imp}\big(t(\alpha,\beta),\gamma\big), \qquad [8.90]$$

hence:

$$T\big[c\big(T(\alpha,\beta)\big),\gamma\big]. \qquad [8.91]$$

Likewise for the rule:

$$\text{IF} \quad (x \text{ is } A \quad \text{OR} \quad y \text{ is } B) \quad \text{THEN} \quad z \text{ is } C$$

its degree of satisfaction is:

$$\operatorname{Imp}\big(T(\alpha,\beta),\gamma\big) = T\big[c\big(T(\alpha,\beta)\big),\gamma\big]. \qquad [8.92]$$

These rules can be used to describe in a qualitative fashion the graph of a fuzzy function using a small number of rules. For example, a function such as that in Figure 8.12 can be described, on a rather crude granularity level, by:

IF X is small THEN Y is small

IF X is medium THEN Y is large

IF X is large THEN Y is small

These rules rely on the concept of linguistic variables, discussed above, and the semantics of the values "small", "medium", "large", are defined by fuzzy sets over the definition domains of X and Y.

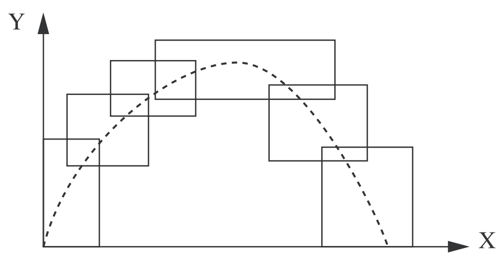


Figure 8.12. *An example of a function's graph*

Fuzzy rule systems have been used in many fields, essentially in fuzzy control, but also for fuzzy reasoning, for modeling flexible criteria in image processing and fusion, etc.

8.7.2. *Possibilistic logic*

Possibilistic logic relies on the definition of a possibility measure Π in a Boolean algebra B with the formulae [DUB 91]:

$$\Pi : B \longrightarrow [0, 1]$$

such that:

– $\Pi(\bot) = 0$;

– $\Pi(\top) = 1$;

– $\forall\varphi, \phi, \Pi(\varphi \vee \psi) = \max(\Pi(\varphi), \Pi(\psi))$;

– $\forall\varphi$, $\Pi(\exists x\varphi) = \sup\left\{\Pi\big(\varphi[a \mid x]\big),\ a \in D(x)\right\}$ (where $D(x)$ is the domain of the variable x and $\varphi[a|x]$ is obtained by replacing the occurrences of x in φ with a).

Now let Ω be the set of interpretations and let π be a normalized possibility distribution:

$$\pi : \Omega \longrightarrow [0, 1]$$

such that:

$$\exists\omega \in \Omega,\ \pi(\omega) = 1.$$

The possibility of a formula is then expressed as:

$$\Pi(\varphi) = \sup\left\{\pi(\omega),\ \omega \models \varphi\right\} \qquad [8.93]$$

where $\omega \models \varphi$ is read "ω is a model of φ", meaning that φ is satisfied in the world ω.

As we did with sets, a necessity measure is defined for formulae using duality by:

$$N(\varphi) = 1 - \Pi(\neg\varphi). \qquad [8.94]$$

We then have the following property:

$$\forall\varphi, \phi,\ N(\varphi \wedge \psi) = \min\big(N(\varphi), N(\psi)\big). \qquad [8.95]$$

This formalism can be used to deal with many situations by modeling them in a very simple way. For example, a default rule such as "if A then B", with possible exceptions, can be simply expressed as:

$$\Pi(A \wedge B) \geq \Pi(A \wedge \neg B) \qquad [8.96]$$

Likewise, possibilistic modus ponens reasoning can be modeled by:

– if we have the rule: $N(A \Rightarrow B) = \alpha$

– and an element of knowledge written as: $N(A) = \beta$

– then the conclusion can be expressed by: $\min(\alpha, \beta) \leq N(B) \leq \alpha$.

The formalism of possibilistic logic is used in many fields, for example, to represent preference or utility models in the form of KBS and then to use these systems for reasoning [DUB 99].

Let us consider a knowledge base of the type:

$$KB = \{(\varphi_i, \alpha_i),\ i = 1 \ldots n\}$$

where α_i is a degree of certainty or priority associated with the formula φ_i (representing an element of knowledge).

Satisfying this set of formulae in each world is represented by a possibility distribution defined as follows. In the case where the knowledge base is comprised of only one formula, we have:

$$\pi_{(\varphi,\alpha)}(\omega) = \begin{cases} 1 & \text{if } \omega \models \varphi \\ 1 - \alpha & \text{otherwise} \end{cases} \qquad [8.97]$$

More generally, for a set of elements of knowledge with priorities, we have:

$$\pi_{KB}(\omega) = \min_{i=1\ldots n} \{1 - \alpha_i, \omega \models \neg\varphi_i\} = \min_{i=1\ldots n} \max\big(1 - \alpha_i, \varphi_i(\omega)\big). \qquad [8.98]$$

This formula is interpreted this way: if a formula is significant (α_i close to 1), this formula's degree of satisfaction in the world ω is taken into account. If, on the other hand, it is not significant (α_i close to 0), then it will not come into play in the overall evaluation of the knowledge base. The min corresponds to the fact that we are trying to know to what extent the formulae of the knowledge base are simultaneously satisfied in ω.

The inconsistency degree of the base KB can be measured using the expression:

$$1 - \max_{\omega} \pi_{KB}(\omega). \qquad [8.99]$$

A base is said to be complete if it can be used to infer whether any formula is true or whether its opposite is true: either $KB \vdash \varphi$ (KB allows us to infer φ), or $KB \vdash \neg\varphi$ (KB allows us to infer $\neg\varphi$).

When a base is not complete, it leads to an absence of knowledge regarding certain formulae φ: $KB \nvdash \varphi$ and $KB \nvdash \neg\varphi$. Using possibilistic logic, this situation can be represented in a very simple way:

$$\Pi(\varphi) = \Pi(\neg\varphi) = 1$$

whereas there is no such simple model in probability theory, for example.

8.8. Fuzzy modeling in fusion

Among the non-probabilistic techniques that have appeared over the past 10 years in fusion, fuzzy set theory provides a very efficient tool for explicitly representing imprecise information, in the form of membership functions [BAN 78, KAU 75, ZAD 65], as we saw above. As a result, the measure $M_i^j(x)$ introduced in Chapter 1 is written in the form:

$$M_i^j(x) = \mu_i^j(x), \tag{8.100}$$

where $\mu_i^j(x)$ refers, for example, to the degree of membership of x to the class C_i according to the source I_j, or the translation of a symbolic element of information expressed by a linguistic variable (see, for example, [DEL 92]).

In other works there are two main methods for using fuzzy sets in image processing [BLO 96a]: the first is more symbolic in nature and expresses in the form of fuzzy rules the membership of certain structures to a class, depending on measurements obtained by image processing; the second uses fuzzy sets to directly represent the classes or structures in the image, spatially covering the objects with a membership function. Let us consider the example of the "road" class in a satellite image. In the first approach, we would describe the road in a linguistic form such as "a road is a rather elongated structure". The membership of an object to the road class will then be represented by a function associating its length with a degree in $[0, 1]$. Any parallel contour detection algorithm can then be used to assign to the objects it detects a degree of membership to the road class, depending on their length. In the second approach, the road is directly represented in the image by a fuzzy set, with membership degrees that are strong in the middle of the road and close to 0 in fields or forest.

These functions are not subjected to the axiomatic constraints imposed by probabilities and hence offer a greater flexibility in modeling them. This flexibility can be seen as a disadvantage since it can easily leave the user helpless to define these functions. The disadvantage of fuzzy sets is that they represent essentially the imprecise nature of information, whereas uncertainty is represented implicitly and can only be obtained by inference from the different membership functions.

Possibility theory [DUB 88, ZAD 78], which is derived from fuzzy sets, allows us to represent both the imprecision and the uncertainty, by using possibility distributions π on a set S and two functions characterizing events: the possibility Π and the necessity N.

A possibility distribution is interpreted as a function that gives the degree of possibility for a variable to have the value s, with S being the domain of the variable's values. The distribution π is then interpreted as the membership function to the fuzzy subset S for the possible values of this variable. In the framework of numerical fusion, a possible application of this theory consists of choosing $S = D$ (the class set) and defining the measure M_i^j by:

$$M_i^j(x) = \pi_j^x(C_i), \qquad [8.101]$$

i.e. as the possibility degree for the class to which x belongs to have the value C_i, according to the source I_j. This defines a possibility distribution for each source and each element x. The possibility and the necessity for each class are then written:

$$\Pi_j(\{C_i\}) = \pi_j(C_i), \quad N_j(\{C_i\}) = \inf\{(1 - \pi_j(C_k)),\ C_k \neq C_i\}. \qquad [8.102]$$

For any subset A of D, the possibility and the necessity are calculated using formulae [8.24] and [8.26].

This modeling assumes that classes are crisp, whereas the fuzzy model defined by equation [8.100] assumes fuzzy classes.

Generally speaking, there are three interpretations of fuzziness in other works, in terms of plausibilities, similarities, preferences [DUB 99]. The same interpretations are used in signal and image fusion. The interpretation in terms of plausibilities is used for the membership to a class, in the definition of a fuzzy spatial object (an object with imprecise limits). The interpretation in terms of similarities is that used for the definition of a fuzzy class in a characteristic space as a function of the distance to a prototype, for example, of linguistic variables representing information or knowledge about spatial objects, or also of degrees of satisfaction of a relation, a constraint. Finally, preferences are used in the expression of choice criteria (for example, for planning applications in robotics), which are often related to constraints or knowledge outside the image.

8.9. Defining membership functions or possibility distributions

Constructing membership functions or possibility distributions can be done in several ways.

In most applications, this construction is done either by taking ideas directly from probabilistic learning methods, from heuristics, from neuromimetic methods used for learning the parameters of particular forms of membership functions, or finally by minimizing classification criteria [BEZ 81]. Below is a description of the major methods.

A first method consists of defining a fuzzy class membership function based on the image's intensity function I (the gray levels):

$$\mu_i(x) = F_i[I(x)], \qquad [8.103]$$

where F_i is a function that is determined according to the problem. The most commonly used are normalization functions or S functions [PAL 92] (which is equivalent to considering that the lighter parts of the image have a high membership to the class), functions Π (monomodal, they associate the class with a range of gray levels with imprecise limits), or also multimodal functions.

These functions are often determined under supervision, but can also be learned, for example, using automatic classification algorithms such as fuzzy C-means [BEZ 81] or possibilistic C-means [KRI 93] (see, for example, [BEZ 99] for an overview of fuzzy classification algorithms). The main drawback of fuzzy C-means is that the membership functions have counter-intuitive forms: the class membership values are non-decreasing with respect to the distance to the center of the class. This problem is avoided with possibility C-means.

Other characteristics can be used to achieve this goal. For example, the set of contours in an image can be defined by a spatial fuzzy set whose membership function is a function of the image's gradient:

$$\mu_i(x) = F_i\big[\nabla I(x)\big], \qquad [8.104]$$

where F is a decreasing function.

If specific object detectors are available, the membership functions of these objects can be defined as functions of the response to these detectors (the case of contours falls into this category). For example, a road detector can provide in a satellite image a response whose amplitude increases with the membership to the road.

In the case of linguistic variables, the forms of membership functions and their parameters are often defined by the user.

The spatial imprecision over the definition of the limits between classes (if the membership functions are defined in the image space) can be introduced based on a preliminary binary detection of the classes. A membership function is constructed as equal to 1 inside the binary area at a certain distance from the edges, as equal to 0 outside this area at a certain distance from the edges and as decreasing between these two limits. For example, an imprecision zone on the edge of the class can be modeled as the zone included between the erosion and the dilatation of this object, since the size of these operations depends on the spatial extension of the imprecision we wish to represent. If R is the binary area we start with, $E^n(R)$ its erosion of size n and $D^m(R)$ its dilatation of size m, the fuzzy class membership function can be defined by:

$$\begin{aligned} &\mu(x) = 1 \text{ if } x \in E^n(R), \\ &\mu(x) = 0 \text{ if } x \in D^m(R)^C, \\ &\mu(x) = F\big[d\big(x, E^n(R)\big)\big] \text{ otherwise} \end{aligned} \qquad [8.105]$$

where F is a decreasing function of the distance from x to $E^n(R)$.

The construction of possibility distributions can also be done from probabilistic learning, followed by a transformation of probability into possibility. Several methods have been suggested for this purpose. The main advantage in signal and image processing is that statistical information is often available, particularly the histogram, which is well suited for applying statistical learning methods. We then get probability distributions p_k. Their transformation into possibility distributions π_k (both distributions are assumed to be discrete and $1 \leq k \leq K$) is achieved according to various criteria [DEV 85, DUB 83, KLI 92], such as not changing the order, normalization constraints, conserving the uncertainty measured by the entropy [KLI 92], the consistency $p - \pi$ expressed by [DEL 87]:

$$\forall k, \ \pi_k \leq p_k,$$

which is not very satisfactory (an unlikely class can be possible), or [ZAD 78]:

$$\sum_{k=1}^{K} p_k \pi_k = c$$

where c is a constant in $[0, 1]$, or also a more general relation involving all of the subsets A [DUB 80]:

$$N(A) \leq P(A) \leq \Pi(A).$$

A comparison of these methods can be found in [KLI 92].

Other methods try to directly estimate the membership functions based on the histogram, in order to optimize entropy criteria [CHE 95] or minimal specificity and consistency criteria [CIV 86].

In any case, these methods attempt to find a similarity between the histogram and the membership functions or the possibility distributions and do not take into consideration interpretations that are specific to fuzziness because they invalidate some of these similarities. For example, the tails of the histogram correspond to classes with little representation, hence with values that can be very low, even if the points involved belong to the corresponding classes. The method suggested in [BLO 97] provides a way of avoiding this problem with the help of a criterion that combines the similarities of membership functions and the histogram where they have meaning, with an *a priori* form of the functions that correspond to the desired interpretation. The parameters of the membership functions are then estimated in order to optimize this criterion by using a simulated annealing method.

8.10. Combining and choosing the operators

One of the advantages of fuzzy set and possibility theory, beyond the fact that it imposes few constraints on modeling, is that it offers a wide variety of combination operators. We will present the main ones, then give a few indications on how to choose a fusion operator according to its properties and its behavior.

An important feature, common to every theory, of these combination operators is that they provide us with a result of the same nature as the functions we started with (the closure property) and therefore with the same interpretation in terms of imprecision and uncertainty. Therefore, they make it possible not to make any partial binary decision before the combination takes place, which could lead to inconsistencies that would be difficult to eliminate. The decision is only made at the very end, based on the result of the combination.

In fuzzy set and possibility theory, a number of combination modes are possible [DUB 85, YAG 91]. Among the major operators, we can mention in particular t-norms, t-conorms [MEN 42, SCH 83], means [GRA 95, YAG 88], symmetric sums and operators that take into account conflict or source reliability measures [DEV 93, DUB 92a], as we saw in section 8.5. From here on, the letters x, y, etc. refer to the values we wish to combine, i.e. values in $[0, 1]$ that therefore represent μ_i^j or $\pi_j(C_i)$ in this case.

The choice of a fusion operator is made according to several criteria presented in [BLO 96b].

A first criterion is the operator's behavior. Strict, lenient or cautious behaviors are expressed in mathematical form as conjunction, disjunction, or compromise. Let x and y be two real numbers (in $[0, 1]$) representing the degrees of confidence to combine. The combination of x and y by an operator F is described as:

– conjunctive if $F(x, y) \leq \min(x, y)$ (corresponding to a strict behavior);

– disjunctive if $F(x, y) \geq \max(x, y)$ (lenient behavior);

– compromise if $x \leq F(x, y) \leq y$ if $x \leq y$ and $y \leq F(x, y) \leq x$ otherwise (cautious behavior).

This distinction is not sufficient to categorize operators whose behaviors are not always the same. This is why the classification defined in [BLO 96b] describes operators not only as conjunctive and disjunctive, but also depending on their behavior according to the values of the information to combine. The three classes suggested correspond to:

– context independent constant behavior (CICB) operators: the result depends only on the values to combine (the calculation involves no other information) and the behavior is the same regardless of what those values are;

– context independent variable behavior (CIVB) operators: the behavior depends on the numerical values of the information to fuse;

– context dependent (CD) operators, for example, of more comprehensive knowledge such as the reliability of the sensors, or the conflict between the sources.

Fuzzy fusion operators fall into three categories. T-norms, which generalize set intersection to fuzzy sets, are conjunctive CICB operators, since for any t-norm t, we have:

$$\forall(x, y) \in [0, 1]^2,\ t(x, y) \leq \min(x, y).$$

On the other hand, t-conorms which generalize union are disjunctive CICB operators, since for any t-conorms T, we have:

$$\forall(x, y) \in [0, 1]^2,\ T(x, y) \geq \max(x, y).$$

Mean operators are also CICBs and have a compromise behavior, since they verify:

$$\forall(x, y) \in [0, 1]^2,\ \min(x, y) \leq m(x, y) \leq \max(x, y).$$

Let us note that Bayesian fusion, in which the operator involved is a product, and fusion of belief functions using Dempster's orthogonal sum are also conjunctive.

In the CIVB operator class we have, for example, certain symmetric sums. Generally speaking, any associative symmetric sum σ (except for medians) has the following behavior [DUB 88]:

– conjunctive if $\max(x, y) < 1/2$: $\sigma(x, y) \leq \min(x, y)$;

– disjunctive if $\min(x, y) > 1/2$: $\sigma(x, y) \geq \max(x, y)$;

– compromise if $x \leq 1/2 \leq y$: $x \leq \sigma(x, y) \leq y$ (and the opposite inequality if $y \leq 1/2 \leq x$).

Non-associative symmetric sums also have a variable behavior, but according to less simple rules [BLO 96b].

In the CIVB operator class, we also have the operators suggested in the MYCIN system for combining certainty factors [SHO 75].

Examples of CD operators are found in possibility theory. Earlier, we presented operators that depended on an overall measure of the conflict between two sources of information [DUB 92a], which are applicable to cases where one of the two elements of information is reliable, but where we do not know which one, so that:

– they are conjunctive if the sources are consonant (low conflict): in this case, the two sources are necessarily reliable and therefore the operator can be strict;

– they are disjunctive if the sources are dissonant (high conflict): a disjunction then favors all of the possibilities provided by both sources;

– they have a compromise behavior in the case of partial conflict: these cases are the most problematic and the operators are then “cautious”.

The difficulty is then to find a good conflict measure. That suggested as the maximum of the intersection between two possibility distributions [DUB 92a] is not always well suited to image processing problems, particularly for the classification of multi-source data. Fuzzy distances (see, for example, [BLO 99]) can provide solutions to this problem.

The advantage of DC operators for image processing is undeniable, since they allow us to take into consideration a wider variety of situations, several of which occur simultaneously in image processing. Here are a few examples:

– sources can be in conflict when they provide information regarding one type of event (a class, for example) and consonant for another class;

– sources can have different overall reliabilities;

– a source can be reliable for one class and poorly reliable for another, etc.

Unfortunately, these operators still have not, in our opinion, been developed far enough in image processing and would deserve specific research.

This classification, which includes all of the commonly used operators, constitutes a first criterion for choosing an operator for a specific application.

A second criterion is given by the properties of operators and their interpretations in terms of uncertain, imprecise, incomplete or ambiguous data fusion.

The commutativity and associativity properties reflect the fact that the result of the combination does not depend on what order the elements of information are arranged in when they are combined. Whereas commutativity is satisfied by all of the commonly used operators, this is not systematic with associativity (means and symmetric sums usually are not associative). These two properties are often laid out as the minimum properties that fusion operators have to satisfy. However, human reasoning does not always comply with them. For example, a photo interpreter often starts by constructing a primary interpretation of the scene based on a single image, then improves this interpretation by using the other images, according to a process that clearly is not commutative.

The existence of an identity element means that a source yielding this value will have no influence on the result of the combination and represents some sort of indifference on the part of the source towards the information sought, or even a complete absence of knowledge regarding it. Such an element exists for t-norms and t-conorms.

Another distinctive element, the zero element, means that a source yielding this value has complete determination over the result of the fusion. Such elements also exist for t-norms and t-conorms.

The property of increasingness is usually imposed on operators and matches what our intuition tells us.

Boundary conditions, which define the behavior of the operators when the information to combine has extreme values, guarantee compatibility with the binary case, where all of the propositions are simply either true or false (this corresponds to the constraint of complying with deductive logic imposed by Cox for defining an inductive logic [COX 46]).

The continuity property satisfied by most operators guarantees the robustness of the fusion. However, this property is not always necessary, since natural phenomena (particularly time phenomena) are not always continuous.

Idempotence means that providing information that is already available will not change the fusion result. This property is not systematically imposed. It is verified by means, the t-norm min and the t-conorm max (and those are the only ones). We might want, on the contrary, to have the combination of two identical values reinforce or weaken the overall result. Let us consider the example of identical simultaneous testimonies. If the witnesses are plotting together, it is not surprising to see them saying the same thing and the associated degrees of confidence will therefore be combined in an idempotent way. Whereas if they are independent, the credibility of what they are saying will be reinforced if they are trusted, or weakened if they are not. Let us note that the combination rules modeling these behaviors have been known since Bernoulli. Generally speaking, it is considered that if sources are dependent (in the cognitive sense), idempotence can be imposed, whereas if they are independent, reinforcement effects can be needed.

Along the same lines, the nilpotence property will be imposed, for example, to combine consecutive testimonies, in order to model the deterioration of information along a chain of witnesses that are not completely reliable. For example, for certain t-conorms, satisfying this property will help achieve a result equal to 1 by combining a certain number of measures, which are not all equal to zero. This type of behavior may be useful when the information is the result of a long processing chain.

The excluded middle and non-contradiction properties, satisfied only for certain operators, have an accepted interpretation in reasoning terms, in the field of artificial intelligence and fuzzy reasoning. There are examples in image processing where the excluded middle is not advisable, whenever there is a need for introducing absence of knowledge regarding an event and its complements and therefore to relax the comprehensiveness constraint applied, for example, in probabilities.

The generalization of all of the above to the combination of two elements of information poses no particular difficulty (in particular, the same types of behavior are found in CIVB operators with rules that are a little more complicated), except for non-associative operators. The main question surrounding these operators is to know in what order the elements of information should be combined. Several situations can occur:

– in certain applications, each element of information has to be combined with the others as soon as it becomes available (for example, in order to make partial decisions based on the data available at every instant): the order is then set by the order in which the elements of information arrive;

– the order can be imposed by priorities on the information to take into account, and operators have been designed to respond to these needs (for example, in order to combine database requests);

– in other situations, criteria have to be determined for finding an order adapted to the application, particularly when the elements of information are in conflict, since the results can be very different depending on whether the consonant or the conflicting elements of information are combined first.

Finally, the study of the behaviors of operators in terms of the quality of the decision they lead to and of their reactions when faced with conflicting situations leads to a final criterion for choice. An important point, however, involves the discriminating power of the operators. Highly conjunctive or disjunctive operators (for example, the Lukasiewicz t-norm and t-conorm) quickly saturate at 0 or 1 and therefore are often poorly discriminatory. For example, with the t-conorm $F(a, b) = \min(a + b, 1)$, we have $F(0.5, 0.5) = 1$, $F(0.1, 0.9) = 1$, or also $F(0.8, 0.8) = 1$, whereas these three situations have quite different interpretations.

The ability of operators to combine information that is quantitative (numerical) or qualitative (for which only the order is known) can also be a criterion for choice. For example, the min, the max and any rank filter are useful in this regard since they can combine both types of information. This is because the calculation of $\min(x, y)$, for example, only requires knowing an order between x and y, but does not require their numerical values to be known. Additionally, ordinal operations are imposed if we want them to remain invariant by an increasing transformation of the membership degrees [DUB 99].

8.11. Decision

The major rule used in fuzzy fusion is the maximum degree of membership:

$$x \in C_i \text{ if } \mu_i(x) = \max\left\{\mu_k(x),\ 1 \leq k \leq n\right\}, \qquad [8.106]$$

where μ_k refers to the membership function to the class k resulting from the combination.

The quality of the decision is measured basically according to two criteria:

– the first involves the "crispness" of the decision: the maximum degree of membership (or more generally the one that corresponds to the decision) is compared to a threshold, which is chosen depending on the applications (and possibly depending on the chosen combination operator);

– the second involves the "discriminating" nature of the decision, which is evaluated by comparing the two highest values.

If these criteria are not met for an element x, then this element is placed in a rejection class, or reclassified according to other criteria, such as spatial criteria, for example (see Chapter 9).

8.12. Application examples

In this section, we will illustrate fuzzy methods with two examples of multi-source classification.

8.12.1. *Example in satellite imagery*

For the first example, we go back to the SPOT images from the example in Chapter 6 (Figure 6.1). The classes that are considered are still cities or urban areas (class C_1), rivers (class C_2) and a class C_3 encompassing all the other structures (mostly vegetated areas). This example was discussed in [CHA 95].

First, a supervised learning phase is conducted based on the histograms conditional to the classes, either simply by smoothing these histograms, or by minimizing the distance to the histograms of parametric functions such as truncated Gaussian distributions or piecewise linear L-R functions (trapezoidal functions). This first phase is illustrated in the first line of Figure 8.13.

However, these functions, denoted by f_i^j for learning the class C_i in the image j, have no satisfactory interpretation in terms of membership functions. Particularly, the tails of the histogram correspond to gray levels that are rare in the images, but whose corresponding points belong without ambiguity to the darkest class (the lightest, respectively). The change from the functions f_i^j to the membership functions μ_i^j is done by a transformation such that:

$$\mu_i^j(x) = \lambda_i^j(x) f_i^j(x) \quad [8.107]$$

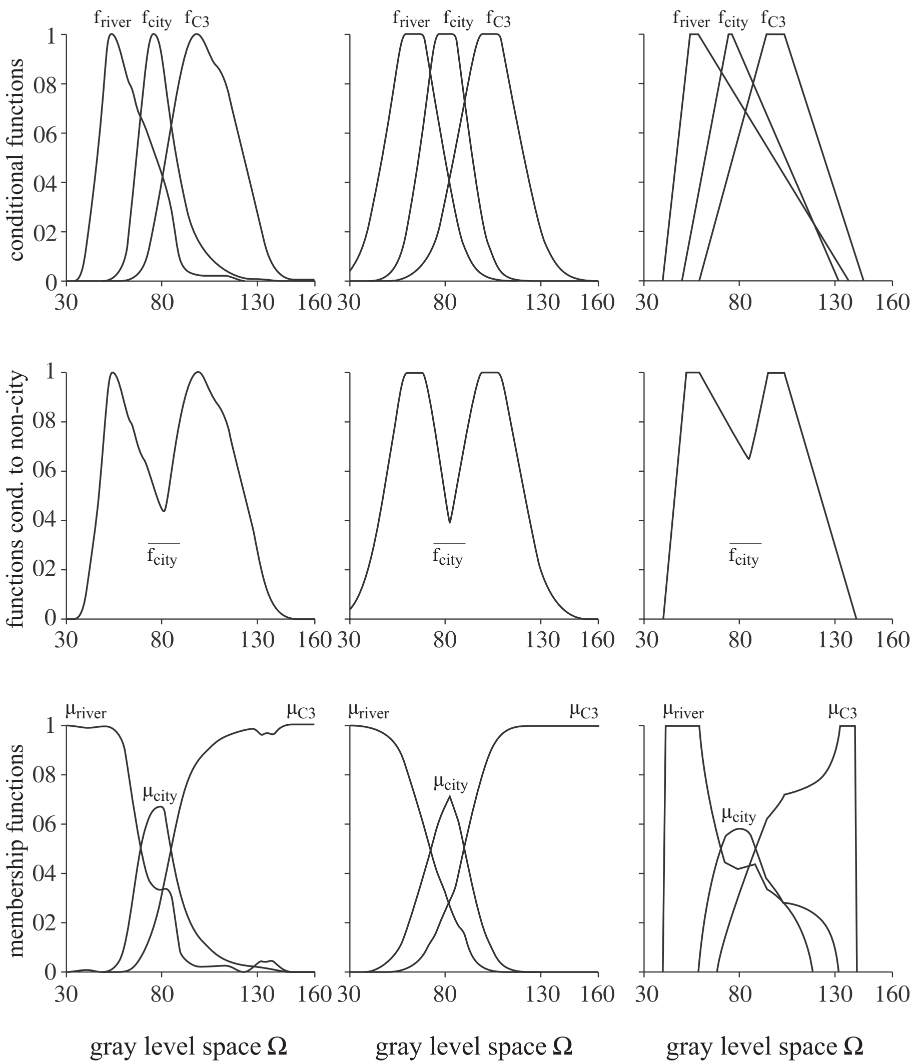


Figure 8.13. *Top: the functions f_i^j conditional to the classes that were estimated under supervision on the learning areas of the images; in the center, the functions $\overline{f_{city}^j}$ estimated from the maxima of the functions f_{river}^j and $f_{C_3}^j$; bottom: the membership functions μ_i^j. Left: estimation from the normalized and smoothed histograms; center, estimation from the truncated Gaussian functions; right: estimation of the functions from the linear L-R functions*

where x denotes the gray level in a point and the λ_i^j are functions determined so that the complement of the classes is defined by the same formula and so that the order in which the functions are in is untouched. By choosing, for example, the simplest complementation $c(a) = 1 - a$ for $a \in [0, 1]$, and by stating that the complement of a class C_i is also the union of the classes C_k for $k \neq i$ (in a closed world), we get:

$$\lambda_i^j(x) = \frac{1}{\overline{f_i^j}(x) + f_i^j(x)} \qquad [8.108]$$

with:

$$\overline{f_i^j}(x) = \max_{k \neq i} f_k^j(x). \qquad [8.109]$$

The results obtained by this method are illustrated in Figure 8.13 and show that the resulting membership functions present the behavior we expected.

For fusion, an adaptive operator is defined by the combination of a t-norm t and a t-conorm T, as $t^{1-\gamma}T^{\gamma}$. This operator, sometimes referred to as the compensation operator, can vary between the t-norm for $\gamma = 0$ and the t-conorm for $\gamma = 1$. The operator defined in [CHA 95] is special because γ is defined locally, in each point, as a function $H^i(x)$ which is a function of a normalized conjunction of the membership degrees of x. This function is illustrated in Figure 8.14 and its values increase with the class membership. The t-norm used here is a min and the t-conorm is a max.

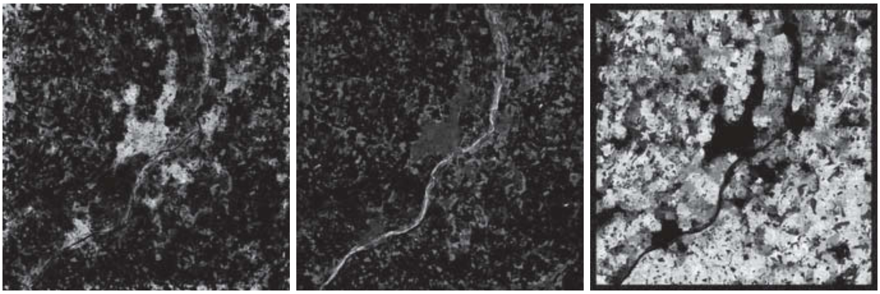


Figure 8.14. *Images of the parameters H^i of the classes of cities on the left, rivers in the center and C_3 on the right*

Figure 8.15 illustrates the results obtained with this operator on a detail of the image (bottom line), which can favorably be compared to those obtained with a simple t-norm. The decision is made by a maximum of membership degrees in each point.

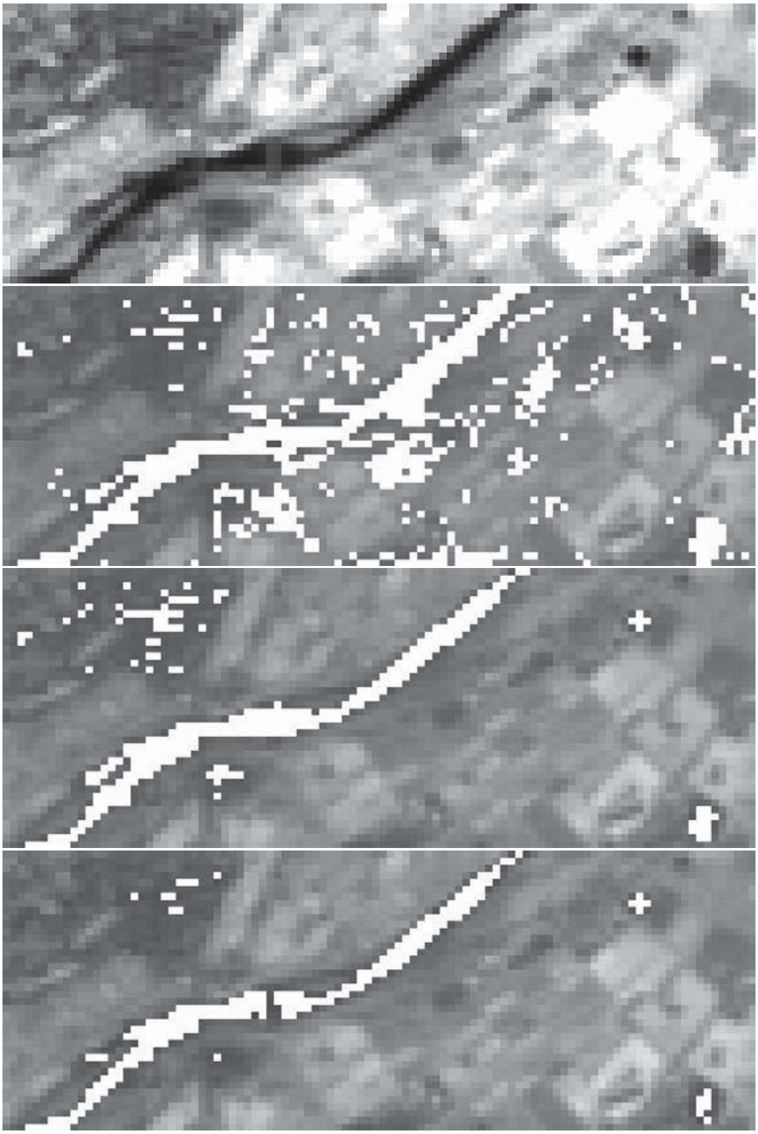


Figure 8.15. *Detail of the river detection images of the Vignola scene: original channel 3 SPOT image, decision based on the images H^i, on the constant behavior t-norms, on the variable behavior compensation operator*

Figure 8.16 shows the result of the fusion for the entire image with the adaptive operator, since the decision is still simply made by a maximum of the membership degrees. The river, as well as the contours of urban areas, are highlighted in white in the original image. The rest corresponds to the class C_3.

Figure 8.16. *Result of the fusion using the variable behavior compensation operator on the Vignola scene*

8.12.2. *Example in medical imaging*

The second example uses the same images as the example in Chapter 7 regarding the classification of multi-echo MRI images of the brain (Figure 7.1). We are still looking at the pixel level and the membership functions are defined based on the gray levels, in an unsupervised fashion [BLO 97]. Three classes are considered: brain, ventricles and cerebrospinal fluid, and the pathological area.

Here the combination operators are chosen in an adaptive fashion, not locally in each point, but depending on the classes. Since the two images provide similar information about the ventricles, the membership functions are fused by a mean operator.

The healthy brain and the pathology cannot be distinguished in the first image and therefore only one class can be learned, denoted by μ_c^1. In the second image, it is possible to learn two classes μ_c^2 (brain) and μ_{path}^2 (pathology). The functions μ_c^1 and μ_c^2 are also combined using an arithmetic mean. For the pathology, μ_c^1 and μ_{path}^2 are combined using a symmetric sum defined by: $ab/(1-a-b+2ab)$. This ensures that the pathology is not detected in the areas where $\mu_{path}^2 = 0$ and otherwise reinforces the membership to that class, thus making it possible to include in the pathological area the entire partial volume area where the points are comprised in part of pathological tissue.

After the combination, the decision is made based on the maximum membership. The result is illustrated in Figure 8.17.

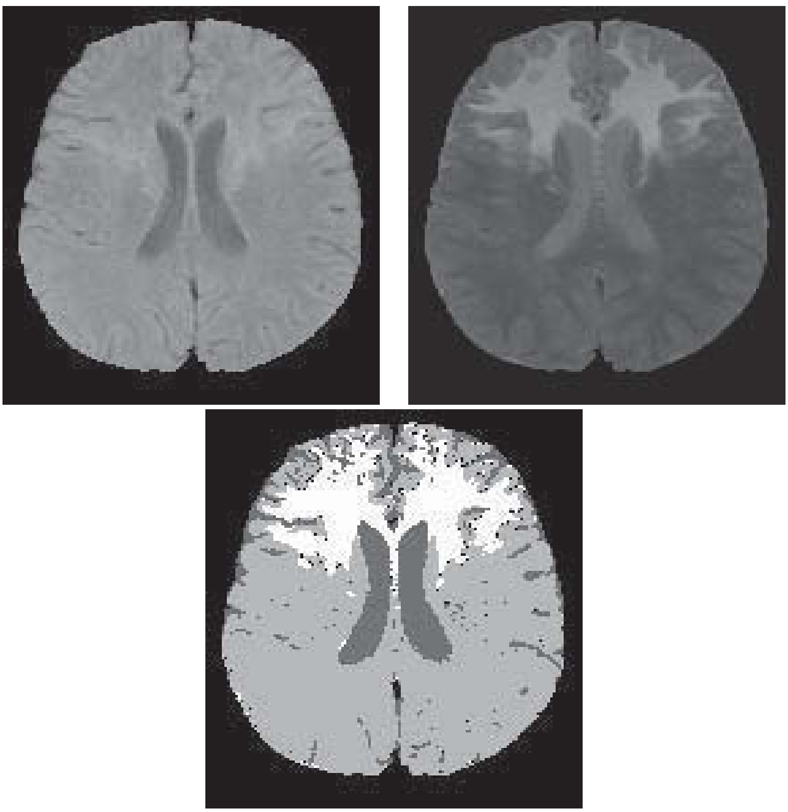


Figure 8.17. *Two MRI acquisitions of the brain and the result of the fuzzy fusion classification (the decision is only made locally in each point, without spatial regularization)*

This example illustrates the advantage of choosing operators in an adaptive fashion, based on information provided by the images about the different classes.

8.13. Bibliography

[BAN 78] BANON G., "Distinction entre plusieurs sous-ensembles de mesures floues", *Colloque International sur la Théorie des Ensembles Flous*, Marseille, France, 1978.

[BEZ 81] BEZDEK J.C., *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, New York, 1981.

[BEZ 92] BEZDEK J.C., PAL S.K., *Fuzzy Models for Pattern Recognition*, IEEE Press, New York, 1992.

[BEZ 99] BEZDEK J.C., KELLER J., KRISHNAPURAM R., PAL N.R., *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, Handbooks of Fuzzy Sets series, Kluwer Academic Publisher, Boston, 1999.

[BLO 96a] BLOCH I., "Image Information Processing using Fuzzy Sets", *World Automation Congress, Soft Computing with Industrial Applications*, Montpellier, France, p. 79-84, May 1996.

[BLO 96b] BLOCH I., "Information Combination Operators for Data Fusion: A Comparative Review with Classification", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 26, no. 1, p. 52-67, 1996.

[BLO 97] BLOCH I., AURDAL L., BIJNO D., MÜLLER J., "Estimation of Class Membership Functions for Grey-Level Based Image Fusion", *ICIP'97*, vol. III, Santa Barbara, California, p. 268-271, October 1997.

[BLO 99] BLOCH I., "On Fuzzy Distances and their Use in Image Processing under Imprecision", *Pattern Recognition*, vol. 32, no. 11, p. 1873-1895, 1999.

[BOU 96] BOUCHON-MEUNIER B., RIFQI M., BOTHOREL S., "Towards General Measures of Comparison of Objects", *Fuzzy Sets and Systems*, vol. 84, no. 2, p. 143-153, September 1996.

[CHA 95] CHAUVIN S., Evaluation des théories de la décision appliquées à la fusion de capteurs en imagerie satellitaire, PhD Thesis, Ecole Nationale Suérieure des Télécommunications and Nantes University, December 1995.

[CHE 89] CHENG Y., KASHYAP R.L., "A Study of Associative Evidential Reasoning", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-11, no. 6, p. 623-631, 1989.

[CHE 95] CHENG H.D., CHEN J.R., "Automatically Determine the Membership Function based on the Maximum Entropy Principle", 2^{nd} *Annual Joint Conf. on Information Sciences*, Wrightsville Beach, USA, p. 127-130, 1995.

[CIV 86] CIVANLAR M.R., TRUSSEL H.J., "Constructing Membership Functions using Statistical Data", *Fuzzy Sets and Systems*, vol. 18, p. 1-13, 1986.

[COX 46] COX R.T., "Probability, Frequency and Reasonable Expectation", *Journal of Physics*, vol. 14, no. 1, p. 115-137, 1946.

[DEL 87] DELGADO M., MORAL S., "On the Concept of Possibility-Probability Consistency", *Fuzzy Sets and Systems*, vol. 21, no. 3, p. 311-318, 1987.

[DEL 92] DELLEPIANE S., VENTURI G., VERNAZZA G., "Model Generation and Model Matching of Real Images by a Fuzzy Approach", *Pattern Recognition*, vol. 25, no. 2, p. 115-137, 1992.

[DEV 85] DEVI B.B., SARMA V.V.S., "Estimation of Fuzzy Memberships from Histograms", *Information Sciences*, vol. 35, p. 43-59, 1985.

[DEV 93] DEVEUGHELE S., DUBUISSON B., "Using Possibility Theory in Perception: An Application in Artificial Vision", *Second IEEE International Conference on Fuzzy Systems*, San Francisco, California, p. 821-826, 1993.

[DUB 80] DUBOIS D., PRADE H., *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, 1980.

[DUB 83] DUBOIS D., PRADE H., "Unfair Coins and Necessity Measures: Towards a Possibilistic Interpretation of Histograms", *Fuzzy Sets and Systems*, vol. 10, no. 1, p. 15-20, 1983.

[DUB 85] DUBOIS D., PRADE H., "A Review of Fuzzy Set Aggregation Connectives", *Information Sciences*, vol. 36, p. 85-121, 1985.

[DUB 88] DUBOIS D., PRADE H., *Possibility Theory*, Plenum Press, New York, 1988.

[DUB 91] DUBOIS D., LANG J., PRADE H., "Fuzzy Sets in Approximate Reasoning, Part II: Logical Approaches", *Fuzzy Sets and Systems*, vol. 40, p. 203-244, 1991.

[DUB 92a] DUBOIS D., PRADE H., "Combination of Information in the Framework of Possibility Theory", in A. MONGI and A. ABIDI (ed.) *Data Fusion in Robotics and Machine Intelligence*, Academic Press, 1992.

[DUB 92b] DUBOIS D., PRADE H., "A Unifying View of Comparison Indices in a Fuzzy Set-Theoretic Framework", in R.R. YAGER (ed.) *Fuzzy Sets and Possibility Theory*, p. 3-13, Pergamon Press, 1992.

[DUB 99] DUBOIS D., PRADE H., YAGER R., "Merging Fuzzy Information", in J. BEZDEK, D. DUBOIS and H. PRADE (ed.) *Handbook of Fuzzy Sets Series, Approximate Reasoning and Information Systems*, Chapter 6, Kluwer, 1999.

[FRA 79] FRANK M.J., "On the Simultaneous Associativity of $F(x, y)$ and $x + y - F(x, y)$", *Aequationes Mathematicae*, vol. 19, p. 194-226, 1979.

[GOE 83] GOETSCHEL R., VOXMAN W., "Topological Properties of Fuzzy Numbers", *Fuzzy Sets and Systems*, vol. 10, p. 87-99, 1983.

[GOE 86] GOETSCHEL R., WOXMAN W., "Elementary Fuzzy Calculus", *Fuzzy Sets and Systems*, vol. 18, p. 31-43, 1986.

[GRA 92] GRABISCH M., MUROFUSHI T., SUGENO M., "Fuzzy Measures of Fuzzy Events Defined by Fuzzy Integrals", *Fuzzy Sets and Systems*, vol. 50, p. 293-313, 1992.

[GRA 95] GRABISCH M., "Fuzzy Integral in Multicriteria Decision Making", *Fuzzy Sets and Systems*, vol. 69, p. 279-298, 1995.

[HAM 78] HAMACHER H., "Ueber logische Verknupfungen Unscharfer Aussagen und deren Zugehoerige Bewertungsfunktionen", *Progress in Cybernetics and System Research*, vol. 3, p. 276-287, 1978.

[KAU 75] KAUFMANN A., *Introduction to the Theory of Fuzzy Subsets*, Academic Press, New York, 1975.

[KLI 92] KLIR G.J., PARVIZ B., "Probability-Possibility Transformations: A Comparison", *Int. J. General Systems*, vol. 21, p. 291-310, 1992.

[KOS 90] KOSKO B., "Fuzziness vs. Probability", *International Jounal of General Systems*, vol. 17, p. 211-240, 1990.

[KRI 92] KRISHNAPURAM R., KELLER J.M., "Fuzzy Set Theoretic Approach to Computer Vision: an Overview", *IEEE Int. Conf. on Fuzzy Systems*, San Diego, California, p. 135-142, 1992.

[KRI 93] KRISHNAPURAM R., KELLER J.M., "A Possibilistic Approach to Clustering", *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 2, p. 98-110, 1993.

[LUC 72] LUCA A.D., TERMINI S., "A Definition of Non-Probabilistic Entropy in the Setting of Fuzzy Set Theory", *Information and Control*, vol. 20, p. 301-312, 1972.

[MEN 42] MENGER K., "Statistical Metrics", *Proc. National Academy of Sciences of USA*, vol. 28, p. 535-537, 1942.

[MUR 89] MUROFUSHI T., SUGENO M., "An Interpretation of Fuzzy Measure and the Choquet Integral as an Integral with respect to a Fuzzy Measure", *Fuzzy Sets and Systems*, vol. 29, p. 201-227, 1989.

[PAL 92] PAL S.K., "Fuzzy Set Theoretic Measures for Automatic Feature Evaluation", *Information Science*, vol. 64, p. 165-179, 1992.

[SCH 63] SCHWEIZER B., SKLAR A., "Associative Functions and Abstract Semigroups", *Publ. Math. Debrecen*, vol. 10, p. 69-81, 1963.

[SCH 83] SCHWEIZER B., SKLAR A., *Probabilistic Metric Spaces*, North Holland, Amsterdam, 1983.

[SHA 76] SHAFER G., *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[SHO 75] SHORTLIFFE E.H., BUCHANAN B.G., "A Model of Inexact Reasoning in Medicine", *Mathematical Biosciences*, vol. 23, p. 351-379, 1975.

[SUG 74] SUGENO M., Theory of Fuzzy Integral and its Applications, PhD Thesis, Tokyo Institute of Technology, Tokyo, 1974.

[YAG 79] YAGER R.R., "On the Measure of Fuzziness and Negation", *International Journal of General Systems*, vol. 5, p. 221-229, 1979.

[YAG 80] YAGER R.R., "On a General Class of Fuzzy Connectives", *Fuzzy Sets and Systems*, vol. 4, p. 235-242, 1980.

[YAG 88] YAGER R.R., "On Ordered Weighted Averaging Aggregation Operators in Multi-Criteria Decision Making", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 18, no. 1, p. 183-190, 1988.

[YAG 91] YAGER R.R., "Connectives and Quantifiers in Fuzzy Sets", *Fuzzy Sets and Systems*, vol. 40, p. 39-75, 1991.

[ZAD 65] ZADEH L.A., "Fuzzy Sets", *Information and Control*, vol. 8, p. 338-353, 1965.

[ZAD 75] ZADEH L.A., "The Concept of a Linguistic Variable and its Application to Approximate Reasoning", *Information Sciences*, vol. 8, p. 199-249, 1975.

[ZAD 78] ZADEH L.A., "Fuzzy Sets as a Basis for a Theory of Possibility", *Fuzzy Sets and Systems*, vol. 1, p. 3-28, 1978.

[ZAD 96] ZADEH L.A., "Fuzzy Logic = Computing with Words", *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 2, p. 103-111, 1996.

[ZIM 80] ZIMMERMANN H.J., ZYSNO P., "Latent Connectives in Human Decision Making", *Fuzzy Sets and Systems*, vol. 4, p. 37-51, 1980.

[ZIM 91] ZIMMERMANN H.J., *Fuzzy Set Theory and its Applications*, Kluwer Academic Publisher, Boston, 1991.

Chapter 9

Spatial Information in Fusion Methods

Spatial information is fundamental in image processing. Including it in fusion methods is crucial and often requires specific developments to adapt the methods used in other fields. One of the most common objectives of these developments it to ensure that the decision is spatially consistent. For example, in multi-source classification, the goal will be to avoid those points which are isolated or scattered in a homogenous class to be assigned to a different class.

9.1. Modeling

Spatial information on the modeling level is generally implicit depending on what level of representation is chosen. If we are reasoning on a pixel level, the information contained in a pixel does not include any spatial information, so this information will have to be added explicitly. The spatial context that is considered is most often the local neighborhood of each point. A simple way of taking it into account is to define the measure $M_i^j(x)$ (see Chapter 1) based on the characteristics of x and of its neighbors also. If we denote by $\mathcal{V}(x)$ the neighborhood of x (containing x), we will define $M_i^j(x)$ as a function of the type:

$$M_i^j(x) = F_i\big[f_j(y), y \in \mathcal{V}(x)\big], \qquad [9.1]$$

where $f_j(y)$ refers to the characteristics of y in the source j. This type of approach can be seen as a spatial filtering problem. In the case of linear filtering, F is expressed as a convolution and the convolution kernel defined on $\mathcal{V}$ is typically a Gaussian function or a rectangular window. If the filtering is not linear, many solutions are suggested

Chapter written by Isabelle BLOCH.

in image processing [MAI 02], the most frequent of which are the median filter, the sigma filter, or morphological filtering. Finally, more elaborate techniques use relaxation, such as Markov fields, which operate either on the level of the measure (this is referred to as restoration), or on the level of the classification, as we will see later on.

If we work on the level of primitives (segments, contours, areas) or on the level of objects or structures in the scene, the local spatial information is implicitly accounted for in the representation. If the detection of these elements or their localization are not precise (for example, because of the imperfection of the registration), it is often advisable to explicitly include this spatial imprecision in the representation, before the fusion. Fuzzy dilation is an operation well suited for this purpose [BLO 95, BLO 96, BUS 00]. This allows the conflict to be reduced to the moment when the fusion takes place and hence to choose a conjunctive combination mode simply and without risk.

In a less local fashion, the spatial relations between primitives constitute important information regarding the structure of the scene [BLO 97, BLO 99a, BLO 99b, BLO 99c, BLO 00b] and they can taken advantage of in fusion, as a source of additional information [BLO 00a, BLO 00c, GER 99]. In this case, the spatial context $\mathcal{V}(x)$ of an element x is a set of primitives or objects whose spatial relations with respect to x are known.

9.2. The decision level

The inclusion of spatial information on the decision level is the easiest. The most common method consists of first establishing a rejection rule (depending on the crispness and the discriminating nature of the decision) then reclassifying the rejected elements according to their spatial context. For example, reclassification can be performed according to the following rule (absolute majority):

$$x \in C_i \text{ if } \big|\{y \in \mathcal{V}(x),\ y \in C_i\}\big| \geq \frac{|\mathcal{V}|}{2} \qquad [9.2]$$

which expresses that at least half of the elements of the neighborhood have to be in C_i in order to put x in C_i. This rule does not always allow x to be assigned to a class. A less severe rule only considers the most represented class in the neighborhood (majority rule):

$$x \in C_i \text{ if } \big|\{y \in \mathcal{V}(x),\ y \in C_i\}\big| = \max_k \big|\{y \in \mathcal{V}(x),\ y \in C_k\}\big|. \qquad [9.3]$$

These rules apply regardless of the level of representation of the elements considered.

An example of fuzzy classification can be found in [BOU 92], but this is a general method, which can be applied in a similar fashion to other theories.

9.3. The combination level

The inclusion of spatial information on the combination is less common and more difficult.

In probabilistic fusion, Markov fields offer a natural framework for this purpose. In the expression of Bayes' rule, the Markovian hypothesis is involved in the *a priori* probability. This probability is combined with the probabilities conditional to the classes by way of a product. This comment leads us to consider that spatial information, in this model, constitutes a source of data like any other.

This is the most common approach and it has been applied on several levels of representation. On the local, pixel level, many examples can be found in other works (for example, [AUR 97b, DES 96]). On a more structural level, Markov fields are defined on graphs that are more general than the pixel graphs (the nodes are primitives or even objects) and examples can be found for road detection in SAR images [TUP 98], for the segmentation of MRI images of the brain [GER 95], for recognizing structures of the cerebral cortex [MAN 95, MAN 96], for the interpretation of aerial images [MOI 95], etc.

With other theories, it would also be possible to develop similar approaches, with spatial information still considered as an additional source of data.

This is, for example, the case of spatial relations mentioned above considered as an additional source of information: recognizing an object can be the result of the fusion of information regarding that object and of information regarding the relations it has to have with respect to other objects. The fuzzy set framework allows both the representation and the fusion of such information [BLO 00c].

We should mention another example: in [HEG 98], a mass function is defined for representing the spatial context and combined with mass functions representing the information extracted from the images according to Dempster's rule. However, there are still few studies in this field, which certainly deserves to be developed further.

9.4. Application examples

9.4.1. *The combination level: multi-source Markovian classification*

Let us consider the same example of the fusion of MRI images of the brain from Chapters 7 and 8, with the objective of segmenting the healthy brain, the pathology and the ventricles, this time using a Markovian approach. This method was developed in [AUR 95, AUR 97a, AUR 97b].

Bayes' rule makes it possible to calculate the *a posteriori* probability of each class conditionally to the two images. The *a priori* probability term is modeled using a Markovian hypothesis regarding the image of the classes and acts as a spatial regularization. Therefore, the *a posteriori* probability is expressed as the product of three terms: two terms expressing the probabilities of the gray levels in each of the images conditionally to the classes (under the conditional independence hypotheses) and a term expressing the spatial regularities of the classes. The Markovian framework allows us to express the problem of the *a posteriori* maximum optimization as the minimization of an energy that includes:

1) data-based terms, that depend on the gray levels of each image, on coefficients weighting the importance of each image according to the classes and on prior knowledge of the positions of the ventricles in the brain;

2) a regularization term, in the form of a Potts potential, that takes into account the neighboring pixels of each point.

Therefore, the energy is written in each point x assigned to the class C_i in the current step:

$$\Phi_i^1\big(f_1(x)\big) + \Phi_i^2\big(f_2(x)\big) + \lambda \sum_{y \in \mathcal{V}(x)} \omega\big(C_i, C_y\big) \qquad [9.4]$$

where Φ_i^1 (Φ_i^2) represents the data-based potential characterizing the class C_i in the first (second) image, which is a function of the gray level $f_1(x)$ ($f_2(x)$) at the point x in this image, $\mathcal{V}(x)$ represents the spatial neighborhood of x, C_y the class to which the neighbor y is assigned in the current iteration and $\omega(C_i, C_y)$ represents the regularization constraints between the classes C_i and C_y. The factor λ makes it possible to weight the influence of the regularization with respect to the data-based term. The data-based potentials are determined automatically based on histograms of gray levels, whose significant modes are selected using a multi-scale approach. Here, the regularization is simple: it favors the membership to the same class as the neighboring points ($\omega(C_i, C_i) = 0$) and puts at a disadvantage the membership of neighboring points to different classes ($\omega(C_i, C_y) = 1$ if $C_y \neq C_i$).

The results (see Figure 9.1) show the spatial homogenity of the obtained segmentation. The spatial information used in this example is still relatively local, since it only involves a small neighborhood around each point.

9.4.2. *The modeling and decision level: fusion of structure detectors using belief function theory*

In this example, developed in [TUP 99], the objective is to interpret a radar image by fusing the results of several structure detectors (roads, slopes, cities, etc.). The

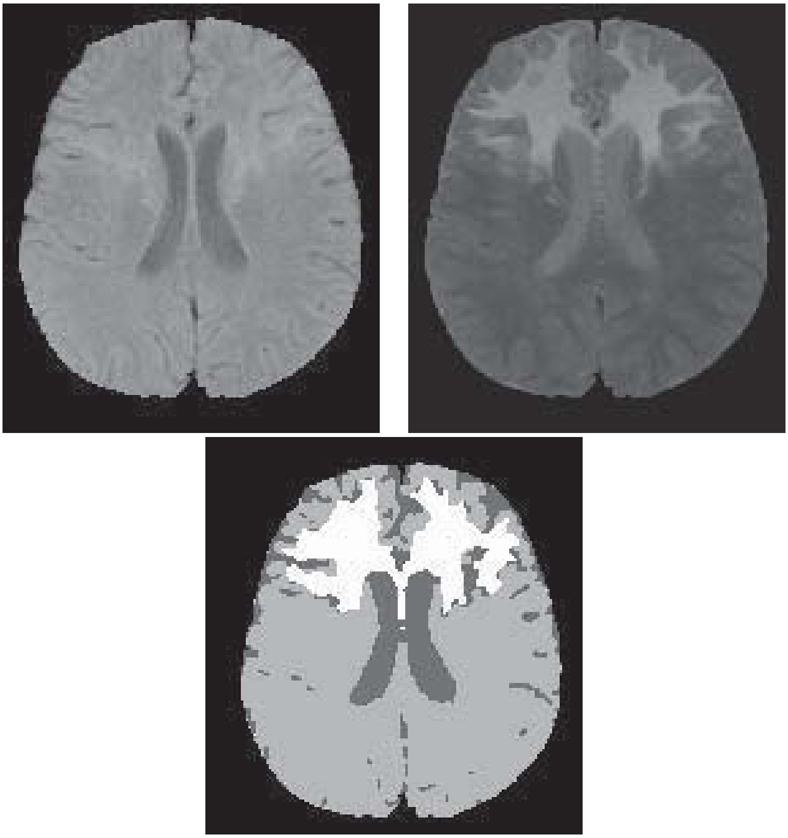


Figure 9.1. *Two MRI acquisitions of the brain and the result of the classification by Markovian fusion*

fusion is performed on an intermediate level, on primitives obtained by these detectors (segments or areas), and therefore which contain spatial information.

Figure 9.2 illustrates an example of an ERS-1 radar image of the Aix-en-Provence region.

Figure 9.3 shows the results of three of the detectors applied to this image.

The results of these detectors serve as the basis for modeling using belief functions. This theory is particularly well-suited here, since it allows us to model the behavior of each detector. Thus, road, slope or river detectors are precise for the objects they are designed to study, but are not capable of distinguishing the other classes. As a result, the focal elements of a road detector are the class "road" and the set "non-road" (hence

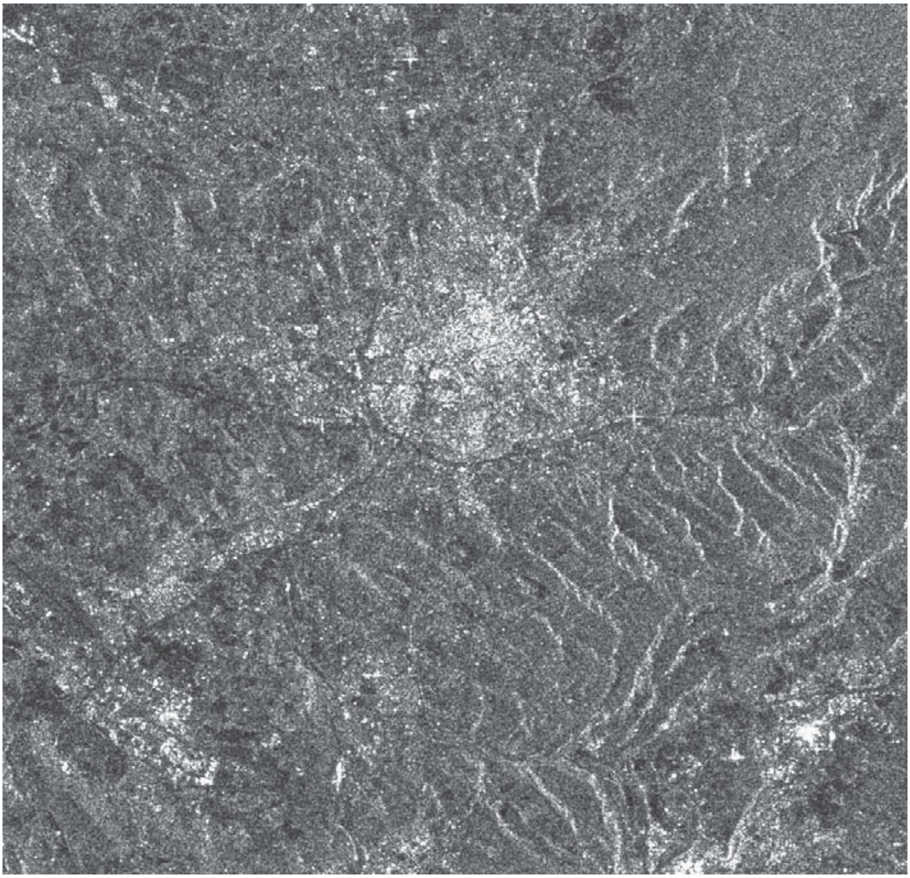


Figure 9.2. *ERS-1 radar image of Aix-en-Provence region*

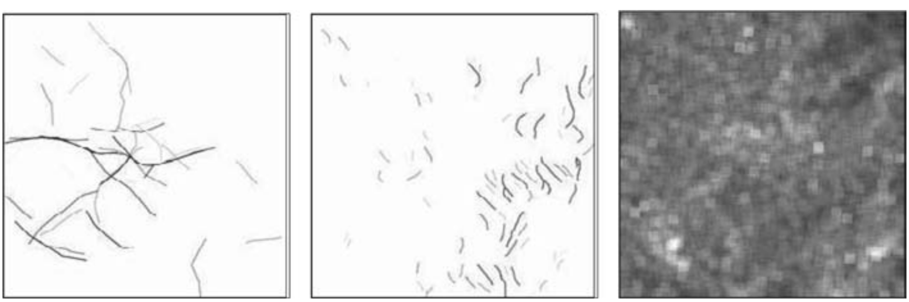


Figure 9.3. *Results of three detectors: roads, slopes and homogenous areas. The gray levels indicate the confidence in the result provided by each detector*

the union of all the other classes considered). On the contrary, the homogenous area detector is not very precise because it uses filters with size 9×9 windows, which erases all of the roads that cut through homogenous areas. It is then natural to choose as the focal elements the set "roads $\cup$ homogenous areas" and the set "non-homogenous". This is an example of a detector for which no focal element is a singleton. Urban area detectors are modeled in a similar fashion. This example shows that knowledge about the detectors and their behaviors is what allows us to conduct the modeling phase and choose the focal elements.

The mass functions are then learned from the response histograms of the various detectors, by minimizing a distance between these histograms and trapezoidal parametric functions.

The fusion is then performed according to Dempster's non-normalized orthogonal rule (conjunctive fusion) since all the imprecisions and ambiguities regarding the detectors are explicitly taken into account in the modeling. This makes it possible to reduce the focal elements to singletons or union of two classes only. Furthermore, this is a typically open world application: it is not possible to predict all the classes that may show up in the image and only those for which detectors have been designed can be detected. The non-normalized combination allows us to represent in the mass of the empty set anything that is not predicted.

Finally, the decision phase is conducted in a Markovian framework, ensuring the addition of spatial consistency between the areas, hence an additional level of spatial information. The pignistic probabilities (see Chapter 7) make it possible to go back to probabilities for singletons, which are then combined to a spatial regularization term.

The result of an interpretation is shown in Figure 9.4.

9.4.3. *The modeling level: fuzzy fusion of spatial relations*

In this last example, the spatial information we are considering is structural information, involving no longer the local consistency of the classes or areas, but instead the relations between the objects we are looking for. The application involves recognizing internal structures of the brain in MRI images, using an anatomical atlas as our guide [BLO 00c, GER 00].

A cross-section extracted from the 3-D volume of the atlas is shown in Figure 9.5; the view of the corresponding cross-section in the 3-D MRI acquisition which needs to be processed is represented in Figure 9.6.

The recognition is performed progressively, with one structure detected at each step. Each step relies on the objects obtained in the previous steps and on various kinds

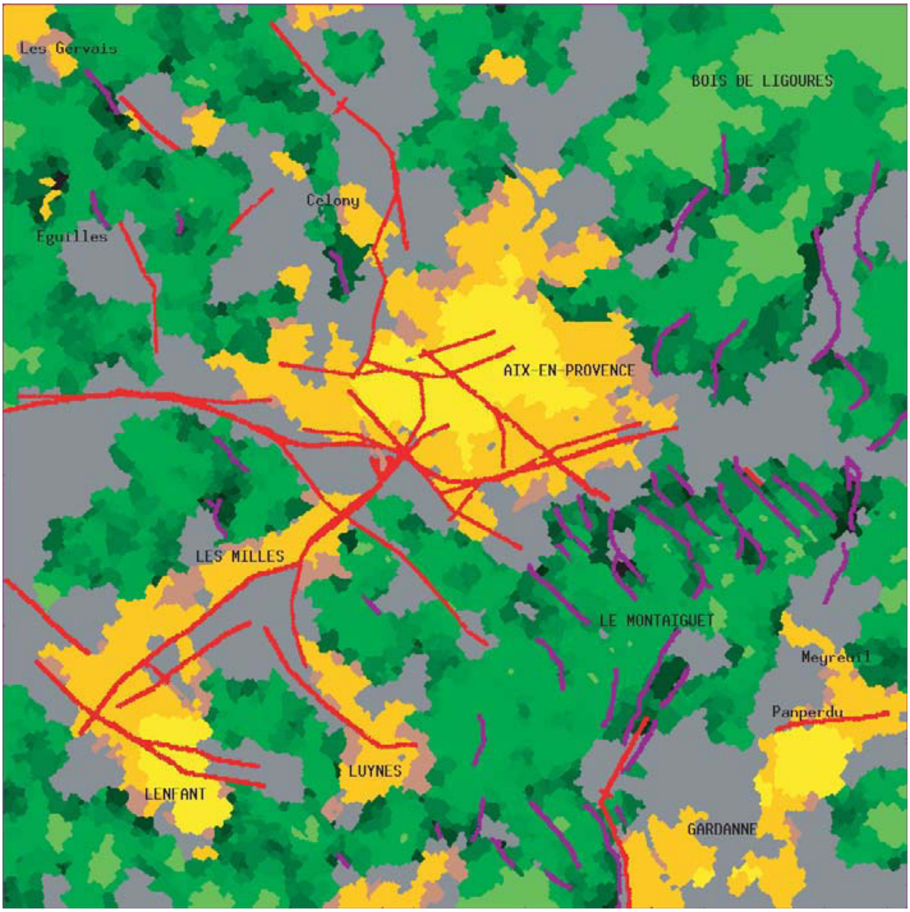


Figure 9.4. *Result obtained from the image in Figure 9.2 by fusing detectors of different structures. The lighter areas represent urban areas. The dark lines correspond to roads, the lighter lines to slopes. The darkest areas (with different shades) are forest areas. Finally, the medium gray areas correspond to the empty set, i.e. unrecognized areas*

of elements of anatomical knowledge. The localization and morphology information of this object are provided by the atlas and symbolic information about this object is expressed with respect to objects identified in the previous steps. This symbolic information concerns spatial relations (pertaining to sets, directions or distances) as well as constitution information (gray matter, white matter or liquid) or radiometric knowledge related to the type of imagery. The fuzzy set framework was chosen for several reasons: it allows us, using a single formalism, to express elements of information with different semantics, of which there can be many for this problem. It also allows us to model the imprecision and the uncertainty, which is particularly important in this field

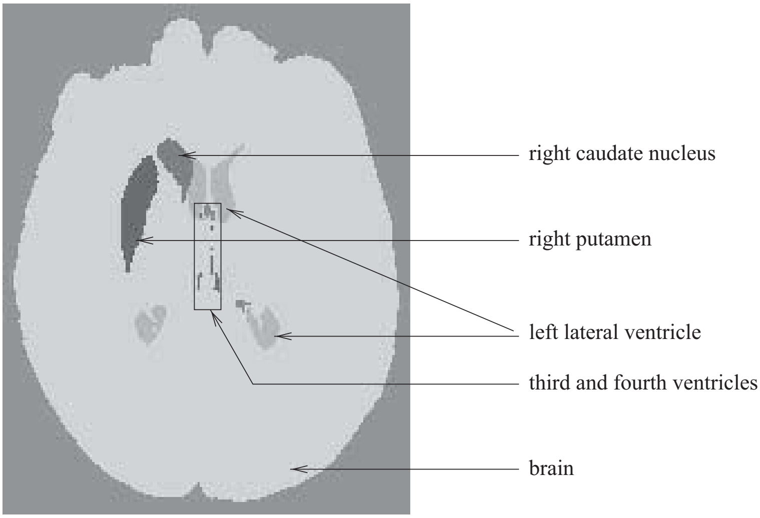


Figure 9.5. *A cross-section of the atlas representing the structures we are trying to recognize in the image*

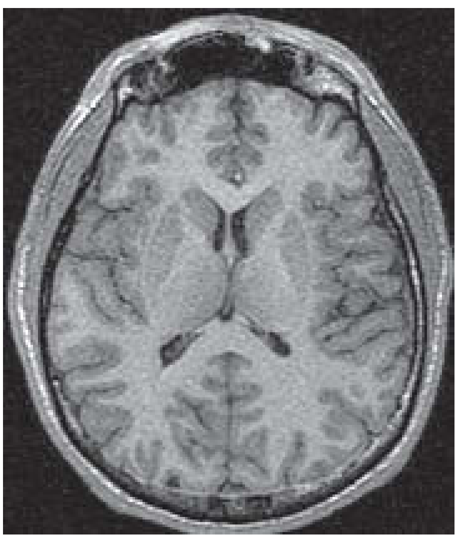


Figure 9.6. *A cross-section of the cerebral volume we wish to recognize in a plane close to that of the atlas in the previous figure. This is an MRI image*

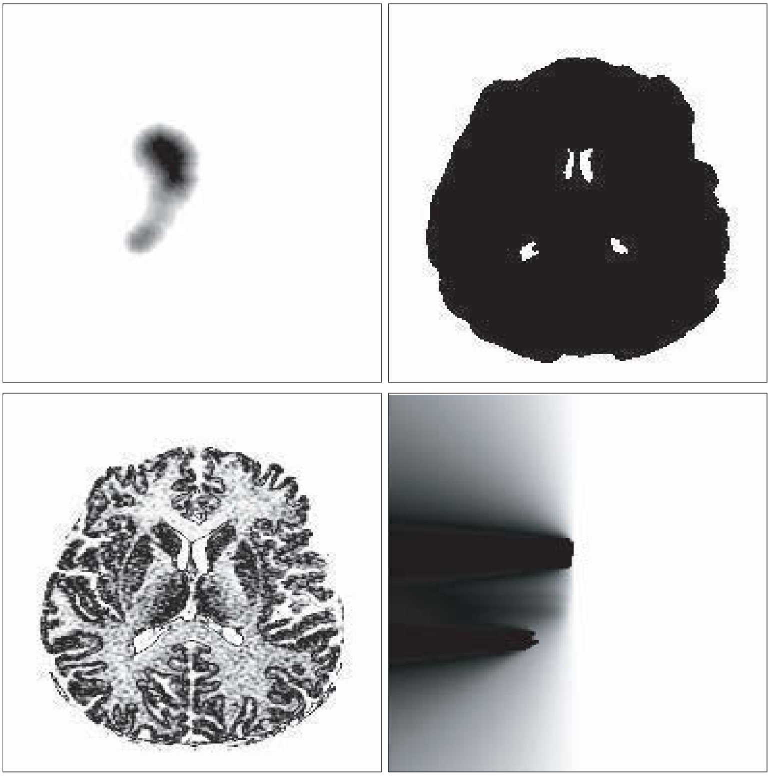


Figure 9.7. *Information expressed in the image space. White and black correspond to minimum and maximum membership values, respectively. Top left: information on the location and the approximate shape provided by the atlas. Top right: binary localization constraint expressing the fact that the caudate nucleus is located in the brain (search area, in black) and outside the lateral ventricles in white. Bottom left: prior radiometric knowledge. Bottom right: relative directional relation "to the left of the lateral ventricles"*

because of the anatomical variability among individuals. Finally, fuzzy fusion which leads to recognition can take advantage of the many operators which make it possible to adequately model the relations we know between these elements of information.

Modeling is achieved by representing, in the image space, each element of information by a fuzzy set. Fuzzy areas of interest are thus defined for each type of knowledge and their fusion allows us to focus our research in a increasingly limited area. We will illustrate this process with the recognition phase of the caudate nucleus. At this phase of the recognition procedure, three objects have been segmented: the brain and the two lateral ventricles. Figure 9.7 illustrates the representation of the knowledge related to the caudate nucleus (information regarding the location and the approximate shape provided by the atlas, location in the brain and outside the lateral ventricles, radiometry and relative position to the left of the lateral ventricles). The formal definition of these representations can be found in [BLO 00b, BLO 00c].

The recognition of certain structures uses information about distance, which is usually represented by fuzzy sets in the image space. Figure 9.8 illustrates this type of representation for three types of knowledge.

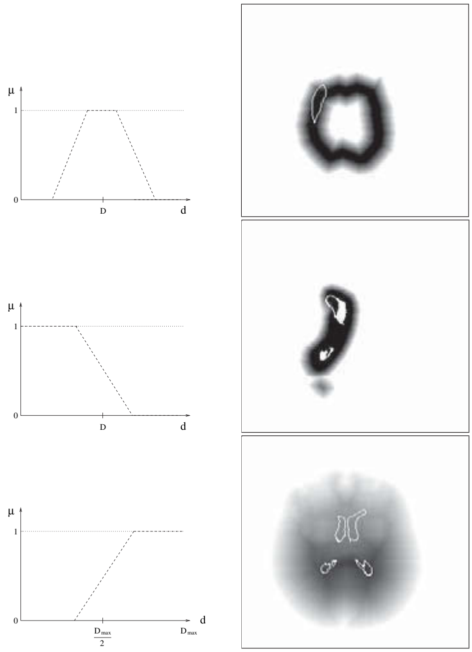


Figure 9.8. *Examples of the representation of knowledge about distances. Left: representation in the form of fuzzy intervals in $\mathbb{R}^+$. Right: the corresponding spatial representations. The putamen is located at an approximately constant distance from the surface of the brain (top), the caudate nucleus at a distance smaller than roughly D from the lateral ventricles (represented in white) (middle), the lateral ventricles are in the brain at a distance greater than roughly $D_{\max}/2$ from the surface of the brain (bottom). The contours of the objects we are looking for (putamen, caudate nucleus and ventricles, respectively) are shown in white and are in fact in the areas that satisfy the relations with a high degree*

The fusion of spatial knowledge is achieved using a conjunctive operator (a t-norm) in order to reduce the area of interest as new knowledge is provided, in order to focus the search. Then these fused elements of spatial knowledge are combined to the radiometric information by a mean operator, reaching this time a compromise between these types of information. The resulting area can then be easily segmented (Figure 9.9).

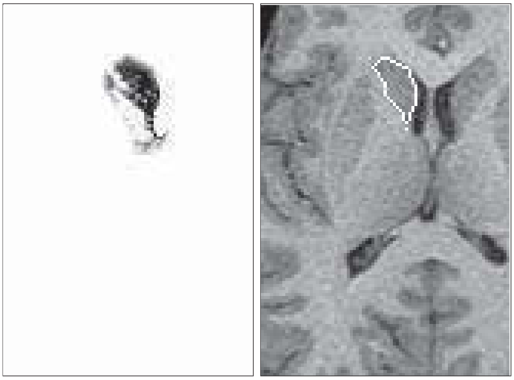


Figure 9.9. *Result of the fusion of all the knowledge available regarding the caudate nucleus (left) and the segmentation of that nucleus (right)*

Figure 9.10 illustrates six objects of the atlas as well as the same objects recognized in an MRI acquisition. They are properly segmented, even though they are significantly different in the image in size, location and morphology from those in the atlas which serves as a model. It is worth noting that the third and fourth ventricles, which are particularly difficult to segment in MRI images, here are properly recognized and segmented, thanks to the use of spatial relations with respect to other structures of the brain.

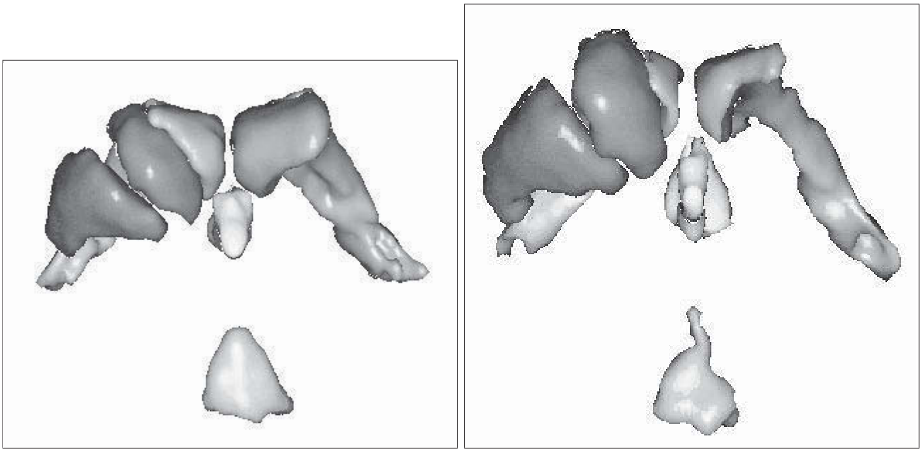


Figure 9.10. *Results of the recognition procedures. Six structures are represented: the lateral ventricles (medium gray), the third and fourth ventricles (light gray), a caudate nucleus and a putamen (dark gray). On the left, the structures from the atlas and on the right, those recognized in a 3-D MRI image*

9.5. Bibliography

[AUR 95] AURDAL L., DESCOMBES X., MAÎTRE H., BLOCH I., ADAMSBAUM C., KALIFA G., "Fully Automated Analysis of Adrenoleukodystrophy from Dual Echo MR Images: Automatic Segmentation and Quantification", *Computer Assisted Radiology CAR'95*, Berlin, p. 35-40, June 1995.

[AUR 97a] AURDAL L., Analysis of Multi-Image Magnetic Resonance Acquisitions for Segmentation and Quantification of Cerebral Pathologies, PhD Thesis, Ecole Nationale Supérieure des Télécommunications, ENST, 1997.

[AUR 97b] AURDAL L., BLOCH I., MAÎTRE H., GRAFFIGNE C., ADAMSBAUM C., "Continuous Label Bayesian Segmentation, Applications to Medical Brain Images", *ICIP'97*, vol. II, Santa Barbara, California, p. 128-131, October 1997.

[BLO 95] BLOCH I., MAÎTRE H., "Fuzzy Mathematical Morphologies: A Comparative Study", *Pattern Recognition*, vol. 28, no. 9, p. 1341-1387, 1995.

[BLO 96] BLOCH I., PELLOT C., SUREDA F., HERMENT A., "Fuzzy Modelling and Fuzzy Mathematical Morphology applied to 3D Reconstruction of Blood Vessels by Multi-Modality Data Fusion", in R. YAGER, D. DUBOIS and H. PRADE (ed.) *Fuzzy Set Methods in Information Engineering: A Guided Tour of Applications*, Chapter 5, p. 93-110, John Wiley & Sons, New York, 1996.

[BLO 97] BLOCH I., MAÎTRE H., ANVARI M., "Fuzzy Adjacency between Image Objects", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 5, no. 6, p. 615-653, 1997.

[BLO 99a] BLOCH I., "Fuzzy Relative Position between Objects in Image Processing: a Morphological Approach", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 7, p. 657-664, 1999.

[BLO 99b] BLOCH I., "Fuzzy Relative Position between Objects in Image Processing: New Definition and Properties based on a Morphological Approach", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 7, no. 2, p. 99-133, 1999.

[BLO 99c] BLOCH I., "On Fuzzy Distances and their Use in Image Processing under Imprecision", *Pattern Recognition*, vol. 32, no. 11, p. 1873-1895, 1999.

[BLO 00a] BLOCH I., "Fusion of Numerical and Structural Image Information in Medical Imaging in the Framework of Fuzzy Sets", in P. SZCZEPANIAK *et al.* (ed.) *Fuzzy Systems in Medicine*, Series Studies in Fuzziness and Soft Computing, p. 429-447, Springer Verlag, 2000.

[BLO 00b] BLOCH I., "Fuzzy Mathematical Morphology and Derived Spatial Relationships", in E. KERRE and N. NACHTEGAEL (ed.) *Fuzzy Techniques in Image Processing*, Chapter 4, p. 101-134, Springer Verlag, 2000.

[BLO 00c] BLOCH I., "Spatial Representation of Spatial Relationships Knowledge", in A.G. COHN, F. GIUNCHIGLIA and B. SELMAN (ed.) *7th International Conference on Principles of Knowledge Representation and Reasoning KR 2000*, Breckenridge, CO, Morgan Kaufmann, San Francisco, California, p. 247-258, 2000.

[BOU 92] BOUJEMAA N., STAMON G., LEMOINE J., PETIT E., "Fuzzy Ventricular Endocardium Detection with Gradual Focusing Decision", *IEEE EMBS Conference*, Paris, p. 1893-1894, 1992.

[BUS 00] BUSCHKA P., SAFFIOTTI A., WASIK Z., "Fuzzy Landmark-Based Localization for a Legged Robot", *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Takamatsu, Japan, p. 1205-1210, 2000.

[DES 96] DESCOMBES X., MOCTEZUMA M., MAÎTRE H., RUDANT J.-P., "Coastline Detection by a Markovian Segmentation in SAR Images", *Signal Processing*, vol. 55, no. 1, p. 123-132, November 1996.

[GER 95] GÉRAUD T., MANGIN J.-F., BLOCH I., MAÎTRE H., "Segmenting Internal Structures in 3D MR Images of the Brain by Markovian Relaxation on a Watershed Based Adjacency Graph", *ICIP-95*, Washington DC, p. 548-551, October 1995.

[GER 99] GÉRAUD T., BLOCH I., MAÎTRE H., "Atlas-guided Recognition of Cerebral Structures in MRI using Fusion of Fuzzy Structural Information", *CIMAF'99 Symposium on Artificial Intelligence*, Havana, Cuba, p. 99-106, 1999.

[GER 00] GÉRAUD T., BLOCH I., MAÎTRE H., "Reconnaissance de structures cérébrales à l'aide d'un atlas at par fusion d'informations structurelles floues", *RFIA 2000*, vol. I, Paris, p. 287-295, 2000.

[HEG 98] HÉGARAT-MASCLE S.L., BLOCH I., VIDAL-MADJAR D., "Introduction of Neighborhood Information in Evidenve Theory and Application to Data Fusion of Radar and Optical Images with Partial Cloud Cover", *Pattern Recognition*, vol. 31, no. 11, p. 1811-1823, 1998.

[MAI 02] H. MAÎTRE (ed.) *Le traitement des images*, Hermès, Traité IC2, Paris, 2002.

[MAN 95] MANGIN J.-F., REGIS J., BLOCH I., FROUIN V., SAMSON Y., LOPEZ-KRAHE J., "A Markovian Random Field based Random Graph Modelling the Human Cortical Topography", *CVRMed'95*, Nice, France, p. 177-183, April 1995.

[MAN 96] MANGIN J.-F., FROUIN V., RÉGIS J., BLOCH I., BELIN P., SAMSON Y., "Towards Better Management of Cortical Anatomy in Multi-Modal Multi-Individual Brain Studies", *Physica Medica*, vol. XII, p. 103-107, 1996.

[MOI 95] MOISSINAC H., MAÎTRE H., BLOCH I., "Markov Random Fields and Graphs for Uncertainty Management and Symbolic Data Fusion in a Urban Scene Interpretation", *SPIE/EUROPTO Conference on Image and Signal Processing for Remote Sensing*, vol. 2579, Paris, p. 298-309, September 1995.

[TUP 98] TUPIN F., MAÎTRE H., MANGIN J.-F., NICOLAS J.-M., PECHERSKY E., Linear Feature Detection on SAR Images: Application to the Road Network, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 2, p. 434-453, 1998.

[TUP 99] TUPIN F., BLOCH I., MAÎTRE H., "A First Step Towards Automatic Interpretation of SAR Images using Evidential Fusion of Several Structure Detectors", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, p. 1327-1343, 1999.

Chapter 10

Multi-Agent Methods: An Example of an Architecture and its Application for the Detection, Recognition and Identification of Targets

The purpose of vision systems is to allow an understanding of the observed scene, based on diverse data obtained from images. This scene is often characterized by a non-homogenous, highly variable environment and any observation condition. Obviously, this general nature leads to very complex systems, including the need for scalability, for the ability to provide intermediate results, the integration of uncertain knowledge, or the possibility to adapt through the choice of strategies, operators and parameters.

The problem of vision that we have just introduced involves the detection, recognition and identification (DRI) of targets such as military ground vehicles, boats and aircraft. In this operational context, we have to be capable of making a decision as quickly as possible, in order to assess the threat. A DRI system needs to constantly search for the relevant information among a wealth of useless elements of information. We propose a method based on multi-agent concepts to solve this problem. At each instant, a population of agents works in parallel in the environment. Each agent has descriptive and operational knowledge at its disposal to allow it to elaborate its own strategy and to conduct processes. The set of results obtained is stored in a world model shared by all the agents.

Chapter written by Fabienne EALET, Bertrand COLLIN and Catherine GARBAY.

This chapter is organized as follows: in section 10.1, we will describe the application and the state of the art in the field, before presenting the fundamental concepts of the system in section 10.2. In section 10.3, we will present the architecture and the platform used. In section 10.4, we will describe the agents of the system and the control scheme we propose. In section 10.5, we will describe the information handled by the agents. Finally, in the last section, we will illustrate this architecture with different results that were obtained.

10.1. The DRI function

The DRI function is comprised of three phases. Detection consists of finding the possible threats in the image. In this phase, little information is available: the target is represented by roughly a dozen pixels. An example is shown in Figure 10.1.

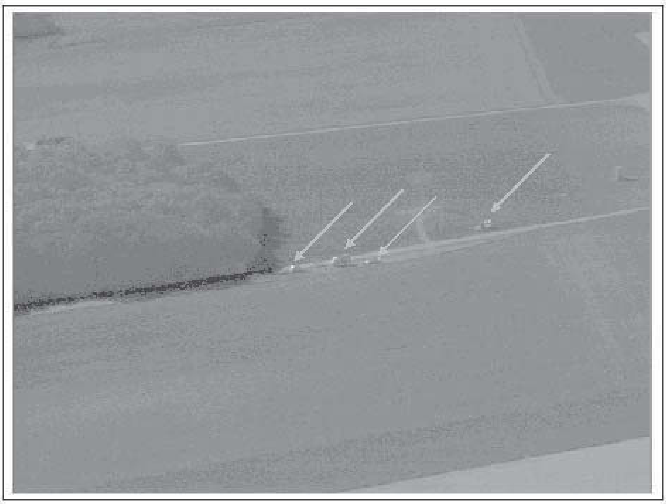


Figure 10.1. *Examples of vehicles in the detection phase. Each target is indicated by an arrow*

In the recognition phase, the mechanism is more precise since it specifies the target's class. The objective is then to answer the question: "is it a ground vehicle, a boat or a ship?" Finally, in the identification phase, the system has to be able to name the target: "is it an AMX10 tank, a frigate or a Mirage 2000 plane?" The consecutive phases of detection, recognition and identification require for the information regarding the target to be expanded in time.

10.1.1. *The application context*

The objective of this work is to find the information regarding the target in an image environment characterized by several levels of complexity:

– unpredictable context: the elements of the observed scene and the targets present are not known beforehand;

– variety of possible scenarios: a scenario corresponds to an observation phase of several targets in the considered environment. The scenarios are constantly evolving in terms of luminosity, possible masking, shadows;

– multi-target application context: when processing a sequence of images, a target can enter the sensor's observation field at any time;

– heterogenous observations: the considered targets are ground vehicles, aircraft or boats. Depending on their positions with respect to the sensor, their aspects and their time signatures are highly variable;

– incompleteness and imprecision: the information provided by the sensor is tainted with imperfections and gives a partial interpretation of the scene.

We have chosen to limit the context of this study to the tracking of military targets in a non-structured (non-urban), outside environment. Depending on the position of the camera with respect to the target, we considered three different missions:

– "ground-to-ground" missions: the camera carrier is placed on the ground and observes a scene on the ground; the possible target maneuvers in the observed space; the sensor can be fixed or mobile; masking phenomena, in this configuration, are very common and often caused by the presence of vegetation;

– "air-to-ground" missions: the sensor is embedded in an airborne carrier (plane, helicopter). It is mobile and observes the scene on the ground; blocking is caused mainly by vegetation-type masking structures;

– "air-to-air mission": as before, the sensor is embedded in an airborne carrier, it is mobile and observes part of the sky; target masking is caused mainly by the presence of clouds, or by the possible presence of other targets.

Depending on the type of mission, the objectives, the target classes and their behavioral characteristics vary. In a surface-to-air mission, the targets are planes. In an air-to-ground or ground-to-ground mission, the focus is more on boats and ground vehicles. This specification provides an element of information regarding the target's possible class.

In the context of this study, we worked with images provided by infrared and visible sensors, acquired during surface-to-air or ground-to-ground missions (fixed sensor). However, the system's design allows the integration of new sources of knowledge and new algorithms adapted to other types of sensors such as the radar or other types of mission.

10.1.2. *Design constraints and concepts*

The vision problem poses a number of difficulties:

– we do not know beforehand which processing chain can solve the vision problem we are dealing with. It is chosen according to the context and the situations encountered. There is never an overall objective defined, on the contrary, local objectives come up depending on the processes being conducted. Furthermore, the aspect variability for a given object or several objects in an image suggests that it is impossible to apply a comprehensive process to the entire image. Thus, the system, depending on the context, needs to adapt its strategy, the operators as well as the parameters according to the local information that is gathered;

– the specification of entities, objects, classes and sub-classes needs to allow a high variability within classes and from one class to another. Of course, this implies that the systems need to be capable of handling uncertainty;

– the information we are interested in is sometimes concentrated in a few pixels, for example, in detection and it is drowned in a considerable amount of data, hence the need to follow an incremental, safe and progressive procedure in order to gather knowledge. The objective is to rely on results that have already been obtained to continue with the process;

– the context of a multi-target application imposes the need for a monitoring alert function that can very quickly inform the system of the presence of a possible threat.

All of these difficulties described here raise a number of questions regarding design that have been discussed in other works, such as the ability to adapt, to focus, to distribute, to handle uncertainty and the system's incrementality.

10.1.3. *State of the art*

Over the past two decades, there has been considerable progress on the subject of DRI from a scientific and technical perspective. Bhanu and Ratches have drawn a particularly detailed review of this progress [BHA 86, RAT 97]. For the most part, the systems suggested in the early 1980s were heuristic. They would typically use the following sequential scheme: segmentation, detection, extraction of parameters, classification and tracking targets.

At the end of the 1980s, a new generation of systems appeared, with the intention of breaking this sequentiality. These methods explicitly integrated the knowledge and techniques of shape recognition (knowledge-based vision and model-based vision). They usually go through the following phases: searching for relevant areas, then recognition and identification of targets in a more limited area [BHA 92].

Today, improving performances requires associating complementary information in order to provide an adequate response to the operational needs of situation analysis.

Among other things, they make it possible to improve the robustness of the analysis with respect to its environment, the acuteness of the information transmitted and the reactivity in time. However, a smart use of the synergies between the different sources of information and the volume of available data requires the implementation of data fusion methods. Different theoretical methods would seem to satisfy this condition since they can be used, for example, to deal with uncertainty, imprecision, incompleteness, reliability, dependence or relevance. Among these methods, we should mention mainly probabilistic [BUE 97], possibilistic, fuzzy set and evidence theories, as well as logical or connectionist methods [HEN 93, ROT 90, YOU 98].

The integration of multi-sensor data, of spatial and temporal knowledge has opened new prospects for the future. The goal is no longer to develop new algorithms but rather to take advantage of partial and imprecise information provided by each one of them. The most common approach remains the prediction and verification of hypotheses [BEV 97, WU 97]. The major drawback is often the handling of a great number of hypotheses, which imposes a large number of combinations to examine. This method of exploration, associated with feedback, has the advantage of questioning hypotheses, and of choosing algorithms and parameters [DRA 95]. In the event of poor performances, it authorizes a change of strategy rather than pursuing the analysis with faulty hypotheses. It allows the system to adapt itself to the current situation. The calculation of a cost/benefit function makes it possible to complete this adaptive scheme by asking the question of how useful an action is, knowing the current situation and uncertainties [ROB 94].

Constraints tied to the heterogenity and the amount of knowledge naturally lead to architectures that distribute knowledge. The objective from now on is to equip the DRI system with the ability to manage, select and activate its own resources so as to conceive efficient control strategies to go through with the various analysis phases. Today, a DRI system is at the intersection between image analysis and cognitive reasoning functions.

10.2. Proposed method: towards a vision system

The problem of vision is stated as an incremental problem of information gathering, in which at each step of the process we ask the questions “where, what, how” [GAR 00, RAO 95]:

– where is the relevant information located in the image?

– what is the relevant information?

– how is this information extracted from the image?

In order to answer these questions and satisfy the constraints stated before, a vision system must constantly rely on the information gathered to construct its own knowledge of the field. The objective is to suggest new areas of focus, different goals and strategies based on the available information. In order to abide by this philosophy inherent to multi-dimensional space, we have made a series of choices which we will justify as we go along.

10.2.1. *Representation space and situated agents*

We suggest focusing the design of our system on the concept of situated agents. Such an agent is embedded in a three-dimensional space: the image space (the "where"), the goal space (the "what") and the method space (the "how"). A representation of the system is shown in Figure 10.2.

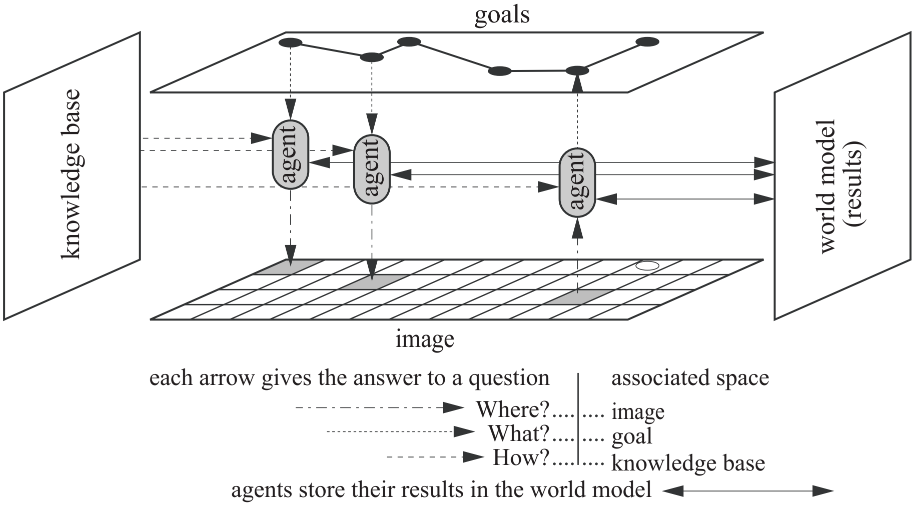


Figure 10.2. *General view of the multi-agent system. Each agent is embedded in a three-dimensional space comprised of the image space, the goal space and the method space. The agents are situated in the image, with a precise goal. They work locally using the data from the knowledge base to produce partial results that will be stored in the world model. This world model is then shared among all of the agents*

The goal space is comprised of a set of concepts that represent the elements in the scene we wish to analyze, such as roads, vegetation, the sky, the sea, etc. In the detection phase, an agent for a given concept (what?) will select a method (how?) in order to extract a region of interest (where?). The population of agents constantly uses the information specified in the knowledge base or gathered in the world model to work out its own strategy:

– an agent is always situated in an area of interest. For example, the system may be searching for vehicles on the roads. Extracted from a previous process, roads then constitute the areas of interest for vehicle tracking;

– agents always have an objective at their disposal. It is set by the knowledge base. Starting with its initial objective, the system navigates through the goal space by means of the focusing rules;

– depending on the concept to search for, agents have one or several methods at their disposal in the knowledge base. For example, when searching for areas that represent the road concept, agents can rely on geometric information (Hough transform, AREA) and radiometric information (gray levels).

Thus, the knowledge base allows connections between different representation spaces:

– a connection between the goal and the "how": the knowledge base associates each concept with localization, detection and focusing methods, which will then be adjusted;

– a connection between the goal and the "where": the knowledge base specifies focusing strategies which enable to define new areas of interest for a new objective.

This variation in the three representation spaces gives the system great qualities in adapting, distributing, incrementing as we will now discuss.

10.2.2. *Focusing and adapting*

According to the philosophy generally advocated by active perception and among others by Bajcsy [BAJ 88], we have chosen a strategy that operates by consecutive focusing steps, in other words a control guided by the search for information [TOU 98]. Thus, when an agent extracts relevant information, it can request the creation of another agent to obtain complementary information in a precise area of the image. As a result, the system works out a strategy based on what it finds at a given instant, and in that sense can be described as opportunistic. Depending on the goal and the context, a method is selected in the knowledge base and its parameters are adjusted locally by the agent.

Focusing and adaptation mechanisms are closely related. They make it possible to improve the performances of a vision system by way of an active behavior of the system. It is only possible to adapt the parameters and strategies of a system if a focusing phase has been conducted beforehand.

These two concepts are also embedded in the three-dimensional "where, what, how" space. As regards focusing, we will use the terms:

– spatial focusing: control in the image space;

– cognitive focusing: control in the goal space;

– operational focusing: control in the task space.

As regards adapting, we will use the terms:

– image adaptation: the system evolves through a series of focusing phases and organizes itself according to the nature of the information found in the image;

– cognitive adaptation: the goals are constructed in a dynamic fashion based on the gathered information;

– operational adaptation: the methods are adapted based on the goal and the location.

This description of the ideas behind focusing and adaptation shows the relation that binds them. There is in fact a permanent interaction between these two concepts. The focusing phase appears in some way as essential to adaptation, regardless of which space is considered. For example, in the image space, focusing defines a window of interest, and adaptation takes advantage of the information found in that window to adjust its processing parameters.

Focusing and adaptation within a system are achieved by way of a constant interaction between the system and its environment. Depending on the information available, the system develops and organizes itself in such a way as to expand its knowledge of the application field.

10.2.3. *Distribution and co-operation*

Distribution and co-operation are two concepts that are readily associated, since distribution in a set of subspaces implies the gathering of partial results. The objective of co-operation is precisely to co-ordinate the activities and to combine the results so as to complete the vision task. According to Hoc [GAR 00, HOC 96], it is possible to distinguish three forms of co-operation: confrontational co-operation, augmentative co-operation and integrative co-operation. It is worth noting the analogy with the three representation spaces of the agents, in order to underline the relation between these three forms of co-operation and the system's forms of distribution. The following analogy becomes apparent:

– between confrontational co-operation and "how": the problem is considered in its operational distribution. Several methods can be considered to reach a goal;

– between augmentative co-operation and "where": the problem is considered as spatially distributed. The system navigates in the image space. It works on areas of the image then fuses the results;

– between integrative co-operation and "what": the problem is considered as functionally distributed. The system navigates in the goal space. In order for the final objective to be achieved, intermediate tasks have to performed.

This concept of co-operation includes of course collective dynamics and a distribution in the design. In vision systems in general there is always at least one level of co-operation [GAT 96]. The fundamental point that characterizes all of these co-operations, whether they are confrontational, augmentative, or integrative is the accumulation of information with the objective of making a decision.

10.2.4. *Decision and uncertainty management*

Decisions have to be made at every step of the vision process. In order to make these choices, the system needs to constantly estimate the confidence in the current hypotheses, whether as regards to location, detection, focusing or recognition. Confidence depends on a great number of parameters such as:

– the quality of the selected operators;

– the presence of significant descriptors for objects;

– knowledge about the context, the environment.

It integrates very different elements of information, such as statistical, geometric or kinematic information, for example. In order to do this, different formalisms have been suggested, such as probabilistic methods [RIM 93], possibilistic methods [NIF 98], Dempster-Shafer theory [LEF 96, NIG 00], or also fuzzy methods [MEE 00]. These formalisms make it possible to approach the problem of the dynamics of beliefs by using inference mechanisms, in order to draw temporary conclusions [FAB 96]. As Bremond and Thonnat have aptly pointed out, "the most commonly used methods are probabilistic .... In the context of scene interpretation ... they provide a strict and rigorous framework that can be applied to any type of uncertainty" [BRE 96].

Graphs constitute a graphic and efficient way of representing the relations that exist between the different possible states of the system: causality, dependence or temporal relations, etc. They also turn out to be well-suited tools for structuring knowledge, thus making it possible among other things to propagate information and uncertainty in order to reinforce or decrease hypotheses.

This is why we have chosen a representation in the form of Bayesian networks. This formalism requires a complete definition of the probabilistic model, but provides a rigorous framework [FAB 96].

10.2.5. *Incrementality and learning*

Incrementality presents itself on several levels of the system:

– in the construction of the system itself since the population of agents is adapted depending on the information available and its relevance;

– in the gathering of information since the information is progressively accumulated in a world model which is made available to the population of agents;

– in the confidence in the information since the system reinforces hypotheses or not through time and weights the uncertainty depending on the information obtained;

– finally, in the knowledge base, which gives an operational description of the scene elements, since the system can refine over time its knowledge that was set beforehand. The term learning is used in this sense.

This incrementality is essential in the context of our application, where the relevant information can be summed up to a few pixels in the detection phase. Therefore, the objective is both to progressively take advantage of the environment and to authorize the system to be adaptive.

We have just described the fundamental points of the design we are proposing. They are associated with a set of keywords: focusing, adaptation, distribution, co-operation, uncertainty, decision, incrementality and learning. These concepts are the basis of the multi-agent architecture we will now describe.

10.3. The multi-agent system: platform and architecture

We propose a decentralized architecture to solve the problem of the detection, recognition and identification of targets. We will focus our attention on the detection phase. We will first present the system's architecture, i.e. the agent in its environment. We will then discuss further the multi-agent platform that allows the exchange of messages.

10.3.1. *The developed multi-agent architecture*

The agents behave independently. They work in an area of the image with their own objective, in competition or in parallel. The knowledge necessary to achieve an objective is specified in the knowledge base shared by all of the agents. Each one of them gathers information and stores it in a world model, which is also shared. This architecture is shown in Figure 10.2.

10.3.2. *Presentation of the platform used*

A community of agents cannot exist without a set of generally complex models defined within a platform. This platform specifies the fundamental elements such as the hardware and software architecture, or also the communication modes used by the agents to communicate with each other and with the data. The different features provided by this structure are summed up in Table 10.1, according to the plan suggested by the ASA (Agent System Architecture) workgroup, a group supported by the AFIA[1] and the GDR-I3[2].

1. French Association for Artificial Intelligence.

2. The Research Group for Information, Interaction and Intelligence of the CNRS.

Platform name	None.
Date and context of its creation	Developed in 1998 by Bouillot and Perez (DGA) as a tool for different research studies.
Development stage	Version 1.0
Current state	Stable version.
Hardware configuration	Operates on a set of UNIX stations (including a 32 processor system).
Software environment	– C++ programming. – Thread library. – TCL/TK interface.
Other works	Internship report [BOU 99]. There are no specific works currently available.
Applications	– Application to mobile robotics [DAL 01]. – Implementation of an image processing chain [ROP 01]. – Application to the detection, recognition and identification of targets [EAL 99, EAL 00].
General architecture	Each agent is an independent thread. These agents communicate by sending messages according to an asynchronous protocol. Any agent can request the creation of another agent or decide to leave the system. The respective activation or destruction messages are sent to the administrator.
Agent model	The generic agent model includes all of the communication features but no particular data regarding the agent's abilities, the representation of outside tasks or the representation of the abilities of other agents of the system. Aside from the communication functions, the agent is therefore an empty structure.
Design tools	There are no design tools, however, there is an interface which makes it possible to preserve a record of the system's operation.
Test	No test.
Availability	Not available.

Table 10.1. *Description of the multi-agent platform*

The platform is comprised of two elements: an agent structure equipped with means of communication and an independent supervisor known as an administrator. This administrator serves as the connection between the various actors of the systems.

In particular, it receives the requests sent by the agents for the creation of new agents. Its role is to:

– take an inventory of the agent population;

– assign to the agents the resources necessary for them to function (allocation of memory space);

– create a thread for each agent;

– provide a common time frame of reference.

The administrator has an information file for each agent that specifies its technical features.

10.4. The control scheme

The objective is to distribute the vision processes among a population of agents. The agents in the system have different roles and different behaviors with respect to the information that is handled and produced. Because we are dealing with the analysis of a sequence of images, we will distinguish two levels of control: intra-image control and inter-image control.

10.4.1. *The intra-image control cycle*

The diagram of the principle of intra-image control is shown in Figure 10.3. It is comprised of two exploration loops: spatial exploration and temporal exploration.

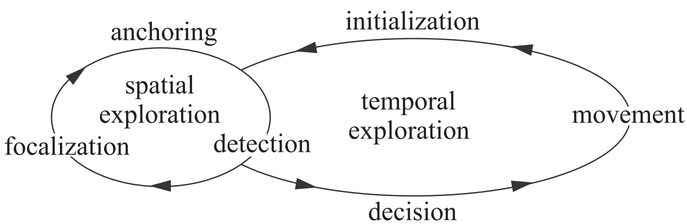


Figure 10.3. *The intra-image control diagram is organized according to a spatial exploration strategy and completed by a temporal analysis of the image*

Spatial exploration of the image takes place in three phases:

– the first phase consists of searching areas in the image for samples, i.e. possible attachment points for a given concept (road, field, vehicle, etc.). In concrete terms, the goal is to search for the areas that comply with a set of clauses specified in the knowledge base (a detailed description of this structure is given in section 10.5.1). This is the localization phase;

– the second phase relies on the samples to extract the supports of the concepts by using techniques based on the growth of areas. This is the detection phase;

– finally, the last phase uses available information to determine new areas of interest in the image and to progress in the concept space. This is the focusing step.

These three phases are repeated until all of the focusing rules specified in the knowledge base have been explored.

We have made the choice of distinguishing the localization and detection phases for two major reasons. On the one hand, it is often easier to find in the image an element that is characteristic of an object rather than that object's entire support. On the other hand, when the objects in the image have been localized, it becomes possible to locally adapt the operators and parameters in order to improve the performances of the segmentation.

Each phase of the process is conducted locally by an agent. The intra-image control cycle is presented in Figure 10.4.

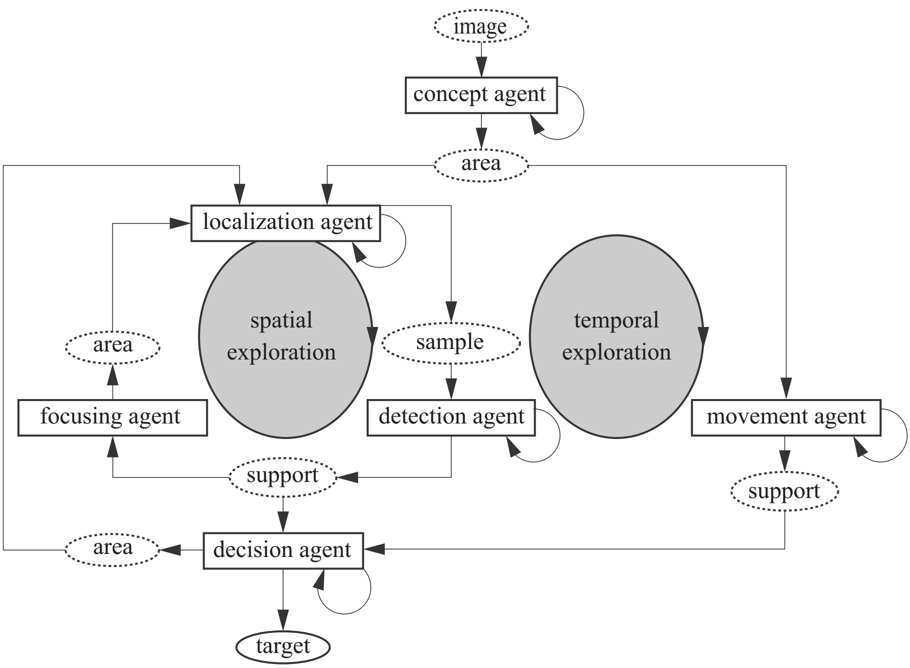


Figure 10.4. *The intra-image control diagram is comprised of two analysis cycles: a spatial exploration cycle and a temporal exploration cycle in order to associate additional information in the context of information fusion*

The temporal exploration makes it possible to associate additional information in the context of an information fusion cycle. For example, in an infrared image, a vehicle can be characterized by a hot area in motion.

Just as with the spatial exploration, each phase of the information fusion is conducted locally by a type of agent. Therefore, two new agents have been specified: the Decision and Movement agents. The Movement agent searches for the support of moving objects while the Decision agent's goal is to make the most of the two supports extracted by the Detection and Movement agents, respectively.

The Decision agent uses the information obtained to assess the need to continue with the analysis or to stop. This makes it possible to either activate the recognition process if a target has been detected, or otherwise to continue the detection phase by activating the loop related to the spatial exploration. As a result, the hypotheses expressed regarding the support's identity and kinematics, respectively, lead to different actions on the part of the Decision agent (see Figure 10.5).

results of the movement agent / results of the detection agent	target	non-target
movement	towards recognition (increased probability)	continue spatial exploration
non-movement	towards recognition (decreased probability)	

Figure 10.5. *The different behaviors of the Decision agent: the hypotheses stated according to the support's identity and kinematics provided by the Detection and Movement agents, respectively, lead to different actions*

At each new image, the system's initialization and the knowledge base's updating protocol are under the control of the Concept agent. This update allows the system to adapt itself according to the gathered information and to progress toward the knowledge specified beforehand. Thus, the parameters of the methods, among other things, can be modified, such as, for example, the validation range related to the image processing operators.

10.4.2. *Inter-image control cycle*

There are three problems to consider in the inter-image analysis cycle:

– using the image sequence;

– using the information gathered from the images;

– updating the knowledge base.

The usage of a sequence of images is carried out by a loop that operates in a set time interval. For each time increment, the intra-image process is started over again by way of the Concept agent.

The objective is then to use the past to our benefit, and by past we mean the information gathered from each image, in order to avoid going through all of the processes for each new image. Thus, under the fixed camera hypothesis, we have to consider the following two situations depending on the kinematics associated with the concept:

– for a fixed concept, it seems wise to keep the satisfactory samples obtained for a given image. These samples will be copied at the time $t + 1$ in the following image. This information copying phase is controlled by a validation threshold that specifies the minimum authorized confidence. Based on these copied samples, a new statistical analysis is conducted by the Detection agents in order to locally adjust the segmentation parameters;

– for a concept in movement, it seems wise to start over a complete analysis. Thus, nothing is copied.

After describing the intra- and inter-image control cycles, which made it possible to present the various agents of the systems and their roles, we are now going to describe in greater detail the information handled by the agents. We will focus on their features, on how they are organized, as well as on possible causality relations.

10.5. The information handled by the agents

Each agent uses and provides information. All of these elements of information have to be organized so as to allow other agents to have access to them. We will distinguish two types of elements of information in our system:

– descriptive and operational knowledge that provides a prior representation of the different concepts present in the application;

– the information gathered by agents during the analysis of the image sequence.

This distinction is important since the representation modes associated with these elements of information are distinct and it is precisely the composition of these information structures that we are going to describe with the knowledge base and the world model.

10.5.1. *The knowledge base*

The knowledge base is specified in order to provide agents with the information required for their execution. It defines, among other things, the field of the process, the initial objective and provides an operational description of the concepts.

The knowledge base is organized in four parts (an example is shown in Figure 10.6).

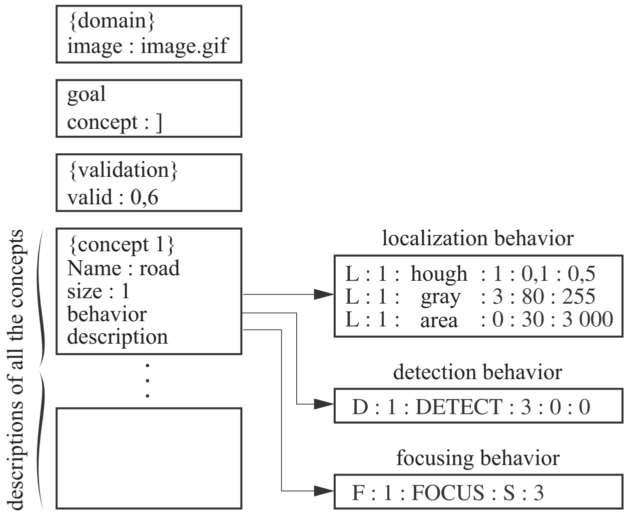


Figure 10.6. *The knowledge base*

– the first specifies the field of our work, i.e. the set of images used;

– the second defines the initial objective given to the system;

– the third part provides the validation threshold as regards the extracted information; this validation threshold will be used in the inter-image control cycles, among other things, to choose the samples to copy;

– the last part provides a description of the methods necessary to the agents and their parameters. The different methods consist of localization, detection, focusing and movement, which are characterized by the tags “L”, “D”, “F” and “M”, respectively. Localization methods rely on local criteria for intensity (“GREY”), homogenity (“HOMOG”), area (“AREA”) and criteria related to the Hough transform (“HOUGH”). Thus, a sample is defined as a set of points that verifies a set of criteria associated with the current concept. Detection methods are based on area growth (“DETEC”). Finally, focusing methods (“FOCUS”) search for new areas of interest, based on the current information, using three different strategies: on (“O”), next to (“N”) or elsewhere (“E”) in the image. A movement detection method has been implemented (“MOVE”) and, based on the optical flow constraint, it is expressed as an optimization problem [KOR 97].

The elements of information specified in the knowledge base depend on the application field and more specifically on the context of the mission. As a result, they are difficult to specify beforehand. To overcome this difficulty, we have chosen an interactive acquisition of knowledge using a graphical interface (developed in TCL/TK) which allows us to specify the names of the concepts we want, the types of methods necessary to their analysis and to calculate in semi-automatic mode the parameters of these methods. In order to do this, the areas of interest that represent each concept (road, vehicle, etc.) are selected in an interactive fashion in the image.

At each new image, this knowledge base is updated based on the information gathered by the agents. This update currently concerns the method parameters. A prospect for the future is of course to extend this "learning" to the choice of methods.

10.5.2. *The world model*

The information gathered by the agents over time progressively adds detail to the world model. A representation of this structure is shown in Figure 10.7.

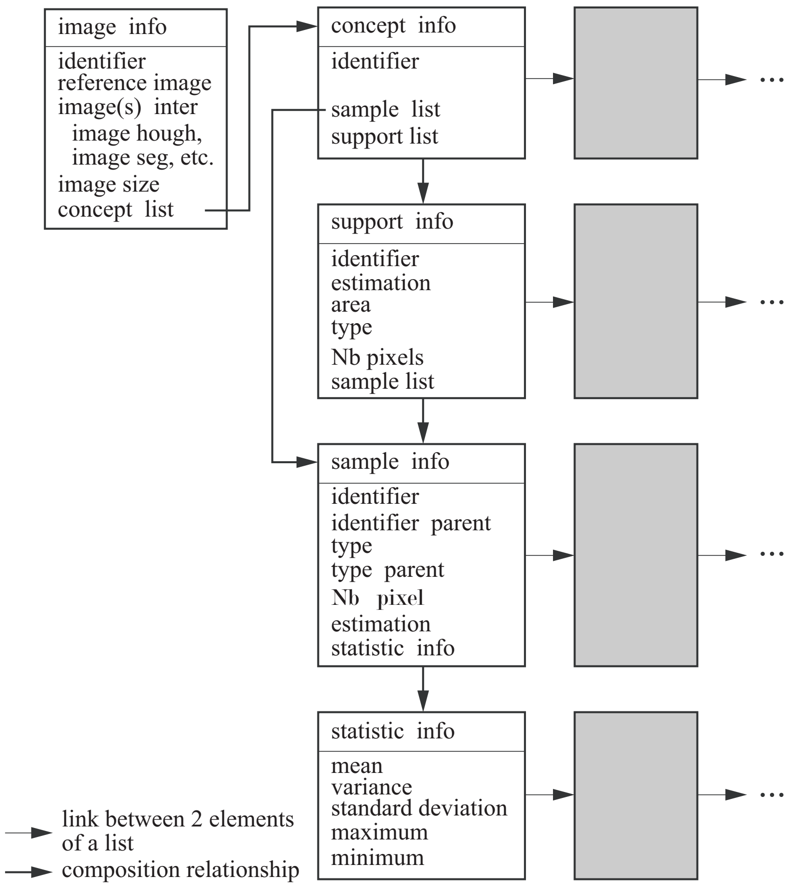


Figure 10.7. *Structure of the world model: for each image, we have access to the list of concepts, of supports and of the associated statistical parameters. A concept is comprised of several supports, which are themselves comprised of samples characterized by a list of statistical information*

This description of the scene is constructed incrementally. The update of the world model is conducted throughout the analysis process. The information gathered is characterized by a confidence measure that makes it possible to evaluate the membership coefficient of an entity to a given concept. This parameter evolves over time and propagates through the information structure using Bayesian networks. The networks were constructed with the “Bayes Net Toolbox[3]”.

For a given support, its confidence depends on three sources of information:

– the samples that have contributed to constructing it;

– the additional information, if there is any, such as the presence of movement;

– the other supports that have induced its construction by focusing (the concept of lineage).

The calculation of confidence associated with each support can be shown in the form of a causality graph. It depends on several sub-networks (Figure 10.8).

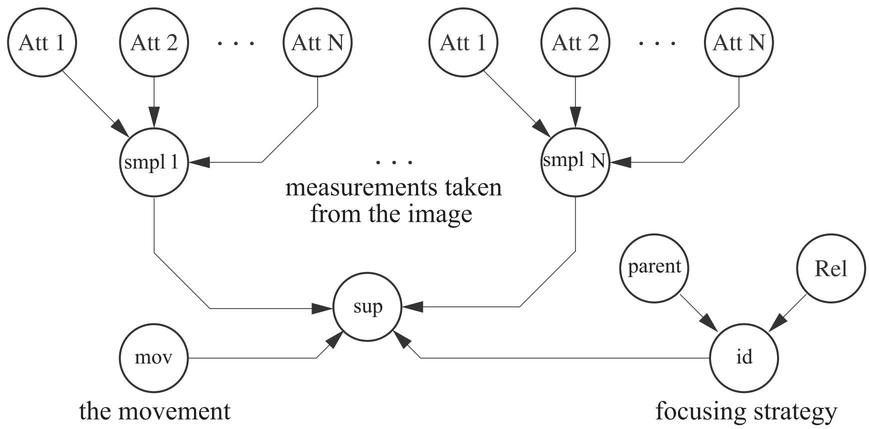


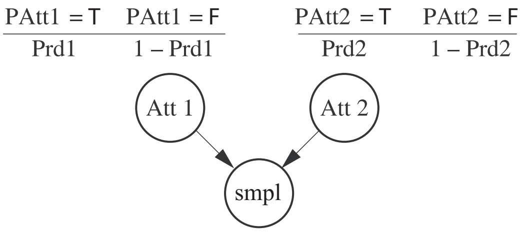
Figure 10.8. *Description of the Bayesian network used for calculating the confidence of a support. This confidence depends on the confidence obtained for each sample, on focusing relations and on the movement*

We are going to describe each one of these sub-networks in detail:

– for each sample (Spl1, . . . SplN) that belongs to the support, a statistical analysis is conducted in order to calculate the measure of confidence for each attribute (Att1, . . . AttN). The system calculates this measure of confidence as a gap between two probability distributions originating from the measure and the model. These measures

3. www.cs.berkeley.edu/%7Emurphyk/Bayes/bnt.html.

are then propagated using the Bayesian network. This is described by the diagram in Figure 10.9.



Att1	Att2	smpl = T	smpl = F
T	T	1	0
T	F	1/2	1/2
F	T	1/2	1/2
F	F	0	1

Figure 10.9. *The computation of the confidence in a sample is conducted using a Bayesian network. In this example, two methods are used to characterize a sample*

The measure of confidence associated with the sample is given by the following relation:

$$\begin{aligned} P(Spl{=}T) = {} & P(Spl = T/Att1 = T, Att2 = T) \cdot P(Att1 = T) \cdot P(Att2 = T) \\ & + P(Spl = T/Att1 = T, Att2 = F) \cdot P(Att1 = T) \cdot P(Att2 = F) \\ & + P(Spl = T/Att1 = F, Att2 = T) \cdot P(Att1 = F) \cdot P(Att2 = T) \\ & + P(Spl = T/Att1 = F, Att2 = F) \cdot P(Att1 = F) \cdot P(Att2 = F) \end{aligned}$$

– additionally, each support has a lineage resulting from the focusing relations. For example, a hypothesis of the type “vehicle on a road” will be reinforced, whereas the probability of having “road in the sky” will be minimized. The possibilities associated to each of the hypotheses are set beforehand and later evolve through the Bayesian network;

– the movement information can also reinforce a hypothesis. For example, the movement information will reinforce a vehicle hypothesis and not a road hypothesis. The concept used for calculating the confidence measure is similar to that described for the samples.

Depending on the information calculated by the agents, the probabilities will evolve and be updated. Information is propagated through the network based on the “causes-consequences” dependency relations that are set.

10.6. The results

Given an initial goal, the system’s objective is to pursue the exploration of the image or sequence as long as it is useful. This exploration is guided by focusing rules and a satisfaction threshold. A goal concept can also be reached by a "direct" strategy if it is the first specified concept, or by an "indirect" strategy if it is specified within the focusing rules. Depending on what strategies are used, the system's responses will be different. We will now illustrate these comments with various experiments:

– direct analysis: the system directly searches for vehicles in the image, without relying on context elements;

– indirect analysis: the system relies on context elements such as roads to search for vehicles. Two focusing strategies are used in the examples mentioned: search for vehicles "on" and "next to" roads.

10.6.1. *Direct analysis*

The first experiment consists of finding vehicles in an image using nothing but data from the knowledge base attached to the vehicle concept. The analyzed image and the associated knowledge base are shown in Figure 10.10.

```
{FIELD}
ROOT:IMAGES/SOURCE/Vol/vol.%02d.gif
EXTENSION:gif:BEGINNING:0:
NB:1:INCREMENT:1:
-------------------------- {GOAL}
CONCEPT:1:------------------------
{VALIDATION} VAL:0.10
------------------------ {CONCEPT
1} NAME:VEHICLE: SIZE:1:
L:1:GREY:1:114:255:
L:1:HOMOG:3:1:10:
L:1:AREA:0:10:100:
D:1:DETECT:1:0:0
{CONCEPT 2} NAME:ROAD:SIZE:1:
L:1HOUGH:1:0.1:0.5:
L:1GREY:4:80:255:
L:1AREA:1:250:4000:
D:1DETECT:2:0:0
```

Figure 10.10. *Direct analysis: the reference image is presented on the left and the associated knowledge base is described on the right*

In this image, two concepts are defined: road and vehicle. These concepts are characterized by localization and detection methods. On the other hand, no focusing rule is specified. Therefore, only the attributes related to the initial concept, namely "vehicle", are used. The models used in the localization phase to find the vehicles in the image are statistical (gray levels and homogenity) and geometric (AREA). The population of agents implemented in this example is shown in Figure 10.11.

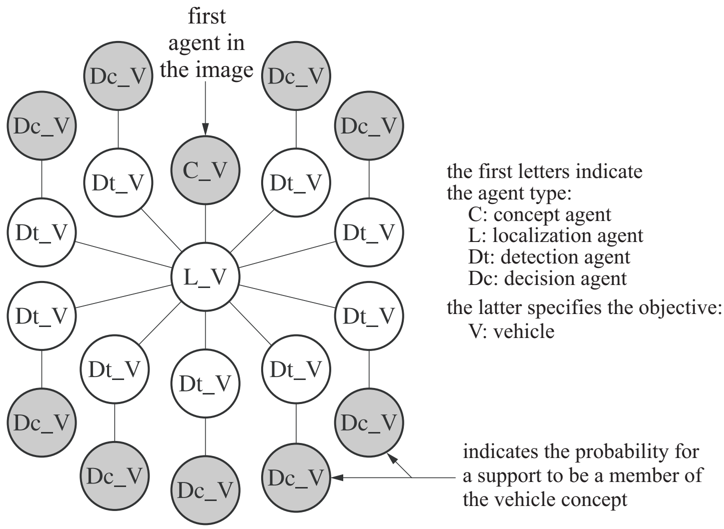


Figure 10.11. *Agent graph generated for detecting vehicles using the direct strategy; four types of agents are implemented: the Concept, Localization, Detection and Decision agents. All these agents have the same objective, namely the vehicle concept. In this example, nine vehicle-type supports are detected*

This population is organized around four types of agents:

– the Concept agent is the first agent that is created. Its role is to initialize the system and activate the exploration loops according to the information at its disposal in the knowledge base. In this example, it requests the activation of spatial exploration using a localization agent, since localization criteria are specified. On the other hand, no movement information is described in the base. As a result, the loop related to temporal exploration is not activated;

– a Localization agent searches the image for pixels that satisfy the set of localization criteria specified in the knowledge base and associated with the vehicle objective. After a labeling phase, this agent provides elementary areas referred to as samples, which are associated with statistical information. As soon as a sample is identified, the Localization agent requests the creation of a new Detection agent. The nine samples that were found led to the creation of nine Detection agents;

– each Detection agent works with the associated sample and performs a local statistical analysis to extract the object's support. No focusing strategy is specified and therefore the spatial exploration comes to an end;

– the Decision agent has no complementary information regarding movement since no knowledge on that subject is specified. Therefore, it transmits the data obtained without requesting complementary analysis and provides a decision.

All of these agents are not present in the image at the same time. As soon as an agent has completed its task, it asks the administrator to delete its information file before terminating itself. The result of the segmentation is shown in Figure 10.12.

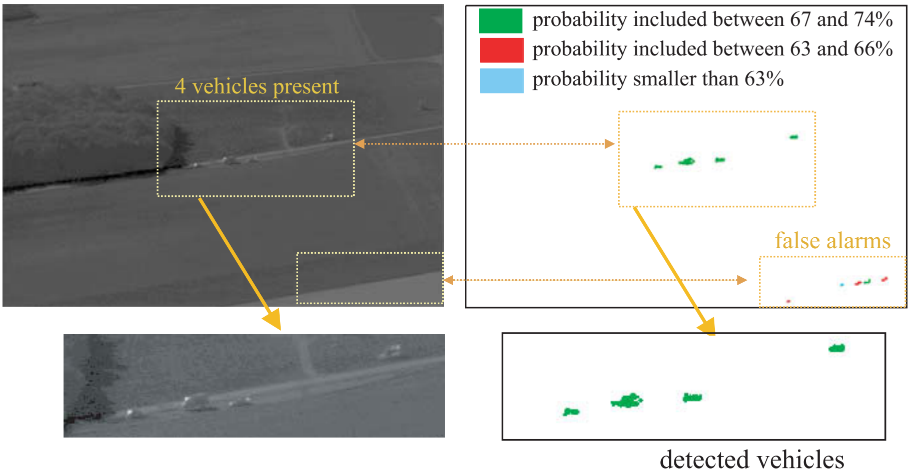


Figure 10.12. *Segmentation of the vehicles supports. On the left, the initial image shows the presence of four vehicles. On the right, the resulting segmentation is shown. The shade associated with each support represents the probability of membership to the concept*

This segmentation image displays the nine "vehicle" supports found in the image. A probability of membership to the vehicle concept is associated with each support. This probability is represented by a color. It is interesting to emphasize two important results:

– the vehicles are detected with a probability of membership of over 67%. On the other hand, these elements of information are drowned in a large amount of false positives with probabilities between 63 and 66%;

– this population of agents shows us a decentralization that appears very early on in the system. The control is conducted locally by each agent. Furthermore, the steps of the process are distributed so as to divide the computational load between the different processes, as shown in Figure 10.13.

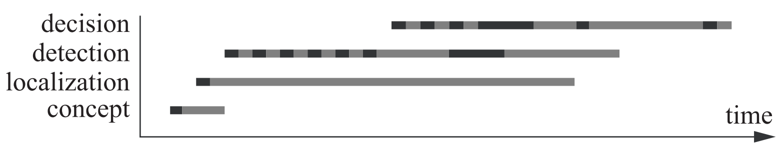


Figure 10.13. *Distribution of the computation times among the different agents: black indicates the activation of an agent and gray indicates that it is running*

Notice how costly the localization phase is in terms of computation time, since it conducts a local analysis on the entire image. This gives us an understanding of how useful the focusing phase is, since it will quickly provide the system with more limited areas of analysis, thus reducing the computation time. This is because localization is only conducted once for the entire image, even in the case of a sequence of images, and only for the first concept we are searching for. The result of the segmentation is mediocre. While vehicles are detected despite how little information is available, the rate of false positives is significant (44%). The next logical step therefore is to better control the analysis by using one or several focusing strategies.

10.6.2. *Indirect analysis: two focusing strategies*

In order to limit the domain we have to search for in order to find relevant information, we have introduced models that express the contextual relations between the various concepts found in the scene. These relations, attached to a concept, are expressed in the form of focusing strategies. They are specified in the knowledge base. As we have said before, we have defined three focusing strategies. It is possible to focus: "on" a support, "next to" a support, or also "elsewhere" in the image. In this section, we present the experimental results obtained with two focusing strategies.

The first idea is to search for large structures in the image where the system is likely to find vehicles. Therefore, in the previous example (section 10.6.1), we chose roads as our initial objective, knowing that the associated focusing strategies will consist of searching for vehicles on and next to roads. Although the knowledge base remains virtually unchanged compared to the previous example, it does, however, include two additional focusing rules that specify whether to search for vehicles (concept 1) on ("O") or next to ("N") roads (concept 2). The file associated with this knowledge base is described in Figure 10.14.

```
{DOMAIN} ROOT:IMAGES/SOURCE/Vol/vol.%02d.gif
EXTENSION:gif: BEGINNING:0: NB:1:
INCREMENT:1: --------------------------
{GOAL} CONCEPT:2: --------------------------
{VALIDATION} VAL:0.10
--------------------------
{CONCEPT 1} NAME:VEHICLE: SIZE:1:
L:1:GREY:1:114:255: L:1:HOMOG:3:1:10:
L:1:AREA:0:10:100: D:1:DETECT:1:0:0
{CONCEPT 2} NAME:ROAD: SIZE:1:
L:1:HOUGH:1:0.1:0.5: L:1:GREY:4:80:255:
L:1:AREA:1:250:4000: D:1:DETECT:2:0:0
F:1:FOCUS:S:3 F:1:FOCUS:A:3
```

Figure 10.14. *Knowledge base for an indirect analysis of the vehicle concept*

The exploration graph associated with this segmentation is shown in Figure 10.15.

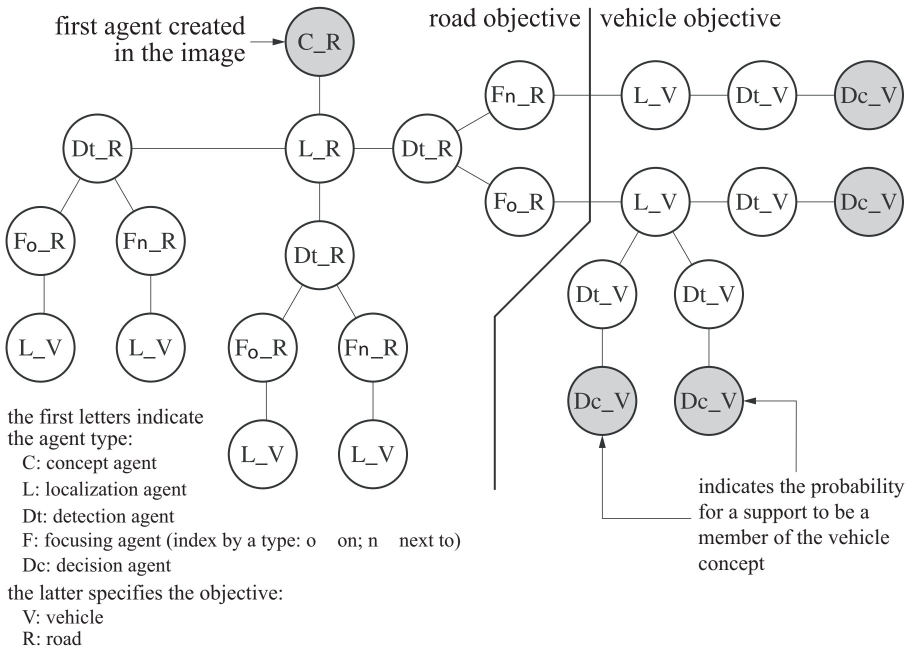


Figure 10.15. *Agent graph related to vehicle detection when two focusing rules are specified. The type of the initial objective is "road". The system searches for vehicles "on" roads or "next to" roads*

The first agent created is the Concept agent. It initializes the system and activates the spatial exploration by creating a localization agent, which searches for elementary areas of the image satisfying all of the localization clauses. Based on three extracted samples, three detection agents are given the task of extracting the road supports. We then note that the focusing branches related to the road objective split in two. This indicates that two focusing agents are activated when a road support is found. The system actually focuses "on" the road support, which is completed by also focusing "next to" road supports.

Five kinds of agents are required in this case: the Concept, Localization, Detection, Focusing and Decision agents. They can be designed with two possible objectives: road and vehicle. The segmentation result is shown in Figure 10.16.

Four vehicles are extracted. Three are located on one of the road supports and the latter is located next to that same support. The system detects all of the vehicles in

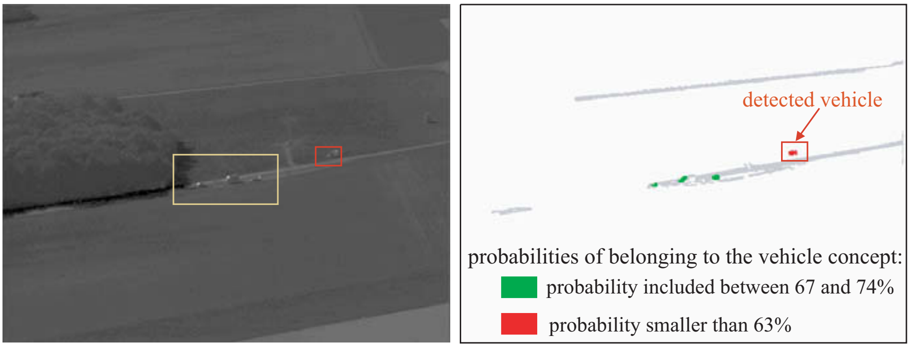


Figure 10.16. *Segmentation of vehicles with two focusing strategies. On the left, the initial image is shown, and on the right, the result of the segmentation. Four vehicles are detected: three on the road, and one next to the road*

the image. The vehicle located next to the road is characterized by a probability of roughly 60%. This probability is much lower than the other probabilities measured (in the range of 70%) that characterize the vehicles on the road. This result illustrates the impact of the focusing relations. The probability of observing a vehicle on a road remains higher than that of observing a vehicle next to the road.

In this example, we focused on the spatial exploration of the image. For the road objective, this exploration is represented by the series of localization, detection, then focusing behaviors, which allows a change of objective and an iteration of the process. All of the results that we have presented are relevant to intra-image processing. We will now consider a sequence of images and present, among other things, the concepts of incrementality and adaptability.

10.6.3. *Indirect analysis: spatial and temporal exploration*

In this section, we give an illustration of the different forms of comprehensive system adaptation. We have chosen to present the results for a sequence of 9 images shown in Figure 10.17. We have introduced a perturbation (a uniform increase in the gray levels) in the images numbered one and three in order to better evaluate the system's adaptive nature.

Using the graphical interface, we have specified the knowledge base presented in Figure 10.18. This is comprised of two concepts: vehicle and field. The vehicle concept is characterized by three localization methods, a detection method and a method related to motion detection. The field concept is characterized by three localization methods, a detection method and a focusing strategy that imposes a constraint on the search for vehicles in fields.

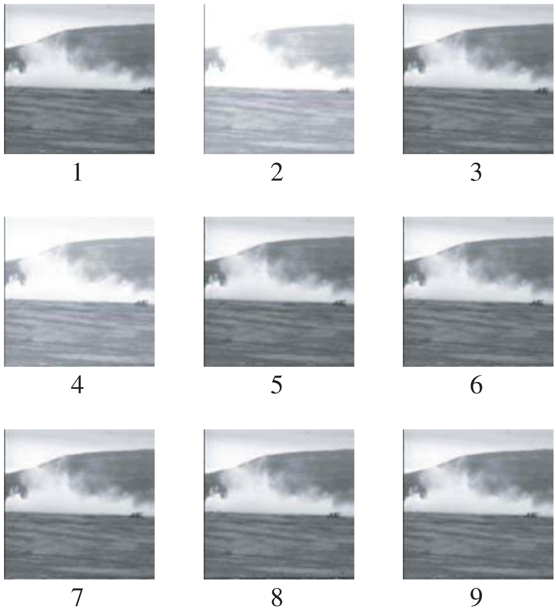


Figure 10.17. *Reference sequence*

```
{DOMAIN} ROOT:IMAGES/SOURCE/Mif1/mif.%02d.gif
EXTENSION:gif: BEGINNING:0: NB:9:
INCREMENT:1: -------------------------
{GOAL} CONCEPT:1: -------------------------
{VALIDATION} VAL:0.1 -------------------------
{CONCEPT1} NAME:FIELD: SIZE:1:
L:0:HOUGH:0:0.1:0.5: L:1:GREY:5:40:120:
L:1:HOMOG:9:0:3: L:1:AREA:0:250:3000000:
D:1DETECT:5:0:0 F:1:FOCUS:S:1
{CONCEPT3} NAME:VEHICLE: SIZE:2:
L:1:GREY:5:10:80: L:1:HOMOG:2:3:15:
L:1:AREA:0:150:1000: D:1:DETECT:1:0:0
M:1:MOVE:1:3:255
```

Figure 10.18. *Knowledge base related to the first image in the sequence*

The resulting agent graph is shown in Figure 10.19. The first agent created is still the Concept agent, but this time it simultaneously activates the spatial exploration by means of a Localization agent and the temporal analysis of the sequence by means of the activation of a Movement agent. The spatial exploration follows the diagram described before in the localization, detection and focusing loop. The temporal analysis is conducted in parallel, in order to quickly arrive at areas of interest. We notice that two Decision agents are activated while the image is being processed. The first one is activated on a request from the Movement agent and it checks to see if the moving areas have already been labeled as vehicles. If they have, this branch would have been interrupted and would have provided the probability of membership to the vehicle concept. In the case that concerns us, it turns out that the moving areas have not yet been labeled by the vehicle concept. Thus, a more in-depth analysis is needed, which results in the creation of five Localization agents. Only one Localization agent found a sample and therefore only one Detection agent is activated to extract the support of the object in question.

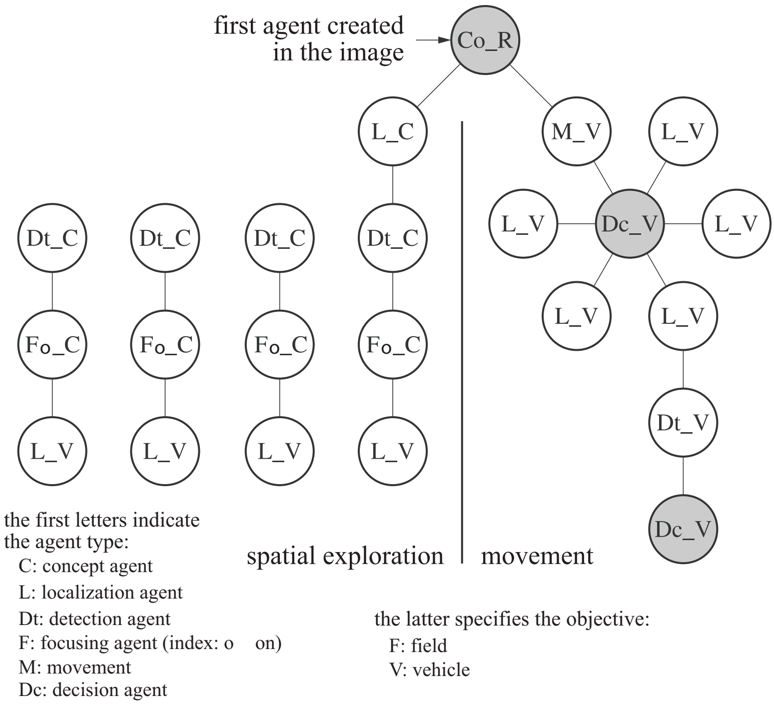


Figure 10.19. *Agent graph in the presence of movement. It represents the population of agents implemented to find the vehicles, knowing that the spatial exploration is conducted along with the search for movement information*

In this example, the object is detected by the loop related to the search for movement. This is because of the speed of execution of this loop compared to the spatial exploration. However, strictly speaking, the vehicle in movement may be found by either one of the two branches.

The result of the segmentation according to movement, obtained for the first image of the sequence, is given in Figure 10.20. Five regions have been labeled, including the vehicle, but also four other regions representing the concept "smoke". Each area appears as a potential target, based on which the analysis will continue.

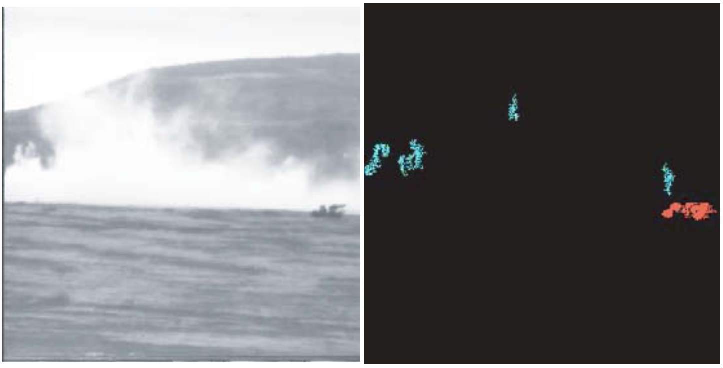


Figure 10.20. *Segmentation of the moving areas: on the left, the reference image is presented and on the right, the resulting segmentation*

Taking into account information regarding movement makes it possible to very quickly reach the area of interest, without waiting for the complete analysis related to the spatial exploration. Furthermore, it leads to an increase in the probability of membership to the vehicle concept. The probability obtained is 79%, whereas without the movement information, the probability is 65%.

The Movement agent is a monitoring agent that allows a quicker decision to be made regarding the possibility of a threat. On the other hand, this analysis in no way alters the benefit of the spatial exploration, which segments the targets that are not detected by the movement branch, either for reasons of computation time (depending, for example, on the machine's load, etc.), or because the vehicle in question is not moving.

Depending on the information available and the execution speed, the paths to the segmentation of the target can be modified. The system can operate using an exploratory process based on contextual relations or on strong information such as movement; this way, it adapts to the situations it encounters.

10.6.4. *Conclusion*

In order to solve the problem of the detection, recognition and identification of targets, we have proposed a multi-agent architecture based on the implementation of agents which are located simultaneously in the image space, the goal space and the method space[4]. The method we propose complies with a series of fundamental principles such as:

– focusing and adaptation;

– distribution and co-operation;

– uncertainty management;

– incrementality and learning.

Each concept allows the system to navigate within a space, but also to switch from one space to another. The suggested system designs its own strategy depending on the information available and proceeds in an incremental fashion. Over time, the system makes hypotheses that are confirmed or not, later on, depending on the information that is gathered. Thus, the suggested method follows an *anytime* philosophy, which means that there is always a possible decision and it is refined over time.

The system is equipped with various adapting capabilities. Here are the major ones:

– local adaptation, which is present, among other places, within the Detection agent, since it uses the statistics associated with the current sample to adjust the segmentation parameters;

– global adaptation which is present in two different forms:

- adaptation of the agent population: the presence of movement in the image can force the system to request a complementary analysis of the areas in question, thus new agents are created to answer to this request,

- adaptation of the data structures: over a sequence of images, the system is capable of managing the updates of the knowledge base and the evolution of the world model by using in particular a set of probabilities to assess the confidence in the results.

Exploration and analysis strategies are not set in stone. They depend on decisions made locally by agents and the modification of knowledge allows the systems to develop different strategies. This operating mode described for the detection phase applies in the same way to the recognition and identification phases [EAL 01]: in this work, the characterization phase was completed by a discrimination phase, since the adaptability relies on the concept of the usefulness of attributes.

10.7. Bibliography

[BAJ 88] BAJCSY R., "Active Perception", *Proceedings of the IEEE*, vol. 76, p. 996–1005, 1988.

4. This study was conducted as part of a CIFRE convention between the E/SCS/V of EADS, the TIMC/IMAG/SIC laboratory and the DGA/CTA/GIP laboratory.

[BEV 97] BEVERIDGE J., DRAPER B., STEVENS M., HANSON A., SIEJKO K., A Coregistration approach to multisensor target recognition with extensions to exploit digital elevation map data, Report no. CS-97-107, Colorado State University, September 1997.

[BHA 86] BHANU B., "Automatic Target Recognition: State of the Art Survey", *IEEE transactions on aerospace and electronic systems*, vol. 22, no. 4, p. 364-379, July 1986.

[BHA 92] BHANU B., JONES T., "Image understanding research for automatic target recognition.", *IUW*, p. 249–254, January 1992.

[BOU 99] BOUILLOT A., Développement d'une architecture multi-agent pour le contrôle d'un robot mobile, CTA 99 R 013, DGA/DCE/CTA/GIP, 1999.

[BRE 96] BRÉMOND F., THONNAT M., "Interprétation de séquences d'images et incertitude", *Rencontres sur la logique floue et ses applications (LFA)*, Nancy, December 1996.

[BUE 97] BUEDE D., GIRARDI P., "A target identification comparison of bayesian and dempster-shafer multisensor fusion", *IEEE transactions on systems man and cybernetics - part A: systems and humans*, vol. 27, no. 5, p. 569-577, September 1997.

[DAL 01] DALGALARRONDO A., Intégration de la fonction perception dans une architecture de contrôle de robot mobile autonome., PhD Thesis, Paris-Sud University, Orsay, January 2001.

[DRA 95] DRAKE K., KIM R., "Hierarchical integration of sensor data and contextual information for automatic target recognition", *Applied Intelligence*, vol. 5, p. 269-290, 1995.

[EAL 99] EALET F., TOUBOUL E., COLLIN B., SELLA G., GARBAY C., "Un système multiagent pour la détection, la reconnaissance et l'identification de cibles.", *ORASIS99*, Aussois, France, p. 289-298, April 1999.

[EAL 00] EALET F., COLLIN B., SELLA G., GARBAY C., "Multi-agent architecture for scene interpretation", *SPIE, Enhanced and synthetic vision 2000*, Orlando, Florida, April 2000.

[EAL 01] EALET F., Une architecture Multi-agent pour la détection, la reconnaissance et l'identification de cibles, PhD Thesis, University of Grenoble, June 2001.

[FAB 96] FABIANI P., Représentation dynamique de l'incertain et stratégie de perception pour un système autonome en environnement évolutif, PhD Thesis, Ecole national supérieure de l'aéronautique et de l'espace, 1996.

[GAR 00] GARBAY C., "Architecture logicielle et contrôle dans les systèmes de vision", in J.M. JOLION (ed.) *Les systèmes de vision*, Hermes, 2000.

[GAT 96] GATEPAILLE S., BRUNESSAUX S., "Dossier I.A. et fusion de données", *Bulletin de l'AFIA*, vol. 24, p. 21–30, 1996.

[HEN 93] HENOCQ H., BUREL G., "Reconnaissance d'objets 3D par analyse de Fourier de la silhouette", *Revue Technique Thomson-CSF*, vol. 25, no. 1, p. 107–125, March 1993.

[HOC 96] HOC J., *Supervision et contrôle de processus: la cognition en situation dynamique*, PUG, Grenoble, 1996.

[KOR 97] KORNPROBST P., DERICHE R., AUBERT G., Image sequence restoration: a PDE based coupled method for image restoration and motion segmentation, Report no. 3308, INRIA, November 1997.

[LEF 96] LEFÈVRE V., COLLET Y., PHILIPP S., BRUNESSAUX S., "Un système multi-agent pour la fusion de données en analyse d'images", *Traitement du signal*, vol. 13, no. 1, p. 99–111, 1996.

[MEE 00] MEES W., Contribution à l'analyse de scènes; Application aux images satellitaires multi-spectrales, haute résolution, PhD Thesis, Henri Poincaré University, Nancy 1, January 2000.

[NIF 98] NIFLE A., Modélisation comportementale en fusion de données; Application à l'identification d'objets ou de situations, PhD Thesis, Orsay University, 1998.

[NIG 00] NIGRO J., LORIETTE-ROUGEGREZ S., ROMBAUT M., JARKASS I., "Driving situation recognition in the CASSICE project towards an uncertain management", *ITSC*, 2000.

[RAO 95] RAO R., BALLARD D., "An active vision architecture based on iconic representations", *Artificial Intelligence Journal*, vol. 78, no. 1-2, p. 461–505, 1995.

[RAT 97] RATCHES J., WALTERS C., BUSER R.G., GUENTHER B., "Aided and automatic target recognition based upon sensory inputs from image forming systems", *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 4, p. 1004-1019, September 1997.

[RIM 93] RIMEY R.D., Control of selective perception using bayes nets and decision theory, Report, University of Rochester, Computer science department, Rochester, New York 14627, December 1993.

[ROB 94] ROBERTS B., BROWN C., "Adaptive and control in an ATR system", *IUW*, p. 467–479, 1994.

[ROP 01] ROPERT V., Proposition d'une architecture de contrôle pour un système de vision, PhD Thesis, Paris V University, 2001.

[ROT 90] ROTH M.W., "Survey of neural network technology for automatic target recognition", *IEEE Transactions on Neural Networks*, vol. 1, p. 28–43, 1990.

[TOU 98] TOUBOUL E., Modèle d'architecture pour la mise en œuvre d'un système de vision générique, PhD Thesis, Conservatoire national des arts et métiers, 1998.

[WU 97] WU X., BHANU B., "Gabor wavelet representation for 3D object recognition", *IEEE Transactions on Image Processing*, vol. 6, no. 1, p. 47-64, January 1997.

[YOU 98] YOUNG S.S., D.SCOTT P., BANDERA C., "Foveal automatic target recognition using a multiresolution neural network", *IEEE Transactions on Image Processing*, vol. 7, no. 8, p. 1122–1135, August 1998.

This page intentionally left blank

Chapter 11

Fusion of Non-Simultaneous Elements of Information: Temporal Fusion

11.1. Time variable observations

Temporal fusion of data allows us to handle instantaneous information, i.e. information whose value is only valid at a given time. This fusion is used when the parameters obtained from observation vary with time.

The problem therefore is to track the evolution of these observations, as opposed to an overall characterization of the evolution. The following example illustrates this instantaneous nature: when observing a beating heart, the temporal fusion of data makes it possible to estimate the instantaneous blood flow. With a non-temporal fusion, it would be possible to estimate the mean blood flow over a long period of time. On the other hand, temporal fusion would be needed to observe the evolution of the mean blood flow depending on, for example, the evolution of the patient's level of stress. This shows that we always have to define concepts related to time such as "moment", "evolution", "period of time" according to the application because a length of time can be "long" in certain cases, or "short" or even "negligible" in others.

Thus, temporal fusion involves the evolution or the modification of data:

– it may involve the modification of the object being observed. For example, when analyzing the heart's movement, focus will be directed to its morphological modifications over time;

Chapter written by Michèle ROMBAUT.

– it may involve the relative evolution of two objects. In robotic manipulation, for example, the objective is to model the effector's movement with respect to the object it has to grip in the scene, in order to be able to determine at any given time the control laws to apply;

– it may involve the relative movement between the observed object and the observing sensor. This is the case, for example, with driver assistance systems when the vehicle is equipped with a sensor capable of detecting and characterizing obstacles;

– finally, it may also be used when the acquisition conditions vary over time and when it is necessary to modify certain acquisition and processing parameters. For example, when the lighting conditions of an observed scene vary, it can be useful to modify a camera's aperture size.

11.2. Temporal constraints

Therefore, the objective of temporal fusion is to determine or evaluate information with a time-limited validity. The concept of an instant is completely related to the application. For example, the global warming timescale cannot be compared to the scale used for locating a plane. In any case, it is necessary to specify the expected performances depending on how the data is going to be used. During the observation of dynamic systems, the observation period (sampling period) has to be compatible with the Shannon theorem: *the sampling period must be no greater that half the response time of the observed system.*

This constraint requires the development of specific processing architectures if the observed system evolves quickly or requires rapid responses, compared to the time it takes to observe the data [ROM 95].

The processing time leads to a delay that corresponds to the time between *the acquisition of measurements* and the *transmission of data* obtained from processing these measurements. For example, if the sensor is a CCD camera, the measurement is an image comprised of pixels which are associated with gray levels. Processing the image can lead to obtaining data of a higher level such as areas associated with physical features. The data, obtained after the processing, involves the moment of observation, i.e. it corresponds to the date of the measurements it is associated with. Processing causes delays that may or may not be neglected depending on the application. For example, if the information is used to determine a control, the delay can cause an instability, which is a result well known to control engineers; if they are used to trigger an alarm, the delay can be fatal. A compromise will often have to be made between the quality of the data and how fast it has to be available.

For each sensor, we need to define the temporal quantities with which it is associated [ALL 01]. Figure 11.1 represents these quantities:

– the moment of acquisition is the measurement that corresponds to a date set by the clock. Usually, the error made for this data is ignored;

– the moment of transmission when the data is available to be fused, for example;

– the processing time that corresponds to the time interval between the acquisition moment and the transmission moment. This time may or may not be constant depending on which processes are applied;

– the sampling period, which corresponds to the time interval between two acquisitions. Usually, this period is constant.

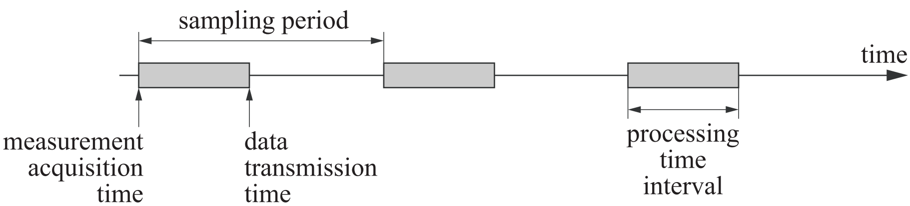


Figure 11.1. *Time related quantities associated with a sensor*

The essential time constraints are the acquisition frequency of measurements and the delay caused by the processing time. If the frequency is high enough and if the delay can be neglected, we can consider that all of the information that can be used in the fusion system is constantly available. It is worth noting that in most fusion applications, temporal constraints are neglected. They are dealt with mostly in military applications for detecting and tracking targets [ANK 01, APP 98, PAO 94], in mobile robotics [KOS 93, PRU 99] and in the context of intelligent vehicles [ETE 94, JOU 99, TRA 93, TRA 94].

11.3. Fusion

Fusion uses data provided by different sources with temporal characteristics that may be different. In the context of temporal fusion, information provided by the same source may also be fused, but at different times.

11.3.1. *Fusion of distinct sources*

In the ideal case, the sources of information are synchronous (same period and same acquisition time) and provide their information with the same processing time. The data available can then be immediately fused.

With real applications, the data is rarely synchronous and cannot be fused directly. Here are the major causes:

1) the measurements are sampled at the same frequency, but with a shift (Figure 11.2);

2) the measurements are sampled at different frequencies (Figure 11.3);

3) the measurements are sampled at the same frequency, but the processing times are different (Figure 11.4);

4) the measurements are sampled at different frequencies and the processing times are different. Therefore, the sensors are completely asynchronous (Figure 11.5).

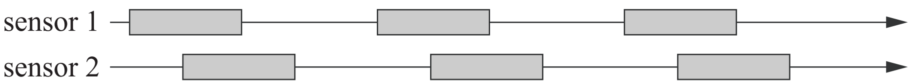


Figure 11.2. *Same period, same processing time, different acquisition times*

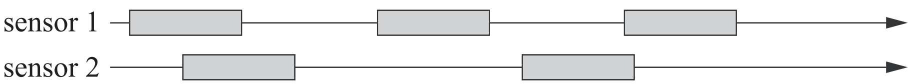


Figure 11.3. *Different periods, same processing time*

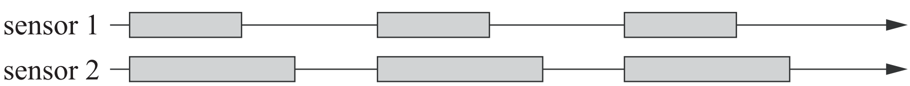


Figure 11.4. *Same period, same acquisition time, different processing times*

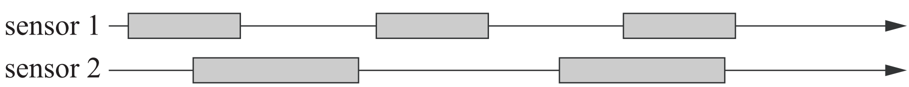


Figure 11.5. *Different periods, different processing times*

11.3.2. *Fusion of single source data*

The fusion of single source data consists of re-evaluating the variable when new data is available by taking into account the previous values. It consists of an update, since the new data corresponds to the innovation of the observed system. The relative importance of the history (the previous data) and of the innovation can be adjusted by the fusion process based on criteria that have to be optimized. It is also possible to define a forgetting coefficient which makes it possible to minimize or eliminate the effect of older measurements.

11.3.3. *Temporal registration*

Sensors observe quantities that vary with time. In order to be able to fuse the data, we have to ensure that it is obtained from measurements that correspond to the same acquisition time. If this is not the case, a temporal registration mechanism has to be implemented in order for the measurements which appear to be acquired at the same time. In multi-sensor fusion, this mechanism will only be implemented if the temporal characteristics of the sensors are too different. In single sensor fusion, the data always has to be registered. This registration mechanism is the major characteristic of temporal data fusion.

11.4. Dating measurements

Whether in single source or multi-source fusion, evaluating variable data requires taking into account the time and particularly the moment when the measurement was taken. Evaluating the acquisition times of measurements is one of the major problems of temporal fusion [ALL 01]. In the most favorable case, the sensors are physically close to one another and the acquisition of measurements is supervised by the same electronic device. The measurements and their associated data are then dated by a single clock common to all of the sensors. On the other hand, in the most common case, the sensors operate independently and all have their own clocks. In this case, it is necessary to account for this multiplicity of clocks. Here are the most common solutions.

Registering the clocks. At certain times, a master clock sends a registration signal to all of the slave clocks so that they may be synchronized. Between two synchronization times, the delays between the clocks are considered to be small and to cause only few mistakes in the measurements.

Common clock. The sensors have the possibility of operating with the help of an external clock which will be the common clock. This requires for the clock signal to travel through the entire device.

It is sometimes very difficult to determine the acquisition date, particularly when the measurement provided by the sensor is evaluated from a time measurement. This is the case, for example, when using an ultrasound sensor for which the distance measurement is determined from the time of flight of the wave traveling from the sensor's transmitter, to the object being studied and finally to the sensor's receiver. The acquisition date is considered equal to either the wave's emission date or its reception date.

Usually, the inaccuracy of measurement dating is not taken into account. The error made on the acquisition date is assumed to lead to a negligible error in the data.

11.5. Evolutionary models

The method used to register data in time is based on the use of evolutionary models. It can also be used to make up for the lack of data in time, for example, when the sampling frequency is too small or in the case of a significant delay.

An evolutionary model is a knowledge model used for estimating a quantity at a given time, knowing the value or values of this same quantity at previous times. The model can be obtained by learning or from expert knowledge based, for example, on the laws of physics, mechanics, physiology, etc. When the model is known, it is then usually possible to predict the data in advance and particularly on the date of the data registration.

This concept of model can be illustrated by the following example. If we know at a time t the position $x(t)$ and the speed v of an object along a line, it is possible, using the well-known laws of kinematics, to predict the position $t+\Delta t$ on the condition that Δt is "small enough" compared to the possible variations of speed:

$$x(t+\Delta t) = v.\Delta t + x(t).$$

The model should also allow the error of the new data to be calculated knowing those of the previous data.

Let us assume, in the previous example, that the error on x and on v is modeled by an interval:

$$x(t) \in \left[x^-(t); x^+(t)\right]$$
$$v \in \left[v^-; v^+\right]$$

Then the lower and upper limits of $x(t+\Delta t)$ are:

$$x^-(t+\Delta t) = v^-.\Delta t + x^-(t)$$
$$x^+(t+\Delta t) = v^+.\Delta t + x^+(t)$$

and therefore:

$$x(t+\Delta t) \in \left[x^-(t+\Delta t); x^+(t+\Delta t)\right] = \left[v^- \cdot \Delta t + x^-(t); v^+ \cdot \Delta t + x^+(t)\right]$$

Note that the interval used to model the error has increased by $\Delta t \cdot (v^+ - v^-)$ over the course of the prediction:

$$\begin{aligned} x^+(t) - x^-(t) &< x^+(t+\Delta t) - x^-(t+\Delta t) \\ &< \big(v^+ \cdot \Delta t + x^+(t)\big) - \big(v^- \cdot \Delta t + x^-(t)\big) \\ &< \big(x^+(t) - x^-(t)\big) + \Delta t \cdot \big(v^+ - v^-\big) \end{aligned}$$

Evolutionary models come in very diverse forms, such as differential equations, recursive equations, evolution graphs, logical rules and others. They are merely approximations of the evolution of the observed variables. Therefore, they lead to errors in the value of the predicted data, errors that need to be estimated and taken into account during the fusion operation with the measured value. Generally speaking, the evolutionary model M_X of the variable X and of the confidence Conf_X assigned to it is defined in terms of accuracy and/or reliability, by equation [11.1], to obtain the predicted values $\widehat{X}$ of this variable and of the associated confidence $\mathrm{Conf}_{\widehat{X}}$:

$$M_X\left(\begin{bmatrix} X(t) \\ \mathrm{Conf}_X(t) \end{bmatrix}, \Delta t\right) = \begin{bmatrix} \widehat{X}(t+\Delta t) \\ \mathrm{Conf}_{\widehat{X}}(t+\Delta t) \end{bmatrix} \qquad [11.1]$$

For example, let us assume that we are observing a discrete system described by the following rule:

$$\text{if } X(t) = E_1, \text{ then } X(t+\Delta t) = E_2$$

where E_1 and E_2 are two possible states. We can introduce uncertainty in this modeled rule by using a probability distribution:

$$\text{if } X(t) = E_1, \text{ then } p\big(X(t+\Delta t) = E_1\big) = p_1 \text{ and } p\big(X(t+\Delta t) = E_2\big) = p_2$$

This rule can be interpreted as the definition of a conditional probability:

$$\begin{aligned} p\big(X(t+\Delta t) = E_1 / X(t) = E_1\big) = p_1 \\ p\big(X(t+\Delta t) = E_2 / X(t) = E_1\big) = p_2 \end{aligned}$$

If we know the probability for the state to be E_1 at t, described by $p(X(t) = E_1) = p$, then, by conditioning, we infer:

$$\begin{aligned} p\big(X(t+\Delta t) = E_1\big) &= p\big(X(t+\Delta t) = E_1 / X(t) = E_1\big) \cdot p\big(X(t) = E_1\big) = p_1 \cdot p \\ p\big(X(t+\Delta t) = E_2\big) &= p\big(X(t+\Delta t) = E_2 / X(t) = E_1\big) \cdot p(X(t) = E_1\big) = p_2.p \end{aligned}$$

In addition, we notice here that the uncertainty has increased, since the probability initially assigned to the state E_1 was divided at the time $t + \Delta t$ between the states E_1 and E_2.

11.6. Single sensor prediction-combination

Associating combination and prediction functions is the basic mechanism for achieving temporal data fusion. Its implementation depends essentially on the formalism used for representing the data (probability, possibility, evidence mass). The best known is the Kalman filter, which is based on probability theory. See [ABI 92, BAR 88, KAL 60] for a detailed description.

In more general terms, the method relies on the alternate use of prediction and combination mechanisms. Let us assume that we have an evolutionary model M_X such as it was defined in the previous section by equation [11.1], as well as a model for the sensor H_X, such that for any acceptable value of X, we can infer the value Y from the sensor's measurement. Finally, let us assume that we know the inverse model H_X^{-1} of this sensor that can be used to estimate X from Y.

First, we initialize at the time t_0 the data $X(t_0)$ at a value as close as possible to the actual value we wish to determine, which is based either on prior knowledge or on a first measurement $Y(t_0)$. We also initialize the confidence $\mathrm{Conf}_X(t_0)$ of this first value in terms of reliability and/or accuracy. A new measurement $Y(t_1)$ is acquired at the time $t_1 > t_0$, to which we assign a confidence $\mathrm{Conf}_Y(t_1)$. The data $X(t_0)$ is predicted up until the time t_1 by using an evolutionary model M_X:

$$M_X\left(\begin{bmatrix} X(t_0) \\ \mathrm{Conf}_X(t_0) \end{bmatrix}, \Delta t\right) = \begin{bmatrix} \widehat{X}(t_1/t_0) \\ \mathrm{Conf}_{\widehat{X}}(t_1/t_0) \end{bmatrix} \qquad [11.2]$$

where $\Delta t = t_1 - t_0$, $\widehat{X}(t_1/t_0)$ is the prediction of X at the time t_1 knowing all of the measurements up until t_0 and $\mathrm{Conf}_{\widehat{X}}(t_1/t_0)$ is the prediction of Conf_X at the time t_1 knowing all of the measurements up until t_0.

We also calculate:

$$H_X^{-1}\left(\begin{bmatrix} Y(t_1) \\ \mathrm{Conf}_Y(t_1) \end{bmatrix}\right)$$

At the time t_1, the data $\widehat{X}(t_1/t_1)$ and its confidence $\mathrm{Conf}_{\widehat{X}}(t_1/t_1)$ are estimated by a conjunctive combination Comb of the data's history, represented by $\widehat{X}(t_1/t_0)$, and the innovation resulting from the measurement $Y(t_1)$:

$$\mathrm{Comb}\left(\begin{bmatrix} \widehat{X}(t_1/t_0) \\ \mathrm{Conf}_{\widehat{X}}(t_1/t_0) \end{bmatrix}, H^{-1}\left(\begin{bmatrix} Y(t_1) \\ \mathrm{Conf}_Y(t_1) \end{bmatrix}\right)\right) = \begin{bmatrix} \widehat{X}(t_1/t_1) \\ \mathrm{Conf}_{\widehat{X}}(t_1/t_1) \end{bmatrix} \qquad [11.3]$$

We notice in equation [11.3] that all of the variables are referenced at t_1 and can therefore be combined. During the prediction phase, the confidence should decrease or, if the model corresponds exactly to reality, remain constant. During the combination phase, since this is conjunctive, the confidence will increase according to the quality of the measurement.

It is common for the sampling frequency to be too small for the application, i.e. intermediate values of X between two sampling times would be needed. In this case, we simply have to use the evolutionary model to register X with the times we are interested in, without having to use new measurements for the combination.

11.7. Multi-sensor prediction-combination

We now wish to combine information gathered from different sources [ELE 96]. We saw in section 11.3.1 that a number of problems could arise from the synchronization of sensors. We will deal with the most general case, where all of the sensors are completely asynchronous, i.e. they have completely independent sampling frequencies and processing times. The only hypothesis, which is usually true, is that the sensors operate "monotonically", i.e. for each sensor, the measurements are transmitted in the same order they were acquired in. In the rest of this section, we will describe the method for variables only, in order to have simpler equations. Obviously, in practice, it is also necessary to deal with the confidence in the variables.

We begin by studying the case of two sensors C_1 and C_2 before generalizing to any number of sensors. Let us assume that these two sensors deliver dated measurements $Y_1(t)$ and $Y_2(t)$. We have at our disposal the inverse methods of these sensors which make it possible to obtain estimates of $X_1(t)$ and $X_2(t)$ based solely on these measurements. From now on, we will say that these sensors directly provide these two estimates, thus implying that they are obtained from measurements. We will denote by t_a the acquisition time and by t_t the transmission time such as they are represented in Figure 11.6.

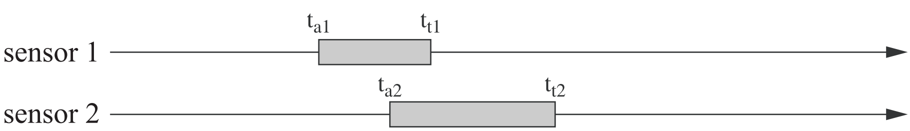


Figure 11.6. *Example of two asynchronous sensor fusion*

Let us assume that we know an estimate of X at the time t_0 denoted by $\widehat{X}(t_0)$ and that $t_{a,1} < t_{a,2}$ and $t_{t,1} < t_{t,2}$.

A $t_{t,1}$

The estimate $X_1(t_{a,1})$ obtained from the measurement $Y_1(t_{a,1})$ becomes available. We assume that this estimate is more recent than that at t_0 and therefore that $t_{a,1} > t_0$. Then, we register $\widehat{X}(t_0)$ until $t_{a,1}$ by using the evolutionary model to obtain:

$$\widehat{X}\left(t_{a,1}/t_0\right) = M_X\left(\widehat{X}\left(t_0\right), \left(t_{a,1} - t_0\right)\right)$$

We can then combine these two estimates that correspond to the same date:

$$\widehat{X}\left(t_{a,1}/t_{a,1}\right) = \mathrm{Comb}\left(\widehat{X}\left(t_{a,1}/t_0\right), X_1\left(t_{a,1}\right)\right)$$

Finally, we predict this value until the present time:

$$\widehat{X}\left(t_{t,1}/t_{a,1}\right) = M_X\left(\widehat{X}\left(t_{a,1}/t_{a,1}\right), \left(t_{t,1} - t_{a,1}\right)\right)$$

However, we bear $X_1(t_{a,1})$ and $\widehat{X}(t_{a,1}/t_{a,1})$ in mind.

A $t_{t,2}$

The estimate $X_2(t_{a,2})$ obtained from the second sensor becomes available. We assume that this estimate is more recent than that at $t_{a,1}$. This time we register $\widehat{X}(t_{a,1})$ until $t_{a,2}$ by using the evolutionary model to obtain:

$$\widehat{X}\left(t_{a,2}/t_{a,1}\right) = M_X\left(\widehat{X}\left(t_{a,1}/t_{a,1}\right), \left(t_{a,2} - t_{a,1}\right)\right)$$

We can then combine these two estimates:

$$\widehat{X}\left(t_{a,2}/t_{a,2}\right) = \mathrm{Comb}\left(\widehat{X}\left(t_{a,2}/t_{a,1}\right), X_2\left(t_{a,2}\right)\right)$$

Finally, we predict this value up until the present time:

$$\widehat{X}\left(t_{t,2}/t_{a,2}\right) = M_X\left(\widehat{X}\left(t_{a,2}/t_{a,2}\right), \left(t_{t,2} - t_{a,2}\right)\right)$$

In this example with two sensors, we have assumed that the measurements on the two sensors arrived in the same order they were in when they were acquired, i.e., if $t_{t,1} < t_{t,2}$, then $t_{a,1} < t_{a,2}$. This condition is not always satisfied, particularly in the case where a sensor has a long processing time compared with the sampling period of another sensor. However, this sensor can provide us with relevant information even if it is delayed and there is no point in working without it. We will now present the method in the general case.

We now have at our disposal a certain number of sensors providing elements of information with the dates $t_{a,i}$. These times are arranged chronologically so that $t_{a,1} < t_{a,2} < t_{a,3} <$, etc. For each of these instants, we have stored $X(t_{a,i})$ obtained directly from the measurement and $\widehat{X}(t_{a,i}/t_{a,i})$, which is the estimate of X that takes into account all of the available measurements up until $t_{a,i}$. Table 11.1 shows all of the variables that are stored.

$t_{a,1}$	$t_{a,2}$	$t_{a,3}$	$t_{a,4}$
$X(t_{a,1})$ $\widehat{X}(t_{a,1}/t_{a,1})$	$X(t_{a,2})$ $\widehat{X}(t_{a,2}/t_{a,2})$	$X(t_{a,3})$ $\widehat{X}(t_{a,3}/t_{a,3})$	$X(t_{a,4})$ $\widehat{X}(t_{a,4}/t_{a,4})$

Table 11.1. *Table of the variables stored at the different times*

At the time t_t^*, we have a new measurement that gives us $X(t_a^*)$, such that $t_{a,2} < t_a^* < t_{a,3}$. Table 11.1 then changes into Table 11.2.

$t_{a,1}$	$t_{a,2}$	t_a^*	$t_{a,3}$	$t_{a,4}$	t_t^*
$X(t_{a,1})$ $\widehat{X}(t_{a,1})$	$X(t_{a,2})$ $\widehat{X}(t_{a,2})$	$X(t_a^*)$	$X(t_{a,3})$ $\widehat{X}(t_{a,3})$	$X(t_{a,4})$ $\widehat{X}(t_{a,4})$	

Table 11.2. *Table of the variables at the current time t_t^**

This new estimate has to be taken into account as well as all of those that followed it. The method consists of alternately using prediction and combination until obtaining the prediction at the time t_t^*. This method is illustrated by Table 11.3

Given that the sensors behave "monotonically", when a measurement is provided by the sensor C_i, the previous measurement from the same sensor is forgotten. As a result, the storing table has a number of columns equal to the number of sensors.

$t_{a,1}$	$t_{a,2}$	t_a^*	$t_{a,3}$	$t_{a,4}$	t_t^*
$X(t_{a,1})$ $\widehat{X}(t_{a,1})$	$X(t_{a,2})$ $\widehat{X}(t_{a,2})$	$X(t_a^*)$ $\downarrow$ $\rightarrow\ \widehat{X}(t_a^*/t_{a,2})\ \rightarrow\ \widehat{X}(t_a^*)$	$X(t_{a,3})$	$X(t_{a,4})$	

Taking the measurement into account at t_a^*

$t_{a,1}$	$t_{a,2}$	t_a^*	$t_{a,3}$	$t_{a,4}$	t_t^*
$X(t_{a,1})$ $\widehat{X}(t_{a,1})$	$X(t_{a,2})$ $\widehat{X}(t_{a,2})$	$X(t_{t_a^*})$ $\widehat{X}(t_a^*)$	$X(t_{a,3})$ $\downarrow$ $\rightarrow\ \widehat{X}(t_{a,3}/t_a^*)\ \rightarrow\ \widehat{X}(t_{a,3})$	$X(t_{a,4})$	

Taking the measurement into account at $t_{a,3}$

$t_{a,1}$	$t_{a,2}$	t_a^*	$t_{a,3}$	$t_{a,4}$	t_t^*
$X(t_{a,1})$ $\widehat{X}(t_{a,1})$	$X(t_{a,2})$ $\widehat{X}(t_{a,2})$	$X(t_{t_a^*})$ $\widehat{X}(t_a^*)$	$X(t_{a,3})$ $\widehat{X}(t_{a,3})$	$X(t_{a,4})$ $\downarrow$ $\rightarrow\ \widehat{X}(t_{a,4}/t_{a,3})\ \rightarrow\ \widehat{X}(t_{a,4})$	

Taking the measurement into account at $t_{a,4}$

$t_{a,1}$	$t_{a,2}$	t_a^*	$t_{a,3}$	$t_{a,4}$	t_t^*
$X(t_{a,1})$ $\widehat{X}(t_{a,1})$	$X(t_{a,2})$ $\widehat{X}(t_{a,2})$	$X(t_{t_a^*})$ $\widehat{X}(t_a^*)$	$X(t_{a,3})$ $\widehat{X}(t_{a,3})$	$X(t_{a,4})$ $\widehat{X}(t_{a,4})$	$\rightarrow\ \widehat{X}(t_t^*/t_{a,4})$

Prediction up to the current time

Table 11.3. *Tables representing the data updating method*

11.8. Conclusion

Whenever trying to track the evolution of variables in time, it is essential to study the temporal constraints:

– the need for information set by the application (frequency of data availability);

– the temporal features of the sensors and of the associated data processes (frequency, delay);

– the temporal characteristics related to the sensors.

After listing all of these, the performances of the sensors can turn out to be good enough as to not require the use of mechanisms specific to temporal fusion (the acquisition frequency is high enough, the delays are negligible). If this is not the case, a system has to be implemented for assigning dates to the data and evolutionary models are needed in order to achieve temporal registration.

There are few studies that have directly focused on temporal fusion, or more generally on time management in data fusion applications. This aspect appears mostly in military applications for target tracking and in robotics. In most cases, time management is limited to the use of the Kalman filter, often in its extended version (linearization around an operating point). The quality of the fusion results could probably be significantly improved if, in critical applications, the temporal aspect was not plainly ignored.

11.9. Bibliography

[ABI 92] ABIDI M.A., GONZALEZ R.C., *Data Fusion in Robotics and Machine Intelligence*, Academic Press, 1992.

[ALL 01] ALLOUCHE M.K., "Classification of Temporal Information in Situation Analysis", ISIF, *Fusion 2001*, vol. 2, International Conference on Information Fusion, p. ThC3 11–18, Montreal, Quebec, Canada, 2001.

[ANK 01] ANKEN C.S., BURNS C.L., "Theater Ballistic Missile (TBM) Reasoner: A Knowledge Base Decision Aid for Time Critical Targeting Situation Assesment", ISIF, *Fusion 2001*, vol. 2, International Conference on Information Fusion, p. FrA2 13–18, Montreal, Quebec, Canada, 2001.

[APP 98] APPRIOU A., "*Uncertain Data Aggregation in Classification and Tracking Processes*", Aggregation and Fusion of Imperfect Information chapter, Physica-Verlag, 1998.

[BAR 88] BAR-SHALOM Y., FORTMANN T., *Tracking and Data Association*, vol. 179, Academic Press, 1988.

[ELE 96] ELETER B., ROMBAUT M., "Intelligent Mapping of the ProLab2 Vehicle Dynamic Environment", *Mathematics and Computer in Simulation*, vol. 41, no. 3-4, p. 329–336, 1996.

[ETE 94] ETER B.E., Etude d'un système temps réel de gestion et de fusion de données, PhD Thesis, Compiègne University of Technology, December 1994.

[JOU 99] JOUANNIN S., Association et fusion de données: application au suivi et à la localisation d'obstacles par radar à bord d'un véhicule routier intelligent, PhD Thesis, Blaise Pascal University, Clermont-Ferrand, 12 January 1999.

[KAL 60] KALMAN R.E., "A New Approach to Linear Filtering and Prediction Problem", *Trans ASME*, vol. 82, p. 34–45, March 1960.

[KOS 93] KOSAKA A., MENG M., KAK A.C., "Vision-Guided Mobile Robot Navigation Using Retroactive Updating of Position Uncertainty", *IEEE International Conference on Robotics and Automation*, Atlanta, USA, 1993.

[PAO 94] PAO L.Y., "Centralized Multisensor Fusion Algorithms for Tracking Applications", *Control engineering practice*, vol. 2, no. 5, p. 875–887, 1994.

[PRU 99] PRUSKI A., HABERT O., "Obstacle Avoidance for the VAHM Smart Wheelchair", *Automation'99*, Warsaw, Poland, March 1999.

[ROM 95] ROMBAUT M., "Génération temps réel de données numériques/symboliques par fusion temporelle multi-capteurs", *Revue Traitement du Signal*, vol. 12, no. 5, p. 317–326, 1995.

[TRA 93] TRASSOUDAINE L., Solutions multisensorielles temps réel pour la détection d'obstacles sur route, PhD Thesis, Blaise Pascal University, Clermont-Ferrand, 16 February 1993.

[TRA 94] TRASSOUDAINE L., HUTBER D., CHECCHIN P., ALIZONA J., GALLICE J., THONNAT M., "Building an Environment Map around the ProLab2 Vehicle using a Controllable Range Sensor", *Intelligent Vehicles'94*, Paris, p. 562–567, 1994.

Chapter 12

Conclusion

Now that we have been through this overview of the major numerical fusion methods and of their use in signal and image processing and in robotics, we will sum up a few achievements and conclusions, as well as a few issues that still pose difficulties.

12.1. A few achievements

The previous chapters have shown that a wide variety of numerical techniques are used for the fusion of imprecise and uncertain information. This diversity is the result of the diversity of the tasks themselves that contribute to the decision in a multi-source information system. Probabilistic methods remain the most commonly used, mostly because they have led to the development of operational tools and great know-how, which is the result of a considerable amount of practice. On the decision level these tools turn out to be particularly efficient, whereas for modeling, some aspects remain limited or even disputed. Fuzzy set theory relies on a type of modeling close to intuition. In the fusion applications mentioned here, there is still little formalism or development in the decision phase. On the other hand, the combination phase is very rich and allows knowledge of any type to be included. Belief function theory offers the most powerful modeling tools, making it possible to simply and efficiently include knowledge, imprecision and uncertainty. Combination, as it is used in signal and image processing, is limited to the conjunctive mode.

In numerical fusion for signal and image processing, the efforts of the past years have led to a better understanding of the different theories taken from fields such as

Chapter written by Isabelle BLOCH.

belief functions and fuzziness. We thus know now which application frameworks are right for these theories, their advantages and their limits in these application fields; we also know how to represent and model information and numerical, symbolic or structural data in each of the formalisms, and how to achieve their combination. There have been many new developments, particularly for the traditional multi-source classification, structure or object recognition in images, tracking, localizing and planning applications.

12.2. A few prospects

Despite this progress, some aspects remain poorly understood or require further development.

Taking into account the origin of the data and knowledge, as well as the relations between sources, is still often performed under supervision and therefore requires little experience. One of the important questions involves the independence between sources and conditioning (particularly in the case of sequential fusion, in dynamic updating processes). The probabilistic framework offers methods to test statistical independence, and those are usually the only tools at our disposal. But this type of independence is often considered too restricting, and in other contexts, such as belief function theory, the preferred concept is cognitive independence, which is related to how the knowledge and data are acquired rather than to their nature [SHA 76, SME 90]. When choosing the operators of fuzzy and possibilistic theories, independence results in the operators being idempotent, whereas dependence requires reinforcement.

A very difficult question is conflict management. Insofar as it is possible, the sources of the conflict have to be identified and made explicit, in order to avoid inconsistencies when making the decision. In particular, it is not always easy to tell the difference between conflict and the complementarity of sources, or to know whether it should be resolved or not. Conflict, which can be referred to as "apparent", is actually a form of complementarity. For example, if a source systematically includes class B in class A whereas another source is good at distinguishing them, these two sources seem to be in conflict. Recognizing complementarity often requires the use of prior (or learned) knowledge regarding the low possibilities for the first source to tell the two classes apart. Resolving such conflicts is easy once they have been properly identified. A second form of conflict, which is real this time, is due to inconsistencies between sources, which are caused by their limited reliabilities, by changes that occur in the scene between acquisitions, or also by the fact that they are not dealing with the same thing. This type of conflict is more difficult to identify and to resolve. It is sometimes even preferable to simply indicate it and not to try to solve it because it often corresponds to a fundamental inadequacy of our knowledge of the problem.

Choosing and evaluating methods is as crucial as it is difficult. Again, there is no general solution for choosing methods that are adapted to the different types of information and knowledge handled, or to the applications we may have in mind. Evaluating methods can be more or less easy depending on whether the truth is accessible or not. Attempts to compare numerical fusion methods, when applied to the same problems, have given contradicting results and therefore have failed. We think the main reason for this is that each problem is expressed more easily in one theory than in another, so solving them with the wrong tools requires these techniques to become distorted, and does not make much sense.

Finally, in the case of image processing applications, and also for certain applications in robotics, the introduction of spatial information in fusion is an important point, for which the set of existing methods could benefit from further development. Particularly, the recent successes achieved by taking into account structural information shows the advantage of combining spatial information from different levels.

We have noticed in the previous chapters that each approach is adapted to a limited set of imperfections in the information to fuse. It is rare for all of the imperfections to be modeled simultaneously and in a simple way in a unique theory. Thus, a field of investigation that remains open involves the fusion of methods or the combined use of different complementary formalisms. These studies on method combination are promising, because their goal is to use the advantages of the different theories in order to make them cooperate with each other. This combination can rely on relations that exist between the different methods. For example, a probability can be interpreted as a particular mass function, a belief function whose focal elements are such that each one is included in the next can be interpreted as a possibility distribution, a possibility distribution can be interpreted as confidence intervals or as a family of probabilities (see, for example, [DUB 99] for details on the connections between possibilities and probabilities), etc. Thus, studies have already been conducted to combine the imprecision represented by fuzzy sets with probabilistic uncertainty (for example, [CAI 93, PIE 94, SAL 95] in a Markovian classification method where the classes are fuzzy), to combine Markov fields with belief functions [BEN 97, HEG 98], or to work with belief functions whose focal elements are fuzzy [SME 81, YAG 82, YEN 90, ZAD 79].

12.3. Bibliography

[BEN 97] BENDJEBBOUR A., PIECZYNSKI W., “Segmentation d’images multisenseur par fusion évidentielle dans un contexte markovien”, *Traitement du Signal*, vol. 14, no. 5, p. 453-464, 1997.

[CAI 93] CAILLOL H., HILLION A., PIECZYNSKI W., “Fuzzy Random Fields and Unsupervised Image Segmentation”, *IEEE Trans. on Geoscience and Remote Sensing*, vol. 31, no. 4, p. 801-810, 1993.

[DUB 99] DUBOIS D., PRADE H., YAGER R., "Merging Fuzzy Information", in J. BEZDEK, D. DUBOIS and H. PRADE (ed.) *Handbook of Fuzzy Sets Series, Approximate Reasoning and Information Systems*, Chapter 6, Kluwer, 1999.

[HEG 98] HÉGARAT-MASCLE S.L., BLOCH I., VIDAL-MADJAR D., "Introduction of Neighborhood Information in Evidenve Theory and Application to Data Fusion of Radar and Optical Images with Partial Cloud Cover", *Pattern Recognition*, vol. 31, no. 11, p. 1811-1823, 1998.

[PIE 94] PIECZYNSKI W., CAHEN J., "Champs de Markov cachés flous et segmentation d'images", *Rev. Statistique Appliquée*, vol. XLII, no. 3, p. 13-31, 1994.

[SAL 95] SALZENSTEIN F., PIECZYNSKI W., "Unsupervised Bayesian Segmentation using Hidden Fuzzy Markov Fields", *IEEE Int. Conf. on Acoustics, Speech and Signal Procesing*, Detroit, Michigan, 1995.

[SHA 76] SHAFER G., *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[SME 81] SMETS P., "The Degree of Belief in a Fuzzy Event", *Information Sciences*, vol. 25, p. 1-19, 1981.

[SME 90] SMETS P., "The Combination of Evidence in the Transferable Belief Model", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, p. 447-458, 1990.

[YAG 82] YAGER R.R., "Generalized Probabilities of The Degree of Fuzzy Events from Fuzzy Belief Structures", *Information Sciences*, vol. 28, p. 45-62, 1982.

[YEN 90] YEN J., "Generalizing the Dempster-Shafer Theory to Fuzzy Sets", *IEEE Transactions on Systems, Man and Cybernetics*, vol. 20, no. 3, p. 559-569, 1990.

[ZAD 79] ZADEH L.A., "A Theory of Approximate Reasoning", *Machine Intelligence*, vol. 9, p. 149-194, 1979.

Appendix A

Probabilities: A Historical Perspective

This appendix is largely inspired by [BLO 96].

Among the different methods for representing knowledge, numerical methods, which attempt to model the imprecision and uncertainty of data and knowledge, are widely used for problems as diverse as multi-criteria aggregation, combining testimonies, or fusing heterogenous images. Probabilistic methods certainly are the most popular, but still give rise to a number of controversies, particularly between frequentist or objectivist methods and subjectivist methods. Although subjectivists seem to be taking over in many fields, frequentist concepts are still of great practical use, particularly when it comes to learning a law based on large samples, for example, to recognize cultivations in an aerial image.

A historical overview of the different meanings of probability can help explain the causes of these controversies and show that the choice of a method can be thought through and justified by the problem at hand and by our interpretation of probability. section A.1 will focus on this historical perspective and section A.2 on the characterization of different classes of probabilities. This is largely based on review articles cited as references.

It seems remarkable that the hypothesis of the additivity of probabilities[1], which is widely recognized today, only appeared so late. This hypothesis is stated as an axiom

Chapter written by Isabelle BLOCH.

1. The additivity relation expresses the fact that for two exclusive events A and B, the probability of the union denoted by $A + B$ is equal to $p(A + B) = p(A) + p(B)$. Particularly, we infer $p(A) + p(\bar{A}) = 1$, where $\bar{A}$ denotes the opposite of A (or its complement in set theory terms).

in Kolmogorov theory. However, the works of Cox show that these axioms can be inferred from a certain number of basic postulates prompted by intuition (see section A.3).

In section 6.10, we gave a few examples showing the limits of additive probabilities, which are often the results of the strict constraints that they impose. The modification of basic postulates in order to overcome these limits leads to different numerical theories, which no longer satisfy the same properties, and thus leads us again to methods such as fuzzy sets or Dempster-Shafer belief theory. The latter has often been criticized on the grounds that Dempster's orthogonal combination rule had no theoretical justifications. Several authors have responded to this criticism, and in Appendix B we will present Smets's arguments, which allow this rule to be inferred from more easily justifiable axioms. We will then establish the relation between these axioms and those given by Cox, in order to explain the origins of the differences between the two theories.

A.1. Probabilities through history

This study was inspired by Shafer's works and particularly by his remarkable review articles on the history of science [SHA 78, SHA 86]. The following historical presentation is largely based on these works. They constitute the basic points, and additional information was added from articles or books [COX 46, DEM 93, DUB 88, GOO 59, HOR 86, JAY 57, JEF 61, KEM 42, NEA 92, SOM 89, TRI 72].

A.1.1. *Before 1660*

The conflict between knowledge and opinion appeared as early as the Classical era, in particular with Plato, in roughly 400 BC, and terms such as necessary, possible, probable, began to be defined. We find for example with Aristotle (roughly 350 BC) assertions of the type "if an event is necessary, then its opposite is impossible" (until possibility theory was developed, no consistent theory was capable of modeling this sentence) or also "what is probable is what usually occurs" (in reference to phenomena repeating themselves, which are the basis of frequentist theory). For the Ancients, there were three epistemological categories. In the first, certain knowledge is possible. This is equivalent to Plato's concept of knowledge or science. The second category is comprised of events for which knowledge is probable or possible. This corresponds to Plato's concept of opinion and, in this sense, probability is seen as an attribute of opinion. The third category, which is absent in Plato's philosophy, corresponds to events for which no knowledge is possible, hence to the realm of chance. This term is used to mean the lack of statistical regularity. Transposed in more modern terms, these concepts correspond, in our opinion, to deductive reasoning for the first category and to inductive reasoning for the second. The third corresponds to phenomena that (apparently) do not follow any law that is predictable or that has been learned. This third category would seem to rule out any possibility of a mathematical theory for chance.

These concepts were still very primitive, and no theory had yet been developed. However, their importance can be disputed, since they are related to the theory of knowledge, considered to be essential, and the focus of the work of many philosophers. The fact that probability was considered as a "guide to life" by Cicero (60 BC) stands as proof of this. It is difficult to resist the pleasure of repeating these few sentences by Seneca, quoted in [MAT 78]: "There are differences of interpretation, however, between our countrymen and the Etruscans, the latter of whom possess consummate skill in the explanation of the meaning of lightning. We think that because clouds collide, therefore lightning is emitted; they hold that clouds collide in order that lightning may be emitted. They refer everything to the will of God: therefore they are strong in their conviction that lightning does not give an indication of the future because it has occurred, but occurs because it is meant to give this indication" (Seneca, *Natural Questions*, II, 32).

This passage is a good illustration of the subjective nature of opinion; it is impossible to imagine an experiment that would make it possible to prove or refute either one of these opinions. This is also an illustration of the difference between causality and logical links, which are often not distinguished. Probabilities express logical links but not causality relations [DEM 93].

These epistemological categories later disappeared, for unknown reasons. During the Renaissance, two completely independent concepts were used: chance, or randomness, and probability, which is seen as an attribute of opinion and with no numerical value attached to it. The concept of chance is strongly related to game theory. Its premises can be found in Dante's *Purgatory* (1310), in which he describes the different sums that can be obtained from rolling three dice. Game theory was later widely developed at the end of the 16th century and during the first half of the 17th century. Cardan (1560) and Galileo (1620) compiled the different results that can be obtained in a game, and counted the number of cases where each of these outcomes occurred. For the first time, the concept of "equiprobable cases" was mentioned. The origins of the mathematical probability theory are attributed to Pascal and Fermat (even though they did not use the word probability), since, in their correspondence (around 1654), they started solving the first non-trivial problems. In 1657, the first book on the subject of game theory, written by Huygens, was published. These three mathematicians and philosophers tried to solve the "point problem": a game between two players, which requires that one player has three points in order to win, remains unfinished; how, then, should the stakes be equitably divided if one player has one point and the other has two? They explained with equiprobable cases why proportional frequencies appeared in a wide series of outcomes. However, they were troubled in solving their problem because of the determinism imposed on them by Christianity. The Ancients, on the contrary, accepted indeterminism very well. To sum up, all of these works dealt with the problem of games and produced a theory of chance, but probability is never mentioned, even if the vocabulary used is slightly more epistemological. There was still significant confusion between statistics and prior knowledge, which led to two classes of probabilities being confused, based on frequencies (related to statistics) and based on equiprobable cases (related to prior knowledge).

A.1.2. *Towards the Bayesian mathematical formulation*

The first link between game theory and probabilities was given in 1662, when it was introduced by Arnauld in the *Art of Thinking*. Arnauld established an analogy between games and everyday life and suggested that an epistemological perspective of chance made it possible to apply the theory to probabilities (which were still considered as attributes of opinion). He stopped short of the concept of numerical probability, but his works clearly mark a milestone in the evolution of the concept of probability. The analogy between games and life was used up until the end of the 17$^{\text{th}}$ century by demographers who calculated life expectancy tables by using game theory, but without including the concept of probability.

A contribution from an entirely different field came from Leibniz, who suggested in the *De Conditionibus* (1665) to represent a person's legal rights using numbers. The absence of law was represented by 0, a pure law by 1 and a conditional law by a fraction between 0 and 1. This classification of rights relies on the condition upon which the law is founded: an impossible condition leads to the absence of law, if it is necessary the law is pure, if it is contingent[2], the law is conditional. These concepts of contingency and necessity are also used by Bernoulli regarding the problem of the combination of testimonies. Leibniz suggests relating the probability for the condition to exist with the "quantity" of the law and thus seemed to lean towards a numerical conception of probability, without relying on game theory. He only later becomes acquainted with this theory. Although his essays on chance provided nothing new from a mathematical perspective, they acknowledged the link between probability and game theory.

In the field of the combination of testimonies, the works of Hooper in 1699 (*A Calculation of the Credibility of Human Testimony*) lead to the definition of non-Bayesian confidence functions, which represent the credibility of a witness, as well as to two combination rules, one for consecutive testimonies and the other for simultaneous testimonies. These two rules, which were very popular in the 18$^{\text{th}}$ century, were completely abandoned in the 19$^{\text{th}}$ century.

2. Contingent is used here meaning that something may or may not occur.

One of the most important contributions to the relation between game theory and probability at the end of the 17th century is probably that of Bernoulli, particularly in his book *Ars Conjectandi* (published in 1713). Bernoulli suggested as early as 1680 a mathematical theory of probabilities and their combinations. He uses game theory to calculate probabilities, but does not remove from probabilities their epistemological aspect and their role in the judgment of individuals. Thus, the first formalisms that allowed numerical probabilities to be used focused on the study of subjective probabilities! For the most part, Bernoulli's theory focuses on probabilities that are subjective, and that are means of measuring knowledge. They are calculated based on the concept of "arguments", and their properties depend on the nature of these arguments. In particular, they are not always additive. The combination rules suggested by Bernoulli, which are more complete than Hopper's, take on different forms, depending on the arguments, and not all of these forms correspond to the usual probabilistic rules. In the last part of his work, Bernoulli established his famous law of large numbers. This law allows a prior unknown probability to be estimated afterwards based on the observation of occurrence frequencies. Therefore, this theorem has more to do with random probabilities than epistemological probabilities (according to the distinction later introduced by Lambert), since a chance that can only be known *a posteriori* is not initially a characteristic of our knowledge.

The successors of Bernoulli simplified his theory and unknowingly reduced its scope. First of all, they were not convinced by Bernoulli whose theory, which often seemed complicated to them, was not as well established as game theory. Additionally, they mostly held on to the law of large numbers and identified probability with the chance of appearing, thus leading to an essentially frequentist approach. We can mention, among the successors of Bernoulli, Montmort (*Essai d'analyse sur les jeux de hasard*, 1708), who attempts to apply game theory to other fields, and Moivre (*De Mensura Sortis*, 1711, *Doctrine of Chances*, 1718) where we find the first explicit additivity rule and a representation of probabilities between 0 and 1. His definition became the classical definition. These two authors mentioned probabilities but their theory often deals with chance. They defined probabilities as the ratio of the number of favorable cases to the number of possible cases, but the authors were faced with the problem of counting the number of cases, which is not possible in every field. Note that Moivre had already discovered the Gaussian distribution.

In the 18th century, only Lambert pursued Bernoulli's work and distinguished random probabilities and epistemological probabilities. The former are those that can be known *a priori*, as in game theory, or *a posteriori*, provided by experience. The latter are assigned to events by inferences based on effects or circumstances and are more subjective in nature. In *Photometrica* (1760), he focused on error theory and suggested a method known today as maximum likelihood. In *Neues Organon* (1764), he generalized Bernoulli's argument theory, corrected and generalized his combination laws and discussed the cases of games, syllogisms and of different types of testimonies. His laws are a specific example of Dempster's combination rule and, again, probabilities here are not additive.

The works of Bayes (*An Essay Towards Solving a Problem in the Doctrine of Chances*, 1763) later suggested a general approach that eliminated the distinction between random probabilities and epistemological probabilities. Bayes reversed Bernoulli's theorem, whereas Bernoulli estimated the number of successful outcomes based on the knowledge of probability, Bayes attempted to calculate the probability knowing the number of successful outcomes in a sample and expressed it in terms of initial, final and likelihood probabilities. He only dealt with additive probabilities. Bayes never tried to make his work known to others. His essay was found only after his death, along with other works (particularly on electrically charged bodies), which were written using abbreviations that have not all been deciphered [HOL 62]. This has led some to wonder about the exact origins of Bayes' theorem (see, for example, Stigler's works [STI 82, STI 83], who suggests a... Bayesian solution to this interrogation).

This work was continued by Laplace (*Théorie analytique des probabilités*, 1812). This era saw the rapid development of inverse probabilities, seen from a subjective perspective. This theory underlined the distinction between the "initial" (or prior) probability of a hypothesis, the "final" probability (after the experiment), and the "likelihood" probability (probability of the experiment knowing the hypothesis). Laplace also introduced the concept of insufficient reason: some outcomes are considered equally probable if there is no reason to think otherwise. This principle was widely adopted until the middle of the 20th century.

A.1.3. *The predominance of the frequentist approach: the "objectivists"*

In the 19th and early 20th century, because of the rapid development of physical sciences, the modeling of human reasoning was neglected. A new discipline emerged at that time: statistics. The concept of probability was often related to the observation of physical phenomena, to their repetition in long sequences. The theories of Bayes and Laplace were criticized for their subjective natures, accused of lacking rigor, and the concept of prior probability was rejected because it seemed too vague.

The works of Cournot (1843), Ellis (1843), Venn (1866) then defined physical probabilities, in terms of frequencies. As emphasized by Good [GOO 59], these works were faced with problems that were impossible to solve. For example, if by flipping a coin, we observe the following sequence of Heads (H) and Tails (T): THTHTHTHTH, etc., we can infer that the probability of getting Tails is 1/2, but this does not mean we can draw conclusions about the game's honesty.

One of the problems raised by these methods involves the length of the sequences used for calculating the frequencies. They have to be long, but how long? The theories were developed for infinite sequences, but did not specify how to proceed in practice. This was the case for Venn's limit, Fisher's hypothetical infinite population (1912), or von Mises's infinite random sequences (1919). Von Mises clearly stated the distinction between abstract mathematical theory and the application of this theory: the essential property of these infinite random sequences has to be that the probability of success must remain the same regardless of the sub-sequence (which is infinite), which is an abstract concept, and is restricted in particular to those fields where this definition is reasonable. His argument was that it is not necessary to actually repeat the experiment indefinitely for the probability to exist and therefore he limited himself to physical probabilities and random processes, excluding problems where we would ask ourselves, for example, the probability for X to die at age 60.

The 20th century also saw the development of the Gaussian distribution. Already known to Moivre, it was obtained by Gauss (1823) by using the maximum likelihood principle in problems about the estimation of the observation error. In the middle of the 19th century, it was rediscovered by Herschel, based on geometric considerations for estimating measurement errors in the positions of stars, and also by Maxwell while he studied the speed distributions of molecules in a gas [DEM 93].

However, despite the strong frequentist context of the time, distinctions similar to those made today were expressed. For example, Poisson, in his research on the probability of judgments (1837), distinguished chance and probability. Chance is specific to the event itself, regardless of our knowledge, whereas probability is related to our knowledge. This distinction is similar to Lambert's. The distinction between objective probability and subjective probability is also explicit in Cournot's *Exposition de la Théorie des Chances et des Probabilités* (1843). But even if the distinction is explicit, theories developed in the 19th century only make it possible to solve problems related to physical or objective probabilities.

Boole's works, in a way, constitute a link between frequentist and epistemological methods. In his book *Laws of Thought* (1854), he attempted to combine, on an epistemological level, evaluations made locally on various attributes of the information. He supported the idea that probabilities are obtained from frequencies, but acknowledged the impossibility of estimating, in many situations, the joint frequencies, which thus need to be generated on a subjective level.

A.1.4. *The 20th century: a return to subjectivism*

In the 20th century, traditional methods continued to be developed, with increasingly strong mathematical foundations, particularly under the impulse of Kolmogorov, and the frequentist method remained present and strong (particularly in signal and image processing), benefiting from the works of Neyman, Pearson, Feller [FEL 66]. At the same time, with the birth of artificial intelligence and its growing importance, human reasoning and the models used to describe it gained interest and led researchers to return to a more subjectivist conception of probabilities. Two schools of thought have appeared, one for additive properties and the other for non-additive properties.

The first school of thought relies on basic postulates to more rigorously and less arbitrarily obtain probabilities, their properties, Bayes' theorem, etc. Among them we can mention the works of Keynes [KEY 29], Kemble [KEM 42], Cox [COX 46], Jaynes [JAY 57], Jeffreys [JEF 61], Tribus [TRI 72]. We will discuss Cox's approach in detail in section A.3. Tribus's approach is directly inspired by it. Jaynes and Kemble worked in the field of statistical mechanics and showed that this requires a subjectivist approach. Jeffreys, based on a method similar to Cox's but in which numbers are included by convention, inferred the properties of probabilities from a certain number of principles. These principles reject others considered to be fundamental in other theories (for example, the definition of probabilities in terms of infinite sets of possible observations, in terms of world properties, the causality principle, etc.). The essence of his theory is that none of the direct probabilities, whether *a priori* or *a posteriori*, is a frequency. Even if the probability is calculated based on a frequency, it is not identical to the frequency and a reasonable degree of confidence is necessary before it can be used. The goal of Jeffreys's theory was not to justify inductive reasoning, but to ensure its mathematical consistency. In a very similar fashion, a great number of researchers studied the philosophical perspective of subjective probabilities with respect to objective probabilities, often without questioning additivity (Keynes, Jeffreys, Ramsey, de Finetti, Koopman, Russel, Carnap, Good, Savage, etc.). In particular, the works of Savage and Finetti showed that the Bayesian theory's subjective approach is the most justified and the most consistent [FIN 37]. In his works, de Finetti adopted a resolutely subjective approach (more than that of Cox) and reasoned in terms of the consistency of individual opinions, and even in terms of collective psychology in order to explain the coincidences in the opinions of different individuals. This method, which is particularly interesting, is described briefly in section A.3.

The second school of thought completely re-examined additivity, relying in particular on the works of pioneers such as Bernoulli and Lambert (see, for example, [GOO 59, SHA 86]). Koopman, in the 1940s, introduced the concept of lower and upper probabilities, thus defining a subjective probability with an inequality and no longer as a precise value, based on the works of Boole (*Laws of Thought*, 1854), which had already foreseen this evolution. Several other researchers followed up on his work (Good, Dempster, etc.). In particular, Dempster generalized Lambert's rules, which are only able to deal with arguments involving a single conclusion, in the case when several hypotheses need to be considered. Applications of these new theories can be found in the field of economics, where Shackle, for example, suggested economic models relying on concepts close to possibility theory, or in the field of legal precedents, particularly the works of Ekelöf (*Rättegang*, 1963) who suggested three operators for combining testimonies: the combination of consecutive testimonies and that of simultaneous corroborating testimonies follow Hooper's rules, whereas the combination of conflicting testimonies is related to a specific case of Lambert's rule.

Starting in the 1960s, theories appeared that were no longer directly related to probabilities. Zadeh invented fuzzy sets in 1965 [ZAD 65], Shortliffe and Buchanan constructed the MYCIN system based on the concept of certainty factors in 1975 [SHO 75], Shafer developed belief theory (*A Mathematical Theory of Evidence*, 1976) [SHA 76] and Zadeh introduced possibility theory in 1978 [ZAD 78].

A.2. Objectivist and subjectivist probability classes

Section A.1 showed that there are several classes of probabilities, summed up here, based on the classification set out by Good [GOO 59]:

1) the traditional definition is provided by game theory and relies essentially on the concept of equally probable cases. Calculating the probabilities is achieved using frequencies of occurrence, by counting all of the cases;

2) a more subjective version of this definition includes additional information related to the knowledge we have, for example, knowledge of the honesty, or lack thereof, of a game. Therefore, this second class only considers conditional probabilities. We can also include in this class Savage's or Finetti's subjective probabilities, which are estimated proportionally to the sum of money that a person would be willing to give if what it claimed turned out to be false [DUB 88, FIN 37];

3) a third class is that of inverse probabilities, according to Bayes and Laplace. It consists of the final probability of a hypothesis (after experiments have been conducted) estimated from the *a priori* probability (in the absence of experiments) and the conditional probability (or likelihood probability), or experiment probability given the hypothesis;

4) the physical probabilities used in the 19th century are no longer subjective in nature and attempt on the contrary to achieve objectivity by calculating probabilities that are conditional to experiments that have been conducted;

5) the purely frequentist approach calculates occurrence frequencies in large sequences (Venn's limit, Fisher's infinite population, etc.);

6) finally, the last class, which Good calls "neo-classic" subjective probability, is the largest. Probabilities represent degrees of confidence, with respect to a state of knowledge, taking into account both subjective and objective information. This definition encompasses all of the others and may be much more general. It relies on a mathematical theory based on a few axioms, which makes it possible to ensure the consistency of all the degrees of confidence. It can be extended to a theory of rational behavior by including utilities. Finally, Good even suggests representing these subjective probabilities with inequalities, which is similar to the Dempster-Shafer belief theory, for example.

The opposition between objectivists and subjectivists actually stems from fundamental differences in the types of problems they try to solve and in their models. Objectivist frequentists search for frequencies in a set, which implies the possibility of infinite repetitions in similar conditions, but also provides an operational means of calculation. The probability is specific to the set and does not exist without it, but the data can be hypothetical (it is not always necessary to conduct all of the repetitions). This leads objectivists to refuse problems that are devoid of meaning, such as events that only occur once. Statements are considered objective if they can be refuted (with counter-examples), even if they cannot be rigorously proven [MAT 78]. On the contrary, subjectivists consider probabilities as measures of confidence, of reasonable expectation [COX 46], of the numerical coding of a state of knowledge [DEM 93], of an appropriate mental subtlety and can therefore deal with problems for which there is no set, particularly unique phenomena. For such phenomena, there is no probability *per se*, but only probabilistic models [MAT 78]. The hypotheses are evaluated according to the observed data and the prior probability, even if the knowledge is incomplete. Subjectivists do not try to achieve the best asymptotic behavior, as statisticians do, but try instead to make the best possible inference given the available data [DEM 93, KEM 42]. In other words, frequentists deal with random probabilities and subjectivists deal with epistemological probabilities [SHA 78, SHA 86]. The former are specific to the event itself and are not modified when knowledge changes [KEM 42]. The latter, on the other hand, are always conditional and change according to knowledge. They enable possible conclusions to be drawn, between certainty and impossibility, and therefore, constitute an extended logic [KEY 29]. Subjectivists reject the principle according to which the same causes produce the same effects, not because they consider it to be false, but because it has no meaning, since the causes are never identical. Oddly enough, objectivity was introduced in order to eliminate the arbitrary and subjective nature of Bayes and Laplace, but it required the use of statistical criteria that are not universal and which have to be chosen somewhat arbitrarily [DEM 93].

Finally, the last difference between the two methods, both mathematical and in meaning, is fundamental because it involves additivity. Random probabilities are necessarily additive, since they are related to the frequentist aspect, while epistemological probabilities do not have to be, although there is still controversy over this. We will discuss this further in the following sections.

To sum up, there are three types of people concerned with probabilities: mathematicians who suggest models without worrying about whether they fit reality or how they will be used, physicists who infer laws from observations and experiments, and philosophers who wonder about the meaning of all this.

A.3. Fundamental postulates for an inductive logic

Rather than accepting the "axioms" of probabilities such as they are presented, for example, in Kolmogorov's traditional approach, more subjectivist methods start off with intuitive postulates, which are directly related to what is expected of an inductive logic[3], from where they infer probability rules. This method was devised by Cox [COX 46], essentially, and was discussed in detail, for example, in [DEM 93, PAR 95, TRI 72], where demonstrations of the major results can also be found. Here, we will present these fundamental postulates and the outline of the reasoning. At the end of this section, we will present the works led by de Finetti [FIN 37]. Less known to signal and images processors, they are appealing in two ways, both for their fundamentally subjectivist aspect and for the simplicity of the demonstration.

A.3.1. *Fundamental postulates*

Here are the fundamental postulates as laid out by Cox [TRI 72] (those suggested by Jeffreys [JEF 61] are very similar[4]):

1) *consistency or non-contradiction*: if a conclusion can be drawn in different ways, they must all lead to the same result; there have to be no contradictory conclusions based on the same data; furthermore, equal confidences have to be attributed to propositions that have the same truth value;

2) *continuity of the method*: the operations performed have to be continuous and if a slight change occurs in the data, it must not lead to sudden changes in the result;

3) *universality or completeness*: it has to be possible to attribute a degree of confidence to any well-defined propositions and to compare degrees of confidence;

4) *unequivocal statements*: propositions have to be well defined, i.e. it has to be theoretically possible to determine whether a proposition is true or false. This is equivalent to what Horvitz refers to as clarity [HOR 86];

5) *no information is refused*: conclusions cannot be drawn based on partial information, meaning that all the information, experience or knowledge available related to the proposition we wish to evaluate has to be taken into account and, in particular, it is important to take into account the dependence of the context. This postulate is

---

3. The objective of inductive logic is to determine the most likely solution given the information available, true and false being the extreme cases, as opposed to deductive logic, for which the only possible cases are true, false, and total lack of knowledge.

4. Jeffreys, by trying to define a general inference method, laid out the following postulates: all of the hypotheses have to be expressed, and the conclusions are inferred from the hypotheses; the theory has to be consistent and not contradictory; every rule has to be applicable in practice; the theory has to provide indicators for pointing out possible false inferences; the theory should not systematically reject empirical information. Additionally, Jeffreys suggests relying on the following guidelines: the number of postulates has to be kept to a minimum; the theory has to be in agreement with human reasoning; because induction is more complex than deduction, we cannot hope to develop it further than deduction. Jeffreys's approach then consists of translating these postulates in more formal axioms, of introducing numbers to represent probabilities and finally of demonstrating the traditional results [JEF 61].

a response to traditional probability theories, where, in order to achieve objectivity, certain types of information have to be dismissed.

Postulates 2 and 3 lead to the use of real numbers for representing and comparing degrees of confidence: a single real number is necessary and sufficient for representing a degree of confidence and the switch from true to false is continuous.

Postulate 1 leads to the existence of functional relations between the degrees of confidence.

Postulate 4 imposes that traditional, deductive symbolic logic is a specific case.

Postulate 5 leads to hypothetical conditioning: the degree of confidence in a proposition A is only know conditionally to a state of knowledge e that represents information related to the confidence in A and assumed (or believed) to be true. Such a degree of confidence is denoted by $[A \mid e]$.

The consistency postulate and hypothetical conditioning impose the existence of a functional equation T relating $[AB \mid e]$ (degree of confidence in "A and B" for the state of knowledge e) and at least two of the quantities $[A \mid e]$, $[A \mid Be]$, $[B \mid e]$, $[B \mid Ae]$, and the existence of a functional relation S between the degrees of confidence in a proposition $[A \mid e]$ and in its negation $[\bar{A}|e]$.

Paris [PAR 95], in a rigorous demonstration of the works of Cox, insists on a hypothesis that is often omitted but is crucial to the demonstration:

$$\forall(\alpha,\beta,\gamma) \in [0,1]^3,\ \forall\varepsilon > 0,\ \exists A,B,C,D,\ \text{verifying the consistency postulate} /$$
$$\big|[D \mid ABC] - \alpha\big| < \varepsilon,\ \big|[C \mid AB] - \beta\big| < \varepsilon,\ \big|[B \mid A] - \gamma\big| < \varepsilon.$$

In particular, this hypothesis cannot be verified in a finite frame of reference.

A.3.2. *First functional equation*

For the relation T, 11 functions are possible (6 with 2 arguments, 4 with 3 arguments and 1 with 4 arguments). Because the roles of A and B are symmetric, these functions can be reduced to 7. With the rest of the arguments, certain types of functions can be eliminated by examining specific cases that lead to absurdities.

If the state of knowledge e stipulates that "A and B are independent", then $[A \mid e] = [A \mid Be]$. Among the 7 forms of T, the one that is a function of $[A \mid e]$ and $[A \mid Be]$ is then a function of $[A \mid e]$ only and no longer depends on B. This form must therefore be eliminated.

If e = "$A = \bar{B}$", then $[AB \mid e] = [B \mid Ae] = i$ (where i represents the degree of confidence assigned to impossible propositions). This constant value eliminates the form of T that depends on $[A \mid e]$ and $[B \mid e]$.

If we now examine the case "A is impossible", then $[AB \mid e] = [A \mid e] = [A \mid Be]$ $= i$ and $[B \mid Ae]$ is undefined. This case allows us to eliminate the 4 forms of T that are functions, respectively, of $[A \mid e]$ and $[B \mid Ae]$, of $[A \mid e]$, $[A \mid Be]$ and $[B \mid Ae]$, of $[B \mid e]$, $[B \mid Ae]$ and $[A \mid e]$, and of $[A \mid e]$, $[A \mid Be]$, $[B \mid e]$ and $[B \mid Ae]$.

Therefore, the only possible form is:

$$[AB \mid e] = T\big([A \mid Be], [B \mid e]\big) = T\big([B \mid Ae], [A \mid e]\big), \qquad \text{[A.1]}$$

in which A and B have interchangeable roles and where the continuity postulate imposes that T is a continuous function.

Traditional deductive logic imposes that for three propositions A, B and C, we have $(AB)C = A(BC)$. By applying this rule, we infer that T must be associative. The general solution to this functional equation is a product:

$$Kf\big[T\big([A \mid Be], [B \mid e]\big)\big] = f\big([A \mid Be]\big)f\big([B \mid e]\big), \qquad \text{[A.2]}$$

where K is a constant that can be chosen as equal to 1 for convenience and f is a monotonic function. The original demonstration of this result [COX 46] assumes that T is twice differentiable. However, Aczél's results on functional equations can be used to reduce these hypotheses [ACZ 48, ACZ 66]: T only has to be associative, continuous and strictly increasing with respect to each of the arguments; differentiability is not required[5].

If we now assume that "$A = B$", then $[A \mid Be] = c$ (where c is the degree of confidence assigned to a proposition that is certain). Therefore, we have $f(c) = 1$. In a similar fashion, by assuming that "$A = \bar{B}$", we find that $f(i)$ must be equal to 0 or to $+\infty$. By convention, we choose $f(i) = 0$. Therefore, the function f is positive and increasing from 0 to 1.

A.3.3. *Second functional equation*

We now examine the functional relation S between $[A \mid e]$ and $[\bar{A}|e]$. By applying S twice, we get $S^2 = Id$. Consistency with the first functional equation implies that

5. Dubois and Prade have shown that by accepting that T is simply non-decreasing, we can choose $T = \min$.

S must satisfy the equation:

$$yS\left[\frac{S(x)}{y}\right] = xS\left[\frac{S(y)}{x}\right] \quad \text{[A.3]}$$

the general solution of which is:

$$f\big([A \mid e]\big)^k + f\big([\bar{A} \mid e]\big)^k = 1 \quad \text{[A.4]}$$

by assuming that S is twice differentiable.

A.3.4. *Probabilities inferred from functional equations*

We then set by convention $p(A|e) = f([A \mid e])^k$, which is referred to as the probability of A conditionally to e. The two functional equations then become:

$$p(AB \mid e) = p(A \mid Be)p(B \mid e), \quad \text{[A.5]}$$

$$p(A \mid e) + p(\bar{A} \mid e) = 1. \quad \text{[A.6]}$$

We have thus demonstrated the relations imposed axiomatically in the traditional approach (Kolmogorov). Furthermore, we are dealing from the start with conditional probabilities (related to a state of knowledge), whereas these probabilities were only introduced later in the traditional theory. Finally, we infer the relation that leads to the probability of the union:

$$p(A + B \mid e) = p(A \mid e) + p(B \mid e) - p(AB \mid e), \quad \text{[A.7]}$$

and therefore the additivity of the probabilities of exclusive events (usually imposed axiomatically in the traditional theory). We also infer Bayes' rule:

$$p(A \mid Be) = \frac{p(B \mid Ae)p(A \mid e)}{p(B \mid e)}. \quad \text{[A.8]}$$

Note that this approach leads to subjective probabilities that are additive, in contradiction with the general conception in the 17$^{\text{th}}$ century and with that of one of the schools of thought in the 20$^{\text{th}}$ century.

A.3.5. *Measure of uncertainty and information theory*

In an approach similar to that of Cox, Jaynes defined a series of criteria in order to obtain a measure of uncertainty [JAY 57]. In his works, he tried to bring together mechanics, statistics and information theory. He expressed his research as a problem involving the specification of probabilities when little information is available. He examined the two objectivist and subjectivist methods, and chose the latter. By representing a state of knowledge, this makes it possible to express possible conclusions if not enough information is available to obtain conclusions that are certain. It is therefore more general, and Jaynes adopted it for statistical mechanics.

The intuitive criteria that Jaynes expects from a measure of uncertainty are the following:

1) the measure has to be positive and continuous;

2) it has to increase when the uncertainty increases;

3) it has to be additive if the sources are independent.

He thus obtains a unique measure of uncertainty represented by a discrete probability distribution that corresponds to these intuitive criteria:

$$H\left(p_1, \ldots, p_n\right) = -K \sum_{i=1}^{n} p_i \log p_i. \qquad \text{[A.9]}$$

In the same way Cox rediscovered probabilistic relations from his postulates, Jaynes thus rediscovered, based on his criteria, the entropy in statistical mechanics and simultaneously Shannon's entropy [SHA 59].

The maximum entropy principle can then be considered analogous to Laplace's principle of insufficient reason. The essential difference is that Laplace's principle is arbitrary in nature and can lead to paradoxes (the concept of equally probable cases changes if the variable is changed), whereas the maximum entropy principle makes it possible to make inferences on the basis of partial information in an unbiased fashion. It can be chosen for the good reason that entropy is determined in a unique way as the value that "implicates itself the least" with respect to the missing information and not for the negative reason that there is no reason to think otherwise. However, it can be criticized because the results it provides depend on how the problem is stated. This criticism also applies to the principle of insufficient reason as we have seen in section 6.10.

A.3.6. *De Finetti and betting theory*

The approach suggested by de Finetti, which predates that of Cox, also relies on simple and intuitive axioms that lead to the properties of probabilities [FIN 37]. We have of course the same axioms of increasingness and universal comparison, and most importantly a consistency axiom which serves as the basis of the demonstration.

De Finetti developed a betting theory to explain his argument: the probability p attributed by an individual to an event E is given by the conditions on which this individual would be willing to bet on this event, i.e. on which he would wager the sum pS in order to win S if the event E occurred. Based on this definition, de Finetti first showed that the sum of the probabilities of incompatible events has to be equal to 1. Let $\{E_1, \dots E_n\}$ be a complete class of incompatible events, let p_i be their probabilities (always assessed by an individual) and let S_i be the stakes that correspond to each of them. If the event E_k occurs, the gain G_k is defined as the difference between the corresponding stake S_k and the sum of the wagers, meaning that:

$$G_k = S_k - \sum_{i=1}^{n} p_i S_i. \qquad \text{[A.10]}$$

We obtain n equations of this type, corresponding to the n possible outcomes. If we consider these equations as a system of n equations with n unknowns, i.e. the S_i, the determinant of this system is equal to:

$$D = \begin{bmatrix} 1-p_1 & -p_2 & \cdots & -p_n \\ -p_1 & 1-p_2 & \cdots & -p_n \\ \cdots & \cdots & \cdots & \cdots \\ -p_1 & -p_2 & \cdots & 1-p_n \end{bmatrix} = 1 - \big(p_1 + p_2 + \cdots + p_n\big). \qquad \text{[A.11]}$$

If the determinant is not equal to zero, the system has a solution for any G_k, even if every one of them is positive. This would not be consistent with the concept of betting. It is difficult to conceive of a game that could always be won or in which it is possible to give the opponent the possibility of certainly winning! Therefore, the only consistent solution is that obtained when the determinant is equal to zero, i.e. when:

$$\sum_{i=1}^{n} p_i = 1. \qquad \text{[A.12]}$$

Furthermore, this condition is sufficient, since we then have:

$$\sum_{i=1}^{n} p_i G_i = 0, \qquad \text{[A.13]}$$

and not all of the gains are positive.

De Finetti interprets the result this way: each evaluation (which is subjective) of the p_i such that $\sum_{i=1}^{n} p_i = 1$ is an acceptable evaluation, in other words one that corresponds to a consistent opinion. The choice of an evaluation among those that are acceptable is then no longer the objective at all.

From equation [A.12], we get the additivity of the probabilities of disjoint events.

The second step of the argument focuses on determining conditional probabilities. To do this, we will consider three bets:

1) a bet on E_1E_2 (E_1 and E_2), with the stake S_1, and the wager p_1S_1;

2) a bet on E_2, with the stake S_2, and the wager p_2S_2;

3) a bet on $E_1 \mid E_2$, with the stake S, and the wager pS, the gain for this bet being:
- $(1-p)S$ if $E_1 \mid E_2$ is true,
- $(-p)S$ if $E_1 \mid E_2$ is false,
- 0 if E_2 does not occur (the game is considered void in this case, and the wager is paid back).

Three outcomes are possible:

1) if E_1 and E_2 occur, then the gain is:

$$G_1 = \left(1-p_1\right)S_1 + \left(1-p_2\right)S_2 + (1-p)S; \qquad \text{[A.14]}$$

2) if E_2 occurs but not E_1, then the gain is:

$$G_2 = -p_1S_1 + \left(1-p_2\right)S_2 - pS; \qquad \text{[A.15]}$$

3) if E_2 does not occur, then the gain is:

$$G_3 = -p_1S_1 - p_2S_2. \qquad \text{[A.16]}$$

Consider equations A.14, A.15 and A.16 as a system of 3 equations with 3 unknowns, as before. The determinant is equal to:

$$p_1 - pp_2 \qquad \text{[A.17]}$$

and, again for reasons of consistency, has to be equal to 0. We then get the relation:

$$p\left(E_1 \mid E_2\right) = \frac{p\left(E_1E_2\right)}{p\left(E_2\right)} \qquad \text{[A.18]}$$

from which we infer Bayes' theorem.

Note that in this case, the expected gain is equal to:

$$p_1G_1 + \left(p_2-p_1\right)G_2 + \left(1-p_3\right)G_3 = \left(p_1 - p_2p\right)S = 0. \qquad \text{[A.19]}$$

Therefore, the expected gain is zero for any S, S_1, S_2.

De Finetti adopts a subjectivist philosophy, in the sense that he considers subjectivist elements, far from having to be eliminated as suggested by objectivists in order to render the concept of probability more "scientific", are essential and inherent to the concept of probability. This coincides with the point of view according to which probability expresses an individual's opinion, and only has significance with regard to that individual, as opposed to the objectivist perspective which considers that probability exists independently of individuals and is a property of the physical world.

A.4. Bibliography

[ACZ 48] ACZÉL J., "Sur les opérations définies pour nombres réels", *Bull. Soc. Math. Franç.*, vol. 76, p. 59-64, 1948.

[ACZ 66] ACZÉL J., *Lectures on Functional Equations and Their Applications*, Academic Press, New York, 1966.

[BLO 96] BLOCH I., "Incertitude, imprécision et additivité en fusion de données: point de vue historique", *Traitement du Signal*, vol. 13, no. 4, p. 267-288, 1996.

[COX 46] COX R.T., "Probability, Frequency and Reasonable Expectation", *Journal of Physics*, vol. 14, no. 1, p. 115-137, 1946.

[DEM 93] DEMOMENT G., Probabilités, modélisation des incertitudes, inférence logique, et traitement des données expérimentales, Report, Paris-Sud University course, Orsay, France, 1993.

[DUB 88] DUBOIS D., PRADE H., *Possibility Theory*, Plenum Press, New York, 1988.

[FEL 66] FELLER W., *An Introduction to Probability Theory and its Applications*, Wiley, New York, 1966.

[FIN 37] DE FINETTI B., "La prévision: ses lois logiques, ses sources subjectives", *Annales de l'Institut Henri Poincaré*, vol. 7, no. 1, p. 1-68, 1937.

[GOO 59] GOOD I.J., "Kinds of Probability", *Science*, vol. 129, no. 3347, p. 443-447, 1959.

[HOL 62] HOLLAND J.D., "The Reverend Thomas Bayes, F.R.S. (1702-61)", *J. Roy. Stat. Soc. (A)*, vol. 125, p. 451-461, 1962.

[HOR 86] HORVITZ E.J., HECKERMAN D.E., LANGLOTZ C.P., "A Framework for Comparing Alternative Formalisms for Plausible Reasoning", *National Conference on Artificial Intelligence*, p. 210-214, 1986.

[JAY 57] JAYNES E.T., "Information Theory and Statistical Mechanics", *Physical Review*, vol. 106, no. 4, p. 620-630, 1957.

[JEF 61] JEFFREYS R., *Theory of Probability*, Oxford University Press, 1961.

[KEM 42] KEMBLE E.C., "Is the Frequency Theory of Probability Adequate for All Scientific Purposes?", *Am. J. Physics*, vol. 10, p. 6-16, 1942.

[KEY 29] KEYNES J.M., *A Treatise on Probability*, Macmillan, London, 1929.

[MAT 78] MATHERON G., Estimer et choisir - Essai sur la pratique des probabilitś, Report, Ecole Nationale Supérieure des Mines de Paris, Geostatistics Center, Fontainebleau, France, 1978.

[NEA 92] NEAPOLITAN R.E., "A Survey of Uncertain and Approximate Inference", in L. ZADEH and J. KAPRZYK (ed.) *Fuzzy Logic for the Management of Uncertainty*, p. 55-82, J. Wiley, New York, 1992.

[PAR 95] PARIS J.B., *The Uncertain Reasoner's Companion, a Mathematical Perspective*, Cambridge University Press, 1995.

[SHA 59] SHANNON C.E., WEAVER W., *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, USA, 1959.

[SHA 76] SHAFER G., *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[SHA 78] SHAFER G., "Non-Additive Probabilities in the Work of Bernoulli and Lambert", *Archive for History of Exact Sciences*, vol. 19, p. 309-370, 1978.

[SHA 86] SHAFER G., "The Combination of Evidence", *International Journal of Intelligent Systems*, vol. 1, p. 155-179, 1986.

[SHO 75] SHORTLIFFE E.H., BUCHANAN B.G., "A Model of Inexact Reasoning in Medicine", *Mathematical Biosciences*, vol. 23, p. 351-379, 1975.

[SOM 89] SOMBÉ L., *Raisonnements sur des informations incomplètes en intelligence artificielle*, Teknea, Marseille, 1989.

[STI 82] STIGLER S.M., "Thomas Bayes's Bayesian Inference", *J. Roy. Stat. (A)*, vol. 145, p. 250-258, 1982.

[STI 83] STIGLER S.M., "Who Discovered Bayes's Theorem?", *The American Statistician*, vol. 37, no. 4, p. 290-296, 1983.

[TRI 72] TRIBUS M., *Rational, Decriptions, Decisions and Designs*, Pergamon Press Inc., 1972.

[ZAD 65] ZADEH L.A., "Fuzzy Sets", *Information and Control*, vol. 8, p. 338-353, 1965.

[ZAD 78] ZADEH L.A., "Fuzzy Sets as a Basis for a Theory of Possibility", *Fuzzy Sets and Systems*, vol. 1, p. 3-28, 1978.

This page intentionally left blank

Appendix B

Axiomatic Inference of the Dempster-Shafer Combination Rule

The Dempster-Shafer belief theory has often been criticized for imposing an *ad hoc* combination rule (Dempster's orthogonal rule [SHA 76], equation [7.26]), with no theoretical justification for it, although it does have a perfectly satisfactory intuitive interpretation, which is in agreement with the concept of the conjunction of focal elements.

This appendix completes the theory presented in Chapter 7 by giving a theoretical justification of the conjunctive combination rule (non-normalized, as it is suggested in [SME 90]).

We will also explain how it is related to the Cox approach to probabilities, which is presented in Appendix A. In particular, we will underline the differences between the axioms, which explain the difference between the models and the combination modes in the two theories.

Several recent works have attempted to justify this rule, for example, those of Dubois and Prade [DUB 86], which mathematically justify the use of the product to combine masses based on the concept of the separability of sources, or those of Smets based on the transferable belief model [SME 90].

Chapter written by Isabelle BLOCH.

We should also mention the works of Gacôgne, which have led to a justification, in the specific cases where the frame of discernment is reduced to two elements, based on the concept of accentuation [GAC 93][1].

The works of Smets offer the most general justification, as far as we know, and his arguments will be described here. Furthermore, his method is similar to that of Cox (see Appendix A). Work similar to that of Smets has been done by Klawonn and Schwecke [KLA 92].

B.1. Smets's axioms

The first observation Smets made involves the indifference principle (or principle of insufficient reason). Assigning the same probability to every simple event implies that different probabilities will be assigned to union of events, which, to Smets, does not correspond to indifference. This should instead be expressed by the existence of a constant that is positive or equal to zero and such that:

$$\forall A \subset D,\ A \neq D,\ \mathrm{Bel}(A) = c, \qquad \text{[B.1]}$$

where D refers to the frame of discernment. This is obviously impossible with probabilities, but it is not in the Dempster-Shafer context with credibilities, since we have:

$$A \cap B = \emptyset \Longrightarrow \mathrm{Bel}(A \cup B) \geq \mathrm{Bel}(A) + \mathrm{Bel}(B) \qquad \text{[B.2]}$$

hence $c \geq 2c$ and therefore $c = 0$. The mass function representing indifference (or complete lack of knowledge) is therefore defined by:

$$m(D) = 1 \text{ and } \forall A \neq D,\ m(A) = 0, \qquad \text{[B.3]}$$

1. Gacôgne shows that the Dempster-Shafer rule (equation [7.26]), in the case where the frame of discernment is reduced to a proposition P and its opposite $\bar{P}$, can be inferred from the concept of accentuation. An accentuation function is such that it reduces the degrees of confidence smaller than 0.5 and increases those greater than 0.5, thus making them more like binary degrees (this is equivalent to the concept of reinforcement found in the algebraic theory of ordered semigroups as well as in fuzzy set theory). The rational function of $[0, 1]$ in $[0, 1]$ with the lowest degree that is an accentuation function is defined by $x^2/(2x^2 - 2x + 1)$, and that is the one that can be used to show the analogy with Dempster-Shafer. This concept is then generalized to pairs (x, y), characterizing a proposition P, such that $0 \leq x \leq y \leq 1$ (referred to as "obligation" and "eventuality", hence similar to the concept of credibility and plausibility or of necessity and possibility). The accentuation of such a couple is defined by: $\#(x, y) = ((2xy - x^2)/(1 - 2x + 2xy), y^2/(1 - 2x + 2xy))$. These two values correspond exactly to those we would obtain by combining, using the Dempster-Shafer rule (equation [7.26]), the credibility (respectively, the plausibility) of a proposition P with itself. If we now have at our disposal two games with the measures (x, y) and (x', y') on P, the same type of reasoning leads to the justification of the Dempster-Shafer rule [GAC 93]. We would then be left with having to generalize this approach to more complex spaces of discernment.

which, this time, is perfectly satisfactory. This particular mass function plays an important part in the combination rule since it is its identity element, which confirms its interpretation in terms of complete lack of knowledge, which cannot modify another mass function.

Smets's second idea is the transferable belief model, which defines conditioning. The problem is stated as follows: given a new information, which makes it possible to state that the truth is located in a subset B of the frame of discernment D, how can a set of masses m be modified to take into account this new information? The expression suggested by Smets is as follows:

$$\begin{aligned} m'(A) &= \sum_{X \subseteq \bar{B}} m(A \cup X) \; \forall A \subseteq B \\ &= 0 \text{ otherwise} \end{aligned} \qquad [B.4]$$

where m' refers to the new set of masses. These formulae can be modified if needed, although this is not necessary and can even be detrimental to the extent that it masks conflict [SME 90], as we saw in Chapter 7. Normalization also poses continuity problems in the vicinity of the total conflict [DUB 86].

This formula is interpreted as follows. If we decompose a subset A into the union $A_1 \cup A_2$ with $A_1 \subseteq B$ and $A_2 \subseteq \bar{B}$, the mass $m(A_1 \cup A_2)$ is entirely transferred to A_1 (hence the model's name). In specific cases where $A_2 = \emptyset$ ($A \subseteq B$), the mass of A is not modified and if $A_1 = \emptyset$ ($A \subseteq \bar{B}$), the mass of A becomes zero.

In Shafer's theory [SHA 76] and in our presentation of it in Chapter 7, the conditioning formula is inferred from the combination rule, whereas here, it precedes it and is constructed simply by logical considerations. Note that the conditioning rule on plausibilities:

$$\text{Pls}(A \mid B) = \frac{\text{Pls}(A \cap B)}{\text{Pls}(B)}$$

can itself be justified by supposing that:

$$\text{Pls}(A \cap B) = T\big[\,\text{Pls}(A \mid B), \text{Pls}(B)\big]$$

and by applying a method similar to the one that lead us to the first functional equation (section A.3.2) [DUB 86].

In a third step, Smets defines two axioms that he wishes to see satisfied by the combination rule, denoted by $\oplus$:

A1: $(\mathrm{Bel}_1 \oplus \mathrm{Bel}_2)(A)$ has to be a function of only the functions m_1 and m_2 and of A.[2]

A2: $\oplus$ has to be commutative.

A3: $\oplus$ has to be associative.

A4: If $m_2(B) = 1$, then $m_1 \oplus m_2$ has to satisfy the conditioning law, meaning that:

$$\begin{aligned}\big(m_1 \oplus m_2\big)(A) &= \sum_{X \subseteq \bar{B}} m_1(A \cup X)\ \forall A \subseteq B \\ &= 0 \text{ otherwise.}\end{aligned} \qquad \text{[B.5]}$$

A5: The law has to satisfy an internal symmetry property (invariance under permutation of the simple hypotheses).

A6: For $A \neq D$, $(m_1 \oplus m_2)(A)$ does not depend on $m_1(X)$ for $X \subseteq \bar{A}$ (auto-functionality property).

A7: There are at least 3 elements in D.

A8: The law has to satisfy a continuity property:

$$\begin{aligned}&m_2(A) = 1 - \varepsilon,\ m_2(D) = \varepsilon,\ m_A(A) = 1 \\ &\quad\Longrightarrow \forall X,\ \lim_{\varepsilon \to 0} \big(m_1 \oplus m_2\big)(X) = \big(m_1 \oplus m_A\big)(X),\end{aligned} \qquad \text{[B.6]}$$

with m_1 being a function of any mass; this property makes it possible to eliminate degenerate cases.

B.2. Inference of the combination rule

Based on the previous axioms, Smets obtains the only possible combination rule that satisfies these axioms. To do this, he relies on communality functions defined by:

$$\forall A \subseteq D,\ q(A) = \sum_{A \subseteq X,\ X \subseteq D} m(X). \qquad \text{[B.7]}$$

2. In [KLA 92], the axioms are for the most part similar to those used by Smets. The major difference lies in the use of relations between spaces of discernment that are included in one another rather than of dependence relations as used by Smets.

His demonstration, which we will not give in detail, relies on the properties of triangular norms and of absolutely monotonic functions and is divided into three parts. First of all, axioms *A1* to *A4* lead to the existence of a function f such that the result of the combination depends only on A and on the communalities of the subsets included in A:

$$\left(q_1 \oplus q_2\right)(A) = f\left[A, \left\{q_1(X);\ X \subseteq A,\ q_2(X);\ X \subseteq A\right\}\right]. \qquad \text{[B.8]}$$

Then, by adding axioms *A5* and *A6*, it is possible to specify the form of f, which now only depends on A and on $q_1(A)$ and $q_2(A)$:

$$\left(q_1 \oplus q_2\right)(A) = f\left[A, q_1(A), q_2(A)\right]. \qquad \text{[B.9]}$$

Finally, the set of axioms *A1* to *A8* allows us to determine the final form of the combination rule:

$$\left(q_1 \oplus q_2\right)(A) = q_1(A)q_2(A). \qquad \text{[B.10]}$$

As expected, we get the Dempster-Shafer rule on communalities and this leads us to the combination of mass or credibility functions.

The advantage of Smets's method is that the axioms it relies on have interpretations that are close to what our intuition tells us. It is also easier to refute or modify them if they do not correspond to the problem at hand.

B.3. Relation with Cox's postulates

In this section, we will try to establish the links between Cox's postulates (see Appendix A) and Smet's axioms, in order to show why they lead to different theories.

First of all, we should specify which framework we chose for this comparison. The works of Cox and those of Smets do not deal with exactly the same problems, since Cox attempted to justify probabilities and their properties, whereas Smets tried to justify a combination rule. However, it is interesting to note certain analogies between the two sets of axioms. Furthermore, Cox's axioms make it possible to infer Bayes' rule (equation [A.8]), which is used in signal and image processing to fuse information by using conditional probabilities (see Chapter 6). We chose the data fusion point of view for this comparison. It would also be interesting to compare Cox's axioms with those introduced by Smets in order to justify credibility and plausibility functions [SME 93], but this comparison would only involve the modeling phases of the fusion process and not the combination phases themselves.

Axiom *A1*, which expresses the dependence between degrees of confidence and their combinations, is not as strict as Cox's postulate. Indeed, the consistency postulate implies the existence of a relation defining the degree of confidence in AB which involves only the propositions A and B, in the form of degrees of confidence assigned to $[A \mid B]$ and $[B]$ (or $[B \mid A]$ and $[A]$) but not to other propositions. Smets's axiom, which is more general, corresponds to the possibility provided by the Dempster-Shafer theory to deal with subsets and not simply with singletons.

Axioms *A2*, *A3* and *A5* correspond to properties of classical propositional logic. Cox's postulates (particularly postulate 4) also imply that deductive logic exists as a specific case. Therefore, the two methods coincide with each other on this point. These axioms are used in Cox's method to eliminate certain forms of functional relations between $[AB \mid e]$ and the other degrees of confidence, in order to keep only the form that is consistent with deductive logic:

$$[AB \mid e] = T\big([A \mid Be], [B \mid e]\big) = T\big([B \mid Ae], [A \mid e]\big). \qquad \text{[B.11]}$$

Likewise, these axioms are used in Smets's demonstration to eliminate dependences and prove that $(q_1 \oplus q_2)(A)$ at first only depends on A, and on $q_1(X)$ and $q_2(X)$ for $X \subseteq A$; then, in a second phase, only on $q_1(A)$ and $q_2(A)$.

Axiom *A4* (conditioning) expresses an idea that is very similar to the hypothetical conditioning obtained from Cox's fifth postulate. The main difference is that conditioning, this time, is expressed more as a compatibility relation than as a conditional probability.

There is no equivalent to Cox's postulate 3 (universality) in Smets's axioms. This is justified by the very basis of belief theory, in which propositions are characterized by two numbers (credibility and plausibility) instead of just one, and in which well-defined propositions are allowed not to have a degree of confidence assigned to them[3]. This flexibility is helpful for solving problems related to lack of information: if a source is not capable of providing information about A, but provides some, for example, about $A \cup B$, this situation is naturally taken into account by belief function theory by assigning a mass to $A \cup B$ and not to A, whereas it would often require including hypotheses or models in probability theory in order to be able to assign a degree of confidence to A. From the perspective of comparing degrees of confidence,

3. This can be done, for example, by assigning a zero mass to this proposition A. This does not mean, however, that a zero confidence is attributed to A, since the credibility $\mathrm{Bel}(A)$ and the plausibility $\mathrm{Pls}(A)$ are not necessarily equal to zero because non-zero masses can be assigned to propositions B such that $A \cap B \neq \emptyset$. This simply means that no degree of confidence is assigned specifically to A.

this can be done on two levels in belief theory, either on credibilities, or on plausibilities, thus leading to conclusions that are not necessarily equivalent.

Smets's axiom *A6* does not imply that A and $\bar{A}$ are interchangeable, whereas this property is explicitly used by Cox to obtain the second functional equation (equation [A.6]), since subsets X can be involved in both $(m_1 \oplus m_2)(A)$ and $(m_1 \oplus m_2)(\bar{A})$. Therefore, no complementarity relation regarding m can be obtained from it. This is replaced by a duality relation between Bel and Pls.

Finally, axioms *A7* and *A8* are considered by Smets himself as technical axioms used in the demonstrations. The regularity imposed on functions can be compared with the regularity hypotheses formulated for Cox's two functional equations [A.5] and [A.6].

These differences between the two theories have consequences on the three levels that traditionally comprise the fusion process, i.e. the modeling of belief functions, the combination of the functions determined from the information provided by several sources and the final decision:

– first in the modeling phase, because this phase is strongly constrained by the two functional relations (equations [A.5] and [A.6]) in probabilistic fusion, whereas belief theory makes it possible to easily adapt to many situations (we mentioned the example of sensors that only provide information regarding the union of two classes, without distinguishing them);

– in the combination of belief functions, postulates impose Bayes' rule on the one hand, Dempster's rule on the other hand, and their differences stem in particular from the more flexible constraints imposed by Smets's conditioning rather than from Cox's hypothetical conditioning;

– finally, in the decision making, i.e. the ultimate phase of the fusion process, differences come mostly from comparing degrees of confidence, which give way to several types of decision in the Dempster-Shafer theory.

B.4. Bibliography

[DUB 86] DUBOIS D., PRADE H., "On the Unicity of Dempster Rule of Combination", *International Journal of Intelligent Systems*, vol. 1, p. 133-142, 1986.

[GAC 93] GACÔGNE L., About a Foundation of Dempster's Rule, Report, Laforia 93/27, 1993.

[KLA 92] KLAWONNN F., SCHWECKE E., "On the Axiomatic Justification of Dempster's Rule of Combination", *International Journal of Intelligent Systems*, vol. 7, p. 469-478, 1992.

[SHA 76] SHAFER G., *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[SME 90] SMETS P., "The Combination of Evidence in the Transferable Belief Model", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, p. 447-458, 1990.

[SME 93] SMETS P., "Quantifying Beliefs by Belief Functions: an Axiomatic Justification", *Int. Joint. Conf. on Artificial Intelligence*, Chamséry, France, p. 598-603, 1993.

List of Authors

Isabelle BLOCH
TSI
Ecole nationale supérieure des télécommunications
Paris, France

Bertrand COLLIN
DGA/DCE/CTA/GIP
Ministère de la Défense
Arcueil, France

Fabienne EALET
DGA/DCE/CTA/GIP
Ministère de la Défense
Arcueil, France

Catherine GARBAY
LIG - CNRS
Grenoble, France

Jean-Pierre LE CADRE
IRISA
CNRS
Rennes, France

Henri MAÎTRE
TSI
Ecole nationale supérieure des télécommunications
Paris, France

Vincent NIMIER
ONERA
Châtillon, France

Roger REYNARD
IEF - DIGITEO Labs
Paris-Sud University, France

Michèle ROMBAUT
GIPSA-Lab
Grenoble, France

Index