ECVision - Specific Action Contribution to CCV Ontology: Dealing with Imprecise Spatial Information in Cognitive Vision

Isabelle Bloch
Ecole Nationale Supérieure des Télécommunications
Département Traitement du Signal et des Images
CNRS URA 820
46 rue Barrault - 75013 Paris, France
E-mail : Isabelle.Bloch@enst.fr

June 10, 2003

Contents

1 Introduction

This specific action aims at extending CCV Ontology with aspects related to imprecise knowledge representation in cognitive vision, to tools and methods for dealing with imprecise spatial information, and to information fusion issues and methods.

It appears in CVOnline, that in the fusion section, there is nothing on non-probabilistic methods such as fuzzy sets and possibility theory, or belief function theory. Since these theories are important in cognitive vision when several sources of information are available, this could appear also as a part of the Reasoning section of CCV Ontology (Section 3). This can be further subdivided in general consideration on fusion for cognitive vision problems (Section 3.2: Issues), and in formal method description (Section 3.3: Methods). This includes also an important aspect of dealing with spatial information in fusion problems. Moreover, dealing with imprecision in images is a very important issue for cognitive vision, and could be addressed at the level of Knowledge representation (Section 2 of CCV Ontology), and of Reasoning (Section 3.3 in particular).

Contributions on all these aspects will significantly extend the material available in CCV Ontology. It can be very useful for students, mostly at graduate levels, as well as to young researchers involved in these areas.

In this work, we concentrate on Issues (fusion) and Methods (Bases of Fuzzy sets and possibility theories, and Fuzzy sets and possibility theory in image processing and vision), which could be included in Sections 3.2 and 3.3 respectively.

2 General definitions of information fusion for decision making

This Section could probably be included in the Reasoning part of CCV Ontology, Section Issues (3.2).

2.1 Choice of a definition

The increasing development of research in information fusion in several domains is motivated by the multiplication of sources of knowledge and data, and of techniques for their acquisition. The huge volume of data to be processed and the more and more complex problems that are addressed induce a real need in the development of fusion techniques.

In particular, data (or information) fusion appeared in the domain of image processing, computer vision and cognitive vision a few years ago, as a necessary stage for several applications like medical imaging, aerial and satellite imaging, quality control, robot vision, vehicle or robot guidance. The need for information fusion in these domains originates from the increasing number of information sources, mostly images issued from different sensors, but other types of information as well (maps, atlases, expert knowledge, cognitive information, etc.). Imprecision and fuzziness being inherent to these pieces of information to merge, information fusion methods have to take this specificity of images into account, in order to improve the decision.

We consider here the term "information" in a broad sense. In particular it covers data (for instance measurements, observations, images, etc.) and pieces of knowledge (on the data, on the domain, on constraints, etc.), which can be either generic or specific.

The definition that we will use here is the following [36, 39]: *Information fusion consists in combining pieces of information issued from several sources in order to improve the decision making step.*

This definition has the interest to focus on the combination and decision steps, which can take different forms depending on the problem at hand and on the considered application. For each type of problem and application, this definition can be made more specific, by answering a number of questions: What is the aim of fusion? How is the decision expressed? What are the pieces of information to be combined? Where are they originating from? What are their characteristics? Which methodology should be chosen? How to evaluate and validate the method and the results? Which are the main difficulties, limitations? etc.

Let us compare this definition to a few others proposed by some working groups that tried to unify terminology and concepts in the fusion community.

This definition is slightly more specific than the one proposed by the European working group FUSION [32], which worked on fusion in different domains from 1996 to 1999. Resulting from discussions in the working group, the proposed informal definition is as follows: fusion consists in conjoining or merging information that stems from several sources and exploiting that conjoined or merged information in various tasks such as answering questions, making decisions, numerical estimation, etc. In this definition, which is also focused on combination and aims, the aims are more general. Here we restrict to the amelioration of knowledge about how the world is, and do not refer to how we would like the world be, a problem addressed for instance in fusion of preferences.

A unified terminology has been proposed by the Joint Directors of Laboratories Data Fusion Working Group. This group was created in 1986 in US and tried to codify the data fusion terminology, producing a Data Fusion Lexicon [1]. The proposed model is dedicated mainly to military applications (such as target tracking, automatic target recognition and identification) and is functionally oriented, identifying processes, functions and techniques. It focuses mainly on a description of a hierarchy of processing steps in a system. Here we take another point of view which is adapted to fusion problems in several domains, including engineering (e.g. in signal and image processing for pattern recognition, diagnosis and decision making in non military applications). Instead of focusing on the system, we define data fusion as the process that combines information issued from different sources in order to take a decision, and therefore we focus on techniques for combining information towards a specific goal.

Another European working group, of EARSeL (Association of Remote Sensing Laboratories), has extended the JDL definition to the more general domain of satellite imaging [173]: data fusion constitutes a formal framework in which data issued from different sources are expressed; it aims at improving information quality; a more precise definition of quality depends on the application. This definition includes several definitions proposed in satellite imaging, summarized in [173]. The definition used here goes farther, since it includes the decision step as ultimate goal.

2.2 General characteristics of imperfect information

Fusion problems occur in particular when the information is imperfect and when several pieces of information have to be gathered or merged in order to overcome the limitations due to these imperfections.

2.2.1 Types of information

A first characteristic concerns the type of information to be combined. It can consist of direct observations, of processing results on these observations, of more or less generic pieces of knowledge, of expert opinions, etc. These pieces of information can be expressed in numerical or symbolic form. A particular attention has to be paid to the scale used for representation. It does not necessarily has to have an absolute meaning, but at least some commensurability or normalization has to be guaranteed. A more detailed description of different types of information and their representation can be found in [32].

The level of information is also an important aspect. Depending if the information is considered at low level (typically the original data), or at higher level, which often calls for preliminary processing, constraints and difficulties that arise are not the same.

Other distinctions can be underlined, since they lead to different types of modeling and processing, for instance distinction between frequent and rare data, between static and dynamic (evolving in time) information, between generic or factual information (the second one being often more specific than the first one). Typically generic information can be ignored if it contradicts a reliable factual (observed) information. On the contrary if no observation can be made, or if it is not reliable, then generic information has to be used, and is then considered as a "default".

Information involved in a fusion process consists on the one hand of pieces of information to be combined, and on the other hand of additional knowledge, which serves to guide or help the combination. It can be information about the information to be combined as information about sources, their dependencies, their reliability, about preferences, etc. It can also be contextual information, on the domain. This additional knowledge is generally not expressed in the same formalism as information to be combined.

2.2.2 Imperfection of information

One important characteristic of information in fusion is its imperfection. It is always present (otherwise fusion would not be necessary). It can take different forms, that are briefly described in this Section. Note that there is no real consensus about these definitions in the literature. The ones that are given here are certainly not universal, but are well adapted to fusion problems.

Uncertainty Uncertainty is related to the truth of some information, characterizing its adequation to reality [71]. It refers to the nature of the considered object, to its quality, or to its occurrence.

Imprecision Imprecision concerns the content of information and describes a quantitative defect of knowledge or measure [71]. It concerns the lack of precision in quantity. This notion is often included in a larger sense of uncertainty, but it may be useful to distinguish between them. More refined classes of uncertainty have also been proposed (e.g. [108]).

Incompleteness Incompleteness of information issued from each source is the main reason that motivated the fusion. Information provided by a source is generally partial, and gives only one point of view or one aspect of the observed phenomenon.

Ambiguity Ambiguity of information extracted from each source has also to be taken into account. For instance an imprecise measurement can prevent to distinguish between two different situations, or incomplete information can induce confusions between objects or situations. One of the objectives of fusion is therefore to solve ambiguities using additional sources of information and knowledge.

Conflict Conflictual situations occur often in fusion problems, and are usually difficult to solve. First, detecting conflicts is not easy. They can be confused with other types of imperfection, or even with complementarity between sources. Then their identification and typology is a question that often arises, but differently depending on the domain. At last, solving conflict can take different forms. It can rely on the elimination of non reliable sources, on introducing additional information, etc. In some cases, it is preferable to delay combination and wait for more information, or even not to fusion at all (for instance if conflict appears because the sources actually do not speak about the same phenomenon). These issues are addressed in details in [32].

2.2.3 Redundancy and complementarity

Other characteristics of information are more positive, and can be exploited in order to limit or reduce imperfections.

Redundancy Redundancy between sources comes from the fact the they give information about the same phenomenon. Ideally, redundancy should be exploited in order to reduce uncertainties and imprecisions.

Complementarity Complementarity between sources comes from the fact that they usually look at different aspects of the observed phenomenon. Ideally, complementarity should be exploited to get a more complete information and to solve ambiguities.

2.3 Numerical / symbolic

There have been large discussions in the fusion community about the duality between numerical and symbolic fusion. Our aim here is not to go deeply in this discussion but rather to present the different levels at which this question may be addressed. Since most of the discussions originate from the fact that the concepts are confusing if this level is not specified, this type of presentation may help to clarify these concepts. The three levels at which we would like to make a distinction between numerical and symbolic are (i) the type of data to be treated, (ii) the type of processing applied to these data, (iii) the role of representations. Examples are taken mainly from the image processing and computer vision domain.

2.3.1 Data and information

By numerical information, we mean data directly given as numbers. These numbers may represent various features, typically physical measures, grey levels, response to an image processing operator, etc. They may be directly read from the sources to be fused, or attached to the domain or contextual knowledge (e.g. wave lengths in satellite imaging, acquisition times in medical imaging, etc.).

By symbolic information, we mean all information given as symbols, propositions, rules, etc. Such information can be related to the data to be combined (e.g. graphical information in a map or in an anatomical atlas, attributes computed on data or objects previously extracted from the images) or related to the domain knowledge (e.g. propositions about the properties of the problem at hand, structural information stating for instance that a road network can be represented as a graph using roads and cross-roads, propositional knowledge stating general rules about the scene like “the ventricles are always inside the white matter”, etc.).

The classification of data and information in symbolic and numerical classes cannot always be done in a crisp way. We may have to deal with “hybrid” kind of information, where numbers are used for coding information that is not necessarily of numerical nature. This is typically the case for the evaluation of some data or treatment, for the quantification of imprecision or uncertainty. In such cases, the absolute values of these numbers are not important, this is rather the ranking which plays an important role. These numbers may be attached to symbolic information as well as to numerical information. In image processing, examples can be found for quantifying the quality of a detector, the evaluation of some symbolic data, of source reliability, of confidence in some measurement or numerical value, etc.

2.3.2 Processing

As far as processing of information is concerned, we mean by numerical treatment any computation on numbers. In data fusion, it concerns approaches that combine numbers by some formal calculus. Note that such kind of treatment does not make any assumption on what kind of data is represented by numbers. Data may be originally of numerical as well as of symbolic nature.

Symbolic types of treatment include formal computation on propositions (logic is but one example), possibly taking into account numerical knowledge. Structural approaches, like graph-based approaches often used in structural pattern recognition, can be considered as belonging to this class.

We consider as hybrid types of treatment the methods where prior knowledge is used in a symbolic way to control numerical treatments, for instance by stating some propositional rules that suggest/allow/prevent specific numerical operations. Typically, a proposition stating in which cases sources A and B are independent can be used in the way probabilities are combined, or knowing that the recognition depends only on some local or contextual knowledge may lead to an appropriate modeling of the scene as a Markovian field. Such kind of hybrid processing is widely used in image processing and image fusion.

2.3.3 Representations

As it appears from the two previous subsections, the representations and their type may play very different roles. Numerical representations can be used for intrinsically numerical data as well as for evaluation and quantification of symbolic data. An important use of numerical representations in data fusion is for quantifying imprecision, uncertainty or reliability of the information (this information may be of numerical as well as of symbolic nature), therefore representing rather information about the information than the data themselves. We will focus on such kind of representations in the description of the main numerical approaches for image fusion. Numerical representations are also often used for degrees of belief attached to numerical and/or symbolic knowledge, and for degrees of consistency or inconsistency in a database. Note that the same (numerical) formalism can be used for representing very different kinds of data or knowledge: the most obvious example is the use of probabilities for representing data as different as frequencies, subjective beliefs, etc. [20].

Symbolic representations can be used in logical systems, or knowledge-based systems, but also as prior knowledge for guiding numerical treatment, as a structural support for image fusion (see e.g. [123]), and of course as semantics attached to the manipulated objects.

In several examples, a strong duality can be observed between the roles of numerical and symbolic representations.

2.4 Fusion systems and types of architecture

A general fusion problem can be stated in the following terms: given l sources S_j representing heterogeneous data on the observed phenomenon, take a decision D_i on an element x, where x can be for instance a pixel in image processing or any other object, and D_i belongs to a decision space $D = \{D_1, ..., D_n\}$ (or set of hypotheses). As opposed to symbolic fusion methods (see e.g. [164, 126, 86]), in numerical fusion methods, the information relating x to each possible decision D_i according to each source S_j is represented as a number M_i^j having different properties and different meanings depending on the mathematical fusion framework. The global fusion scheme consists in taking a decision on x as a function of all $M_i^j, 1 \leq i \leq n, 1 \leq j \leq l$. Typically, M_i^j evaluates how much source j supports decision i. In this scheme, all information is taken into account, but is complex and hardly tractable. Therefore, two degraded systems are often defined, that will be be referred as centralized and distributed schemes in the following.

In the distributed scheme, decision is taken locally on each source S_j separately, from $M_i^j, 1 \leq i \leq n$.

Then the local decisions are merged in a global one. Many distributed systems have been developed for real-time and military applications, since partial decision are taken as soon as information is available. This scheme is also called decision fusion [166, 58]. Methods in distributed decision often assume 2 possible decisions only D_0 or D_1 (or sometimes 3, the third one representing no decision). In this case, the problem amounts to find an application f from $\{D_0, D_1\}^l$ into $\{D_0, D_1\}$ [143]. The main methods rely on probabilistic models, empirical estimations based on samples, nearest neighbor rules, etc. In this scheme, imprecision can be taken into account for each sensor or each information source separately. Models are then designed adaptively for each type of information, and the addition of a new source of information is easily done. However, no process in such schemes leads to a cooperation between sensors for reducing imprecision or ambiguity provided by one source by using information provided by another one, except at the decision level, which is quite rough for this task. Moreover, conflicts between local decisions are difficult to solve, since all numerical information on information imperfections is lost[1].

In the centralized scheme, the measures related to each possible decision i and provided by all sources are combined in a global (still numerical) evaluation of this decision, taking the form, for each i: $M_i = F[M_i^1, M_i^2, ..., M_i^l]$, where F is a fusion operator. Then a decision is made from the set of $M_i, 1 \leq i \leq n$. In this scheme, the final (binary) decision is issued at the end of the processing chain. Therefore we avoid to make decisions at intermediate steps with partial information only, and thus we diminish contradictions and conflicts, which usually require a difficult control or arbitration step. In centralized schemes, imprecision and more generally imperfection of information is combined at a numerical level, and therefore fusion between sources can be better exploited to reduce these imperfections before the final decision is made.

The centralized scheme is often the most efficient when real time decision is not mandatory[2]. The main steps of fusion can then be described as [36, 39]:

1. modeling information and its imperfections,
2. estimation of the $M_i^j, 1 \leq i \leq n$ according to the chosen mathematical framework,
3. combination, i.e. choice of an appropriate fusion operator F [21],
4. decision.

Although the final aim is decision, the previous steps are very important and should be designed properly in order to exploit as well as possible the specific properties of the data, their complementarity and redundancy. Different choices in the modeling and combination steps may lead to drastically different decisions.

3 Fusion in image processing and cognitive vision

THIS SECTION COULD PROBABLY BE INCLUDED IN THE REASONING PART OF CCV ONTOLOGY, SECTION ISSUES (3.2).

3.1 Objectives

Fusion in image processing and cognitive vision is closely linked to decision making, and has actually little to do with registration, that can be considered as but a preliminary step.

[1]One possible solution consists in keeping several possible individual decisions for each source, and in considering all of them for the fusion.

[2]A scheme that can be seen as intermediary between centralized and distributed decision consists in choosing the necessary information among the $M_i^j, 1 \leq i \leq n, 1 \leq j \leq l$ in order to solve some specific decision problems. Such hybrid schemes often involve expert knowledge and are mainly used in rule-based and multi-agent systems.

Decision can take various forms, ranging from hypothesis evaluation, to classification, object recognition, scene interpretation and understanding, reasoning about events and structures, etc.

Several decision tasks can benefit from fusion of multi-source data, including estimation, hypothesis testing, or model validation. Decision can also take the form of detecting the presence of some phenomenon (e.g. a particular form of an electrocardiogram, or some rupture in a statistical model).

A field where data fusion has been widely used concerns tracking, i.e. the decision consists in assessing the movement of some objects. This task may include several partial decisions like detection of the objects of interest, localization, and possibly recognition.

An important field concerns diagnosis, classification, and pattern recognition. In such problems, the decision consists in labeling objects in a scene, in attributing each pixel or object in an image to a class, in detecting unexpected features, etc. In image fusion, classification and recognition is probably one of the most typical type of decision we may find in the literature. This is a problem closed to multi-criteria aggregation problems. These tasks are part of the more general problem of scene interpretation. The aim of fusion is then to get a global point of view on a scene, by gathering the different points of view provided by the different sources (the sources are then considered as experts or observers, each of them looking at the scene under a different aspect, using its own criteria and specificities). The sources can be different images but also different features extracted from the same image using image processing tools dedicated to specific aspects of the image. Under this form, signal and image fusion can be compared to expert opinion pooling problems. Another part of scene interpretation concerns updating, which aims at improving the information using additional sources, i.e. at getting more precise and more reliable information about the scene. In the case of dynamic scenes, evolving in time, updating also involves processes able to introduce new information or changes in the objects in the global interpretation.

3.2 Fusion situations

Depending on the application at hand, fusion problems can occur in very different situations, where different types of information are available. The main fusion situations are the following:

- Several images from a unique sensor: this can typically be the case of several channels of one satellite, of multi-echo images in magnetic resonance imaging (MRI), or of image sequences in dynamic scenes. Data are then quite homogeneous since they correspond to similar physical measures.
- Several images from different sensors: this is the most frequent case, where different physical principles lead to different and complementary points of view on the scene. Let us mention for instance fusion of SPOT and ERS images, of US and MRI data, etc. Heterogeneity is then a major feature of the data. Each image provides a partial view and are not informative about characteristics they are not dedicated to (for instance an anatomical MRI does not provide functional information).
- Several pieces of information extrated from one image: this refers to situations where several detectors, operators, classifiers, etc. are applied on the image in order to extract various pieces of information. They rely on different data characteristics which usually make the data to be combined rather heterogeneous. The extracted information can concern the same object (fusion of contour detectors for instance), or different objects and then a global interpretation of the scene is searched. Information can be obtained at different levels (local or more structural when dealing with spatial relationships between objects).
- Image and another source of information: the additional source of information can be a model, which can be either particular as a map, or generic as an anatomical atlas, knowledge bases, rules, expert information, etc. Again the pieces of information are heterogeneous, both in their nature and in their initial representation (image for a digital map, but also knowledge bases, linguistic descriptions, etc.).

3.3 Data and knowledge characteristics

In this Section, we summarize some aspects of image information that have to be taken into account in image fusion. Some of them are very specific to image processing and vision, and have to be introduced in fusion schemes and mathematical frameworks, which are generally issued from other domains.

First, the information to handle in image fusion is often **heterogeneous**. In several applications, several imaging techniques have to be used together to answer a specific question. They provide different aspects and different points of view on the problem by exploiting different physical properties. For instance when planning some surgical operation in medical imaging, necessary images can be as heterogeneous as anatomical images (provided by MRI or CT), angiographic images (MRA, spiral CT, etc.), functional images (PET, functional MRI). These images are not informative about the features they are not dedicated to. Similar examples can be found in other domains, in particular in satellite or aerial imaging. An additional cause of heterogeneity comes from the fact that image information needs to be combined with external information to make sense. This can be information on acquisition conditions, on the observed phenomenon, or more generally expert knowledge related to the problem at hand. This knowledge can be expressed either in "iconic" terms, under the form of atlases or maps, or in "propositional" terms, under the form of linguistic expressions or rules. This is an aspect that is quite specific to vision. Indeed, unlike in other domains where data fusion is used, especially in artificial intelligence, symbolic information is not only the information injected in the process by the expert in a propositional way, but it may also be derived from graphical documents since such representations carry a rather abstract interpretation of the objects contained in the scene. Such information is considered as symbolic, as opposed to the numerical information constituted by location, intensity, etc. that may be found in images.

Duality between numerical and symbolic information emerges at this level. Let us take the example of a map and an aerial image of the same area [123]. The numerical information carried by the image provides a quite accurate description of the scene, but the interpretation attached to it is hard to derive. For instance, it is generally difficult to assess the type of a building, although its drawing on the image is accurate. On the contrary, the map carries symbolic information as a semantic meaning of the objects represented on the map but its shape is often sketchy. Here lies also an example of duality between imprecision and uncertainty (imprecision on location and uncertainty on the nature of the objects). This example illustrates how imprecision may be quite different depending on the level of representation (pixel in an image, object in a map for instance).

The situation is even more complex if we include in this scheme image processing algorithms used to extract information about objects, and the attributes describing the objects. These attributes can be numerical or symbolic, and they contribute to the nature of the objects. Therefore, the information we have to deal with is embedded in a complex net, where images, objects, algorithms, measures, interpretations are closely related to each other. Imprecision can be attached to any of these elements. One of the problems to be solved is therefore to understand and model the influence, on the nature of one element, of the imprecision associated with the others. In particular, in order to define accurately these imprecisions and their relationships, we found it difficult to provide some reasonable constructive intermediate between generic definitions and naive instances.

Numerical-symbolic fusion and it specificities when applied to image information under imprecision is a very complex problem, that still did not find a definite solution and that remains a very active field of research.

However, a first step towards understanding and management of imprecisions and their relationships can be performed by studying the different **causes of imprecision**. Indeed, imprecision in image information may be due to several factors, ranging from the observed phenomenon to the algorithms artifacts. In biology, a soft transition between tissues (e.g. healthy and pathological tissues) is surely a cause of imprecision inherent to the nature of the observed objects. Also, if tissues have similar characteristics, images that represent these characteristics will poorly discriminate these tissues. This will result in an uncertainty on the belonging of a pixel to one or the other tissue. Another cause of imprecision is due to the discrete nature of numerical images, resulting in a delocalization of information contained in a small volume at only one point. The partial volume effect (the presence of several tissues in one pixel or voxel) belongs also to this type of spatial imprecision. Other image imperfections can be caused by numerical reconstruction algorithms in computed imaging. One example is the Gibbs effect that may appear in MRI around sharp transitions. At the processing level, imprecision is often induced by the chosen algorithms (e.g. filtering, contour detection, registration between images, etc.), another place where the duality between imprecision and uncertainty can be found [159].

Another aspect that is quite specific to image processing and vision, as compared to other application domains of data fusion, is the **complexity of the information**. This is partly due to the previously mentioned characteristics but also to the increasing number of acquisition techniques and to the huge data sets that have to be dealt with. Typically, one MRI brain image contains $256 \times 256 \times 128$ voxels, one satellite image contains 6000×6000 pixels, and several images of this size have to be combined in a fusion process. The large data volumes, and the statistical measures that are therefore made possible, may explain the use of statistical approaches in most image fusion schemes. The complexity of the fusion process also comes from the simultaneous redundancy and complementarity between images, closely related to the heterogeneity aspects. One of the main tasks of image fusion is to exploit redundancy, in order to increase the global information, and complementarity, to improve certainty and precision. The decision is thus improved by the fusion in terms of both quantity and quality.

The main information in image processing and vision, which is specific to this domain, is the **spatial information**. The previous examples have already shown how it appears at all levels, and the types of imprecision attached to it. When using a fusion method, often issued from another domain, we thus have to incorporate spatial information in the process.

3.4 Constraints

From the architecture point of view, centralized systems are rarely imposed. The most frequent case is off-line fusion, where all information is simultaneously available. Centralized systems can then be used.

Real-time constraints are quite rare too, except in surveillance or multimedia applications.

On the contrary, strong constraints are linked to the spatial characteristic of the data. Spatial consistency is a strong aspect that deserve important research work. It can be achieved either at a local level, using the local spatial context, or at a more structural level, using spatial relations between structures of objects in the scene.

The huge volume of data can impose computation time and algorithmical complexity constraints. Therefore at pixel level, only simple operations are usually implemented. More sophisticated operations often require higher level information as well as structured representations.

The complexity and quantity of data also imposes to make some choice in the information and the knowledge to be fused. This choice is of course mainly guided by relevance criteria with respect to the decision objectives, but also by criteria on the relative ease in the access and representation of information and knowledge, as well as on their quality.

Finally one of the main problems in image fusion (like in other fields of image processing) concerns the **validation and evaluation** of methods. Although the truth generally more or less exists, it is often difficult or impossible to access[3].

3.5 Numerical and symbolic aspects

In image fusion, like in most image processing and vision domains, several levels of information representation are distinguished. At the lowest level, information carried by each pixel is considered. Fusion schemes at this level aim at taking a decision for each pixel, e.g. attributing each pixel to a class of

[3] In that sense, aerial or satellite imaging is probably a better field of investigation than medical imaging. At least, it allows an easier evaluation.

interest. All raw image information is then used, but such schemes are lacking in structural information and spatial coherence [52]. A precise registration between all images to be combined is generally considered as unavoidable at this level. At an intermediate level, the fusion is performed on features extracted from each image using some image processing tools. These features are typically segments, contours, or regions. Precise registration is then crucial. At the highest level, objects extracted from images are combined. This involves a preliminary recognition step and is therefore related to decision fusion. Structural information can be easily introduced at this level, and therefore makes registration less mandatory. However, the richness of the basic numerical information is lost, and conflicts may be difficult to solve, unless some numerical evaluation is attached to the objects and structures extracted from each image. Applications where segmentation results are fused (e.g. [55], [50]) can be considered as intermediate level fusion if the segmented regions or contours have no semantic interpretation (i.e. no real decision is taken on each image), or as high level fusion otherwise.

3.6 Fusion in vision vs. fusion in other domains

Let us now briefly illustrate another reason why image fusion is so different from other application domains of data fusion (see [32] for a description of fusion problems in several domains). Of course, the nature of information is an important point, and the spatial aspects as well. But, to our opinion, with respect to multi-criteria aggregation and optimization, the main difference relies on the fact that in that field, the aim is to try to find a solution that satisfies as well as possible some flexible constraints. In image fusion, the sensors do provide (more or less explicitly) a degree of satisfaction (of membership to classes for instance) and the decision amounts rather to choose the best one. With respect to voting problems, the difference is that in those problems, no truth exists, and that probably subjectivity has more to be taken into account, but according to some "ethic" rule. Subjectivity is also part of combination of expert opinions. In that domain, information is moreover usually more sparse, which makes learning perhaps more difficult than in image processing. However, this last problem is probably the closest one to our.

4 Bases of fuzzy sets and possibility theory

This Section could probably be included in the Reasoning part of CCV Ontology, Section Methods (3.3).

This part intends to provide the reader with the bases of fuzzy set and possibility theory. It contains the basic definitions and properties that are needed in the following when dealing specifically with spatial information.

Many papers and books have been published that extensively present the bases of fuzzy set theory. Beside Zadeh's seminal paper [186], the reader may refer to [103, 67, 191].

This part is organized as follows. In Section 4.1, the basic definitions of fuzzy sets are given, as long as the original fuzzy set operators proposed by Zadeh and the concept of fuzzy number. In Section 4.2, fuzzy measures are introduced, along with some examples. This leads in Section 4.3 to a short introduction to the possibility theory. More fuzzy operations are summarized in Section 4.4. They can be seen as set operations, as well as combination operators in information fusion systems. They also provide a theoretical basis for logical operators. The concepts of linguistic variable and of modifier are introduced in Section 4.5. Fuzzy relations are defined in Section 4.6. The examples of similarity and ordering relations are detailed, since they are the starting point of several studies in fuzzy image processing, in particular for pattern recognition and scene interpretation applications. The principles of fuzzy logic and approximate reasoning are shortly given in Section 4.7. Finally we present the main methods for extending crisp operations to fuzzy ones in Section 4.8.

4.1 Definition of fuzzy set fundamental concepts

4.1.1 Fuzzy sets

Let $\mathcal{U}$ be the universe, i.e. the space of objects of interest. It is a classical (or crisp) set. We denote by x, y, etc. its elements (or points). In image processing, $\mathcal{U}$ can be typically the space $\mathcal{S}$ on which the image is defined (usually $\mathbb{Z}^n$ or $\mathbb{R}^n$, with $n = 2, 3, ...$)[4]. Then the elements of $\mathcal{U}$ are the points of the image (pixels, voxels). The universe can also be a set of values taken by some image characteristics, for instance the scale of grey levels. Then x is a value (a grey level). The set $\mathcal{U}$ can also be a set of features or objects extracted from the images (e.g. segments, regions, objects), leading to a higher level representation of the image content.

A subset X of $\mathcal{U}$ is defined by its characteristic function μ_X, such that:

$$\mu_X(x) = \begin{cases} 1 & if \quad x \in X \\ 0 & if \quad x \notin X \end{cases} \tag{1}$$

The characteristic function μ_X is a binary function, specifying the crisp membership of each point of $\mathcal{U}$ to X.

Fuzzy set theory aims at dealing with gradual membership. A fuzzy subset of $\mathcal{U}$ is defined through its membership function μ from $\mathcal{U}$ into $[0, 1]$[5]. For each x of $\mathcal{U}$, $\mu(x)$ is a value of $[0, 1]$ representing the membership degree of x to the fuzzy subset, i.e. to which extent x belongs to it. Although the correct terminology would be to speak of "fuzzy subset", commonly the simplified term "fuzzy set" is used. We keep this term in the following, for sake of simplicity.

Various notations are used for denoting a fuzzy set. A fuzzy set is completely defined by the set:

$$\{(x, \mu(x)), x \in X\}, \tag{2}$$

which can be noted as:

$$\int_{\mathcal{U}} \mu(x)/x \tag{3}$$

or in the discrete case:

$$\sum_{i=1}^{N} \mu(x_i)/x_i \tag{4}$$

where N denotes the cardinality of $\mathcal{U}$.

Since the set of all couples $(x, \mu(x))$ is completely equivalent to the definition of the function μ, we have chosen here to simplify notations and to refer always to the functional notation μ as a function of $\mathcal{U}$ into $[0, 1]$, where μ will be called indifferently fuzzy set or membership function.

The support of a fuzzy set μ is the set of points that have a strictly positive membership to μ (it is a crisp set):

$$Supp(\mu) = \{x \in \mathcal{U},\ \mu(x) > 0\}. \tag{5}$$

The core of a fuzzy set μ is the set of points that belongs completely to μ (it is a crisp set):

$$Core(\mu) = \{x \in \mathcal{U},\ \mu(x) = 1\}. \tag{6}$$

A normalized fuzzy set μ is such as at least one point belongs completely to μ (i.e. $Core(\mu) \neq \emptyset$):

$$\exists x \in \mathcal{U},\ \mu(x) = 1. \tag{7}$$

A unimodal fuzzy set μ is such that there exists a unique point x such that $\mu(x) = 1$. A less constraining definition allows the core of μ to be a compact set and not only one point.

[4] In the following, $\mathcal{U}$ will denote a general space, while $\mathcal{S}$ will denote more specifically the spatial domain.

[5] The interval $[0, 1]$ is the most used. However, any other interval, or other set (typically a lattice) could be used.

4.1.2 Set operations: original definition of L. Zadeh

Since fuzzy sets have been introduced by L. Zadeh in [186] in order to generalize sets, the first operations that have been proposed are set operations. We recall here the original definitions proposed by L. Zadeh. Further operations are defined later, in Section 4.4.

The equality of two fuzzy sets is defined by the equality of their membership functions:

$$\mu = \nu \Leftrightarrow \forall x \in \mathcal{U}, \mu(x) = \nu(x). \tag{8}$$

The inclusion of a fuzzy set in another one is defined as an inequality on their membership functions:

$$\mu \subset \nu \Leftrightarrow \forall x \in \mathcal{U}, \mu(x) \leq \nu(x). \tag{9}$$

The intersection (respectively union) between two fuzzy sets is defined as the pointwise minimum (respectively maximum) of their membership values:

$$\forall x \in \mathcal{U}, (\mu \cap \nu)(x) = \min[\mu(x), \nu(x)], \tag{10}$$

$$\forall x \in \mathcal{U}, (\mu \cup \nu)(x) = \max[\mu(x), \nu(x)]. \tag{11}$$

The complementation of a fuzzy set is defined as:

$$\forall x \in \mathcal{U}, \mu^C(x) = 1 - \mu(x). \tag{12}$$

The main properties of these definitions are the following:

- they are all consistent with binary set operations: in the particular case where the membership functions only take values 0 and 1 (i.e. they are crisp sets), these definitions reduce to the classical binary definitions; note that this property is important since it is the least we can ask to the fuzzy extension of a binary operation;
- $\mu = \nu \Leftrightarrow \mu \subset \nu$ and $\nu \subset \mu$;
- the 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 the complementation: $(\mu \cap \nu)^C = \mu^C \cup \nu^C$;
- if we consider the empty set $\emptyset$ as a fuzzy set having membership values all equal to 0, then we have $\mu \cap \emptyset = \emptyset$ and $\mu \cup \emptyset = \mu$, for all fuzzy set μ defined on $\mathcal{U}$;
- if we consider the universe as a fuzzy set having membership values all equal to 1, then we have $\mu \cap \mathcal{U} = \mu$ and $\mu \cup \mathcal{U} = \mathcal{U}$, for all fuzzy set μ defined on $\mathcal{U}$.

These properties are the same as the corresponding binary operations. However, some binary properties are lost, in particular the excluded-middle and non-contradiction laws, since:

$$\mu \cup \mu^C \neq \mathcal{U}, \tag{13}$$

$$\mu \cap \mu^C \neq \emptyset. \tag{14}$$

4.1.3 Structure and types of fuzzy sets

Let us denote by $\mathcal{C}$ the set of all crisp subsets of $\mathcal{U}$, and by $\mathcal{F}$ the set of all fuzzy subsets of $\mathcal{U}$. The set $\mathcal{C}$ is a Boolean lattice for intersection and union (i.e. a complemented distributive lattice). It can be considered as the lattice induced by the structure of $\{0, 1\}$. The interval $[0, 1]$ is a pseudo-complemented distributive lattice (in the lattice terminology, the complementation to 1 is a pseudo-complementation), which induces a pseudo-complemented distributive lattice structure on $\mathcal{F}$.

Several types of fuzzy sets can be considered. Until now, we considered that membership functions take values that are numbers. These are called type-1 fuzzy sets. But membership values are not necessarily numbers. They can also be fuzzy sets, by making use of the lattice structure of the set of fuzzy sets. A type-2 fuzzy set is a fuzzy set whose membership values are type-1 fuzzy sets. More generally a type-m fuzzy set is a fuzzy set whose membership values are type-$(m-1)$ fuzzy sets (for $m > 1$). Such extensions of fuzzy sets are particularly useful when the membership value that we can attach to an element is imprecisely defined. In the sequel, mainly type-1 fuzzy sets are considered. The operations that are defined below can be generalized to type-m fuzzy sets, using the extension principle which is introduced in Section 4.8.

4.1.4 α-cuts

The α-cut of a fuzzy set μ is the crisp set defined as:

$$\mu_\alpha = \{x \in \mathcal{U},\ \mu(x) \geq \alpha\}. \tag{15}$$

Strict (or strong) α-cuts are defined as:

$$\mu_\alpha = \{x \in \mathcal{U},\ \mu(x) > \alpha\}. \tag{16}$$

A fuzzy set can be considered as a "stack" of its α-cuts. It can be reconstructed from them using different formulas, the main ones being:

$$\mu(x) = \int_0^1 \mu_\alpha(x) d\alpha, \tag{17}$$

$$\mu(x) = \sup_{\alpha \in]0,1]} \min(\alpha, \mu_\alpha(x)), \tag{18}$$

$$\mu(x) = \sup_{\alpha \in]0,1]} (\alpha \mu_\alpha(x)). \tag{19}$$

Let us now look at the links with Zadeh's operators. The following relationships hold:

$$\forall(\mu,\nu) \in \mathcal{F}^2,\ \mu = \nu \Leftrightarrow \forall \alpha \in]0,1], \mu_\alpha = \nu_\alpha, \tag{20}$$

$$\forall(\mu,\nu) \in \mathcal{F}^2,\ \mu \subset \nu \Leftrightarrow \forall \alpha \in]0,1], \mu_\alpha \subset \nu_\alpha, \tag{21}$$

$$\forall(\mu,\nu) \in \mathcal{F}^2,\ \forall \alpha \in [0,1], (\mu \cap \nu)_\alpha = \mu_\alpha \cap \nu_\alpha, \tag{22}$$

$$\forall(\mu,\nu) \in \mathcal{F}^2,\ \forall \alpha \in [0,1], (\mu \cup \nu)_\alpha = \mu_\alpha \cup \nu_\alpha, \tag{23}$$

$$\forall \mu \in \mathcal{F},\ \forall \alpha \in [0,1], (\mu^C)_\alpha = (\mu_{1-\alpha})^C. \tag{24}$$

We draw the reader's attention to the last equation, which is not as straightforward as the previous ones.

Taking the α-cut of a fuzzy set, for some given value of α, amounts to select the elements of $\mathcal{U}$ that belongs at least to the degree α to the fuzzy set. It is therefore a kind of thresholding process on the membership function. It can also be seen as a "defuzzification" process, i.e. a mean to return in the domain of crisp sets, which is used for instance to take a decision in a fuzzy system.

4.1.5 Cardinality

In this Section, we consider only fuzzy sets that are defined over a finite universe, or that have a finite support. This is not restrictive when applying fuzzy sets theory to image processing, since in this domain, we are working mainly with finite (discrete) universes.

The cardinality of such a fuzzy set μ is defined as:

$$|\mu| = \sum_{x \in \mathcal{U}} \mu(x), \tag{25}$$

or, if $\mathcal{U}$ is not finite but the support of μ is finite:

$$|\mu| = \sum_{x \in Supp(\mu)} \mu(x). \tag{26}$$

Again this definition is consistent with the cardinality of a crisp set. It can be interpreted as counting each point for an amount corresponding to its membership to the fuzzy set. It is also called the power of the fuzzy set (e.g. in [122]).

This definition can be extended to the case where $\mathcal{U}$ is not finite but measurable. Let M be a measure on $\mathcal{U}$ (such that $\int_{\mathcal{U}} dM(x) = 1$). The fuzzy cardinality of μ is defined as:

$$|\mu| = \int_{\mathcal{U}} \mu(x) dM(x). \tag{27}$$

4.1.6 Convexity

In this Section, the universe $\mathcal{U}$ is a real Euclidean space (of any dimension).

The convexity of a fuzzy set is defined from its α-cuts: a fuzzy set μ is convex iff its α-cuts are convex (for all α in $[0,1]$). This definition is not equivalent to the convexity of the membership function in an analytical sense[6]. The analytical equivalent expression for fuzzy convexity is: μ is convex iff:

$$\forall (x,y) \in \mathcal{U}^2,\ \forall \lambda \in [0,1],\ \min(\mu(x), \mu(y)) \leq \mu(\lambda x + (1-\lambda) y). \tag{28}$$

4.1.7 Fuzzy numbers

In this Section, we set $\mathcal{U} = \mathbb{R}$.

A fuzzy quantity is a fuzzy set μ on $\mathbb{R}$. A fuzzy interval is a fuzzy quantity which is convex (i.e. its α-cuts are intervals). There is an equivalence between the upper-semi-continuity of μ and the fact that its α-cuts are closed intervals.

A fuzzy number is a fuzzy interval upper-semi-continuous (u.s.c.) with compact support and unimodal. An example of fuzzy number representing "about 10" is shown in Figure 1.

Less strict definitions are also considered in the literature, in particular by accepting not only one single modal value, but an interval of modal values, i.e. there exist four reals a, b, c, d, with $a \leq b \leq c \leq d$ such that $\mu(x) = 0$ outside the interval $[a,d]$, increasing on $[a,b]$, decreasing on $[c,d]$ and equal to 1 on $[b,c]$ [91, 92].

A fuzzy number can be interpreted as a flexible representation of an imprecise quantity, which is more general than a crisp interval.

Let us return to the definition of the fuzzy cardinality. It has been defined previously as a number. However, when considering a fuzzy set, it can be interesting to define it as a fuzzy number, since the

[6] The convexity of a function f is defined as $\forall (x,y), f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda) f(y)$.

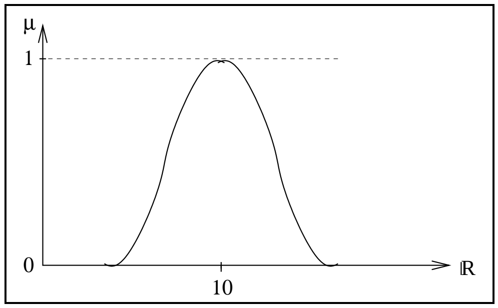


Figure 1: Fuzzy number representing "about 10".

cardinality of an imprecisely defined set may be considered as imprecise too. The cardinality of a fuzzy set as a fuzzy number (see [67]) is defined as:

$$|\mu|_f(n) = \sup\{\alpha \in [0,1],\ |\mu_\alpha| = n\}. \tag{29}$$

This constitutes a first example of fuzzy number.

A widely used type of fuzzy numbers is called L-R fuzzy numbers. They are defined through a parametric representation of their membership function. A L-R fuzzy number μ is defined as:

$$\forall x \in \mathbb{R},\ \mu(x) = \begin{cases} L(\frac{m-x}{\alpha}) & for \quad x \leq m \\ R(\frac{x-m}{\beta}) & for \quad x \geq m \end{cases} \tag{30}$$

where α and β are strictly positive numbers called left and right spreads, m is a number called mean value of the fuzzy number and L and R are functions (referred to as the reference functions of the fuzzy number) having the following properties:

- $\forall x \in \mathbb{R}, L(x) = L(-x)$,
- $L(0) = 1$,
- L is non-increasing on $[0, +\infty[$,

and similar properties for R.

One of the main advantages of such fuzzy numbers is that they have a compact representation and computations on fuzzy numbers can be made easily.

4.2 Fuzzy measures

In this Section, fuzzy measures are introduced, following the approach by Sugeno [165], and several examples of fuzzy measures are given. More details can be found in [165, 67].

4.2.1 Fuzzy measure of a crisp set

A fuzzy measure is a function f from $\mathcal{C}$ (i.e. defined on crisp sets) into $[0, 1]$ satisfying the following conditions:

1. $f(\emptyset) = 0$;
2. $f(\mathcal{U}) = 1$;
3. monotonicity: $\forall (A, B) \in \mathcal{C}^2,\ A \subset B \Rightarrow f(A) \leq F(B)$;

4. continuity:

$$\forall i \in \mathbb{N}, \forall A_i \in \mathcal{C},\ A_1 \subset A_2 ... \subset A_n ... \text{ or } A_1 \supset A_2 ... \supset A_n ...$$
$$\Rightarrow \lim_{i\to\infty} f(A_i) = f\left(\lim_{i\to\infty} A_i\right).$$

Noticeable properties of fuzzy measures are:

$$\forall (A,B) \in \mathcal{C}^2,\ f(A \cup B) \geq \max[f(A), f(B)], \tag{31}$$

$$\forall (A,B) \in \mathcal{C}^2,\ f(A \cap B) \leq \min[f(A), f(B)]. \tag{32}$$

Note that the definition of a fuzzy measure does not assume any additivity constraint. It could also be called non-additivity measure, since the link with fuzzy set theory is rather loose.

4.2.2 Examples of fuzzy measures

Several families of fuzzy measures can be found in the literature. Among the most used, let us mention:

- the probability measures;
- the λ-fuzzy measures, obtained by relaxing the additivity constraint of a probability measure as:

  $$\forall (A,B) \in \mathcal{C}^2,\ A \cap B = \emptyset \Rightarrow f(A \cup B) = f(A) + f(B) + \lambda f(A) f(B) \tag{33}$$

  with $\lambda > -1$;
- the belief functions and plausibility functions, used in the evidence theory of Dempster-Shafer [158];
- the possibility measures [189], that are introduced in Section 4.3.

Links between several types of fuzzy measures have been investigated by Banon [7] (see also [67]).

4.2.3 Fuzzy integrals

Fuzzy integrals [165, 95] are the counter-part of Lebesgue's integrals when the integration of a function is made with respect to a fuzzy measure. Two types of fuzzy integrals are distinguished.

The Sugeno's integral of a measurable function f, defined from $\mathcal{U}$ into $[0,1]$, with respect to a fuzzy measure μ is defined as:

$$S_\mu(f) = \int f \circ \mu = \sup_{\alpha \in [0,1]} \min[\alpha, \mu(\{x \in \mathcal{U}, f(x) > \alpha\})]. \tag{34}$$

In the finite case ($|\mathcal{U}| = N$), this expression is equivalent to:

$$S_\mu(f) = \int f \circ \mu = \max_{i=1}^{N} \min[f(x_{p(i)}), \mu(A_i)], \tag{35}$$

where p is a permutation on $\{1, 2 ... N\}$ such that:

$$0 \leq f(x_{p(1)}) \leq ... \leq f(x_{p(N)})$$

and where $A_i = \{x_{p(1)}, ..., x_{p(N)}\}$.

The Choquet's integral of a measurable function f, defined from $\mathcal{U}$ into $\mathbb{R}^+$, with respect to a fuzzy measure μ is defined by:

$$C_\mu(f) = \int f d\mu = \int_0^{+\infty} \mu(\{x, f(x) > \alpha\}) d\alpha. \tag{36}$$

In the finite case, we obtain:

$$C_\mu(f) = \int f d\mu = \sum_{i=1}^{N} [f(x_{p(i)}) - f(x_{p(i-1)})]\mu(A_i), \tag{37}$$

with $f(x_p(0)) = 0$.

The properties of these integrals have been studied by several researchers [165, 131, 95]. The main ones in the finite case are the following:

- for the measure $\mu_{\min}$ defined as $\forall A \subset \mathcal{U}, A \neq \mathcal{U}, \mu_{\min}(A) = 0$ and $\mu_{\min}(\mathcal{U}) = 1$, $S_{\mu_{\min}}(f)$ and $C_{\mu_{\min}}(f)$ are equal to the minimum of the values taken by f;
- for the measure $\mu_{\max}$ defined as $\forall A \subset \mathcal{U}, A \neq \emptyset, \mu_{\max}(A) = 1$ and $\mu_{\max}(\emptyset) = 0$, $S_{\mu_{\max}}(f)$ 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 any fuzzy measure μ, the following monotony property holds:
$$(\forall x \in \mathcal{U}, f(x) \leq f'(x)) \Rightarrow \begin{cases} S_\mu(f) & \leq & S_\mu(f') \\ C_\mu(f) & \leq & C_\mu(f') \end{cases} \tag{38}$$
this property being also true in the infinite case;
- for any measurable function f and any two fuzzy measures μ and μ', the following monotony property holds:
$$(\forall A \subset \mathcal{U}, \mu(A) \leq \mu'(A)) \Rightarrow \begin{cases} S_\mu(f) & \leq & S_{\mu'}(f) \\ C_\mu(f) & \leq & C_{\mu'}(f) \end{cases} \tag{39}$$
this property being also true in the infinite case;
- as can be deduced from the previous properties, the following inequalities hold for any measurable function f and any fuzzy measure μ:
$$\min_{i=1}^{N} f(x_i) \leq S_\mu(f) \leq \max_{i=1}^{N} f(x_i); \tag{40}$$
$$\min_{i=1}^{N} f(x_i) \leq C_\mu(f) \leq \max_{i=1}^{N} f(x_i); \tag{41}$$
- for any additive measure (or σ-additive in the infinite case), the Choquet's integral coincides with the Lebesgue's integral; in this sense, fuzzy integral can be considered as an extension of Lebesgue's integral;
- for any fuzzy measure μ, Sugeno's and Choquet's integrals satisfy the following continuity property: for any sequence of measurable functions f_n on $\mathcal{U}$ such that:
$$\lim_{n \to +\infty} f_n = f$$
we have:
$$\lim_{n \to +\infty} S_\mu(f_n) = S_\mu(f), \tag{42}$$
$$\lim_{n \to +\infty} C_\mu(f_n) = C_\mu(f). \tag{43}$$

Applications of fuzzy integrals can be found in multi-criteria aggregation, data fusion, pattern recognition.

4.2.4 Measure of fuzzy sets

In the previous Sections, the measures were always applied on crisp sets. If we now consider fuzzy sets, we may also need measures to provide a quantitative evaluation of a fuzzy set. Such measures are called fuzzy set measures [45] or evaluation measures [75]. There is no real consensus on the definition of such measures. We consider here the less constraining one, as proposed in [45].

A fuzzy set measure is a function M from $\mathcal{F}$ into $\mathbb{R}^+$ such that:

1. $M(\emptyset) = 0$;
2. $\forall (\mu, \nu) \in \mathcal{F}^2, \ \mu \subset \nu \Rightarrow M(\mu) \leq M(\nu)$.

One or several of the following requirements may also be added, depending on the application:

- M takes values in $[0, 1]$;
- $M(\mathcal{U}) = 1$;
- $M(\mu) = 0 \Leftrightarrow \mu = \emptyset$;
- $M(\mu) = 1 \Leftrightarrow \mu = \mathcal{U}$.

Simple examples of fuzzy set measures are fuzzy cardinality, cardinality of the support of μ, supremum of μ, etc. Measures of fuzziness, as described in the next Section, are additional examples.

4.2.5 Fuzziness measures

One question related to the evaluation of a fuzzy set concerns the degree of fuzziness. De Luca and Termini [122] have proposed to define a degree of fuzziness as a function f from $\mathcal{F}$ into $\mathbb{R}^+$ such that:

1. $\forall \mu \in \mathcal{F}, \ f(\mu) = 0 \Leftrightarrow \mu \in \mathcal{C}$ (i.e. crisp sets are completely non-fuzzy and they are the only ones that satisfy this property);
2. $f(\mu)$ is maximum iff $\forall x \in \mathcal{U}, \mu(x) = 0.5$;
3. $\forall (\mu, \nu) \in \mathcal{F}^2, \ f(\mu) \geq f(\nu)$ if ν is more contrasted than μ (more crisp), i.e.:

$$\forall x \in \mathcal{U}, \left\{ \begin{array}{lll} \nu(x) \geq \mu(x) & if & \mu(x) \geq 0.5 \\ \nu(x) \geq \mu(x) & if & \mu(x) \leq 0.5 \end{array} \right.$$

4. $\forall \mu \in \mathcal{F}, \ f(\mu) = f(\mu^C)$, i.e. a fuzzy set and its complement are equally fuzzy.

In the same paper, De Luca and Termini proposed a function, called entropy of a fuzzy set [122], as degree of fuzziness, in the finite case. It is defined as:

$$E(\mu) = H(\mu) + H(\mu^C), \tag{44}$$

where $H(\mu)$ obeys a similar definition as Shannon's entropy:

$$H(\mu) = -K \sum_{i=1}^{N} \mu(x_i) \log \mu(x_i). \tag{45}$$

Obviously, E satisfies the axioms of a degree of fuzziness. Moreover, the following relationship holds:

$$H(\max(\mu, \nu)) + H(\min(\mu, \nu)) = H(\mu) + H(\nu), \tag{46}$$

and

$$E(\max(\mu,\nu)) + E(\min(\mu,\nu)) = E(\mu) + E(\nu). \tag{47}$$

A lot of other measures of fuzziness have been proposed in the literature, sharing similar properties. Let us mention the most used ones:

- the Hamming distance to the closest binary set, which is nothing else than the 0.5-cut [103]:

$$f(\mu) = \sum_{i=1}^{N} |\mu(x_i) - \mu_{1/2}(x_i)|; \tag{48}$$

- the Hamming or quadratic distance between μ and its complementary set [178] or more generally:

$$\begin{aligned} f(\mu) &= \left[\sum_{i=1}^{N} |\mu(x_i) - \mu^C(x_i)|^p\right]^{1/p} \\ & \left[\sum_{i=1}^{N} |2\mu(x_i) - 1|^p\right]^{1/p}; \end{aligned} \tag{49}$$

- the measure proposed by Kosko [111], that compares the intersection of μ and μ^C with their union using the following formula:

$$\frac{|\min(\mu,\mu^C)|}{|\max(\mu,\mu^C)|}; \tag{50}$$

- the generalized entropy defined using some generating function, either in an additive form or in a multiplicative one[11]:
  - the additive form is defined as:

$$f(\mu) = \sum_{i=1}^{N} g[\mu(x_i)] + g[1 - \mu(x_i)] \tag{51}$$

    where g is a function from $[0,1]$ into $\mathbb{R}^+$ 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 leads to the fuzzy entropy has defined in [122]).
  - the multiplicative form is defined as:

$$f(\mu) = \sum_{i=1}^{N} g[\mu(x_i)]g[1 - \mu(x_i)] \tag{52}$$

    where g is a function from $[0,1]$ into $\mathbb{R}^+$ such that:

$$\forall t \in [0,1],\ g'(t) > 0 \ and \ g''(t) < 0.$$

    Examples of generating functions are $g(t) = te^{1-t}$, $g(t) = t^\alpha$.

4.3 Elements of possibility theory

Possibility theory is derived from fuzzy set theory. It has been first introduced by Zadeh in [189], and then widely studied by several authors, in particular Dubois and Prade [67, 71].

4.3.1 Necessity and possibility

A possibility measure is a function Π from $\mathcal{C}$ into $[0,1]$ such that:

1. $\Pi(\emptyset) = 0$,
2. $\Pi(\mathcal{U}) = 1$,
3. $\forall I \subset \mathbb{N}, \forall A_i \subset \mathcal{U}(i \in I),\ \Pi(\cup_{i\in I} A_i) = \sup_{i\in I} \Pi(A_i)$.

In the finite case, a possibility measure is a fuzzy measure. It appears as a limit case of equation 31 which was deduced from the continuity property of a fuzzy measure.

By duality a necessity measure is defined as a function from $\mathcal{C}$ into $[0,1]$ such that:

$$\forall A \subset \mathcal{U},\ N(A) = 1 - \Pi(A^C). \tag{53}$$

This duality means that if an events is necessary, its contrary is impossible.

The necessity measure satisfies the following properties:

1. $N(\emptyset) = 0$,
2. $N(\mathcal{U}) = 1$,
3. $\forall I \subset \mathbb{N}, \forall A_i \subset \mathcal{U}(i \in I),\ N(\cap_{i\in I} A_i) = \inf_{i\in I} N(A_i)$.

Conversely, from any measure satisfying these properties, a possibility function can be derived which satisfies the duality relationship.

Additional useful properties of possibility and necessity measures are:

- $\forall A \subset \mathcal{U},\ \max(\Pi(A), \Pi(A^C)) = 1$, which expresses the fact that either one of the two subsets A and A^C is completely possible;
- $\forall A \subset \mathcal{U},\ \min(N(A), N(A^C)) = 0$, which expresses that two contrary events cannot be simultaneously necessary;
- $\forall A \subset \mathcal{U},\ \Pi(A) \geq N(A)$ (an event has to be possible before being necessary);
- $\forall A \subset \mathcal{U},\ N(A) > 0 \Rightarrow \Pi(A) = 1$ (since $N(A) > 0 \Rightarrow \Pi(A^C) < 1$, and $\max(\Pi(A), \Pi(A^C)) = 1$);
- $\forall A \subset \mathcal{U},\ \Pi(A) < 1 \Rightarrow N(A) = 0$;
- $\forall A \subset \mathcal{U},\ N(A) + N(A^C) \leq 1$;
- $\forall A \subset \mathcal{U},\ \Pi(A) + \Pi(A^C) \geq 1$.

The two last properties are non-additive properties. It means that given $\Pi(A)$, it is not enough to determine completely $N(A)$ (this contrasts with probability theory). Therefore the uncertainty attached to an event is expressed as two numbers and not only one.

4.3.2 Possibility distribution

A possibility distribution is a function π from $\mathcal{U}$ into $[0,1]$ with the following normalization condition:

$$\sup_{x\in\mathcal{U}} \pi(x). \tag{54}$$

A possibility measure can be derived from a possibility distribution in the finite case as:

$$\forall A \in \mathcal{C},\ \Pi(A) = \sup\{\pi(x), x \in A\}. \tag{55}$$

Conversely, a possibility measure induces a possibility distribution by:

$$\forall x \in \mathcal{U},\ \pi(x) = \Pi(\{x\}). \tag{56}$$

By duality, we have the following relationship between necessity and possibility distribution:

$$\forall A \in \mathcal{C},\ N(A) = 1 - \sup\{\pi(x), x \notin A\} = \inf\{1 - \pi(x), x \in A^C\}. \tag{57}$$

These definitions find a simple interpretation if we consider the problem of representing the value taken by a variable. Then $\mathcal{U}$ represents the domain of variation of this variable. A possibility distribution on $\mathcal{U}$ describes the degrees to which the variable can take each possible value. It is actually a fuzzy set, i.e. the fuzzy set of all possible values for this variable. The membership degree of each value to this set corresponds to the degree of possibility that the variable be equal to this value. A possibility distribution may therefore represent the imprecision attached to the exact value of a variable. Typically a fuzzy number is a possibility distribution describing the possible values that can take this number.

Let us consider for instance a classification problem in image processing. We give below a few examples where possibility distributions can be defined (they are of course far from being exhaustive):

- Let $\mathcal{U}$ be the set of classes. A possibility distribution on $\mathcal{U}$, defined for each object to be classified (point, region, etc.), may represent the degrees to which this object possibly belongs to each of the classes.
- Let $\mathcal{U}$ be a characteristic space (e.g. the grey-level scale). A possibility distribution on $\mathcal{U}$ may be defined for each class and represents, for each grey-level, the possibility that this class appears in the image with that grey-level.
- Let $\mathcal{U}$ be the image space. A possibility distribution on $\mathcal{U}$ can be defined for each class, and represents the possible locations of this class in the image.

The normalization corresponds to the fact that one value at least is considered as completely possible. Non-normalized possibility distribution can also be useful. In such cases, we do not have $\Pi(\mathcal{U}) = 1$ anymore. Also the properties $N(A) > 0 \Rightarrow \Pi(A) = 1$ and $\Pi(A) < 1 \Rightarrow N(A) = 0$ do not always hold.

In the above definition, we always considered possibility and necessity of crisp subsets of $\mathcal{U}$. Now if we consider a fuzzy subset μ of $\mathcal{U}$ ($\mu \in \mathcal{F}$), the notion if possibility has to be extended [189]. This corresponds to the following interpretation: given a possibility distribution π on $\mathcal{U}$, associated with a variable X taking values in $\mathcal{U}$, we are interested in knowing to which extent "X is μ". This is defined as:

$$\Pi(\mu) = \sup_{x \in \mathcal{U}} \min(\mu(x), \pi(x)). \tag{58}$$

The possibility of μ therefore combines the degree to which the variable X takes the value x and the membership degree of x to the fuzzy set under evaluation.

4.3.3 Semantics

Membership functions and possibility distributions can have several semantics, the main ones being the following:

- similarity degree semantics (based on the notion of distance, to a prototype for instance);
- plausibility degree semantics (plausibility that an object from which only an imprecise description is known is actually the one we want to identify);
- preference degree semantics (a fuzzy class is then the set of "good" choices), this interpretation being close to the notion of utility function.

4.4 Operators

After the early work of Zadeh [186], a lot of operators have been proposed in the fuzzy set theory, to combine membership functions, or possibility distributions[7]. This operators are also named connectives, aggregation or combination operators. Several review papers summarize the main classes of operators (e.g. [70, 181, 21]).

The following Sections provide the definitions of the main classes of operators, examples of the most used forms in each class, and simple interpretations (according to [70, 71]). Further interpretations in terms of set operations and in terms of data fusion can also be provided.

Since most operators work pointwise (i.e. combine membership degrees or possibility degrees at the same point of $\mathcal{U}$), it is enough to define the operators on the values taken by membership functions or possibility distributions. They are therefore defined as functions from $[0,1]$ or from $[0,1]\times[0,1]$ into $[0,1]$.

4.4.1 Fuzzy complementation

A fuzzy complementation is a function c from $[0,1]$ into $[0,1]$ such that:

1. $c(0)=1$,
2. $c(1)=0$,
3. c is involutive, i.e. $\forall x \in [0,1], c(c(x))=x$,
4. c is strictly decreasing.

The most obvious example is the one introduced in Section 4.1 as:

$$\forall x \in [0,1],\ c(x) = 1-x. \tag{59}$$

The general form of continuous complementations is:

$$\forall x \in [0,1],\ c(x) = \varphi^{-1}[1-\varphi(x)], \tag{60}$$

where φ is any function from $[0,1]$ into $[0,1]$ such that:

- $\varphi(0)=0$,
- $\varphi(1)=1$,
- φ is strictly increasing.

If, for some n, φ takes the form:

$$\forall x \in [0,1],\ \varphi(x) = x^n, \tag{61}$$

then the derived complementation is:

$$\forall x \in [0,1],\ c(x) = (1-x^n)^{1/n}. \tag{62}$$

This form becomes more binary when n increases (for $n>1$) or when n decreases (for $n<1$). In the first case, almost all values (but those close to 1) have a complement close to 1, and in the second one, almost all values (but those close to 0) have a complement close to 0.

[7] We draw the reader's attention to the fact that since a possibility distribution and a membership functions have similar mathematical expression and since there exist some links between them, the same operators can apply on both of them. However, they have different meanings and origins, and this should not be under-estimated.

If, for some real a in $]0,1]$, φ takes the form:

$$\forall x \in [0,1],\ \varphi(x) = \frac{ax}{(1-a)x+1}, \tag{63}$$

then the derived complementation is:

$$\forall x \in [0,1],\ c(x) = \frac{1-x}{1+a^2x}. \tag{64}$$

Another example, depending on four parameters a, b and c such that $0 \leq a < b < c \leq 1$, and n, is:

$$\forall x \in [0,1],\ c(x) = \begin{cases} 1 & if \quad 0 \leq x \leq a \\ 1 - \frac{1}{2}[\frac{x-a}{b-a}]^n & if \quad a \leq x \leq b \\ \frac{1}{2}[\frac{c-x}{c-b}]^n & if \quad b \leq x \leq c \\ 0 & if \quad c \leq x \leq 1 \end{cases} \tag{65}$$

Some of these examples are illustrated in Figure 2.

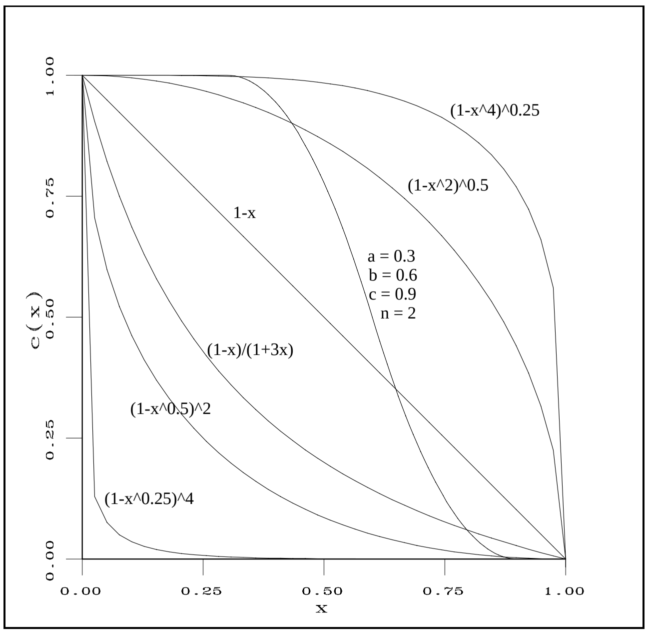


Figure 2: Several examples of fuzzy complementations.

4.4.2 Triangular norms and conorms

In the context of stochastic geometry ([127, 156]), a t-norm t is defined as a function of two variables from $[0,1] \times [0,1]$ to $[0,1]$ satisfying the following properties:

1. commutativity, i.e. $\forall (x,y) \in [0,1]^2,\ t(x,y) = t(y,x)$;
2. associativity, i.e. $\forall (x,y,z) \in [0,1]^3,\ t[t(x,y),z] = t[x,t(y,z)]$;
3. 1 is unit element, i.e. $\forall x \in [0,1],\ t(x,1) = t(1,x) = x$;
4. increasingness with respect to the two variables:

$$\forall (x,x',y,y') \in [0,1]^4,\ (x \leq x' \text{ and } y \leq y') \Rightarrow t(x,y) \leq t(x',y').$$

From these properties, limit conditions can be derived: $t(0,1) = t(0,0) = t(1,0) = 0$ and $t(1,1) = 1$, and it can be easily shown that 0 is null element ($\forall x \in [0,1], t(x,0) = 0$).

A continuity property is often added to these properties.

T-norms generalize intersection to fuzzy sets, as well as logical "and". Examples of t-norms are $\min(x,y)$, xy, $\max(0, x+y-1)$.

It is easy to prove the following result: for any t-norm t, the following inequality holds:

$$\forall (x,y) \in [0,1]^2, t(x,y) \leq \min(x,y). \tag{66}$$

This shows that the "min" is the greatest t-norm and that any t-norm has a conjunctive behavior (an operator is said conjunctive if the result of the combination is less than each of the combined values).

On the opposite, any t-norm is always greater than t_0, which is the smallest t-norm, defined as:

$$\forall (x,y) \in [0,1]^2,\ t_0(x,y) = \begin{cases} x & if \quad y = 1 \\ y & if \quad x = 1 \\ 0 & else \end{cases} \tag{67}$$

Moreover, the following inequalities hold between the mentioned t-norms:

$$\forall (x,y) \in [0,1]^2,\ t_0(x,y) \leq \max(0, x+y-1) \leq xy \leq \min(x,y). \tag{68}$$

Some parametric functions allow to vary between some of these common operators. For instance the t-norm defined in [179] as:

$$\forall (x,y) \in [0,1]^2,\ t(x,y) = 1 - \min[1, [(1-x)^p + (1-y)^p]^{1/p}] \tag{69}$$

varies from 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 3.

Given a t-norm t and a complementation c, another operator T can be defined by duality, called t-conorm:

$$\forall (x,y) \in [0,1]^2,\ T(x,y) = c[t(c(x), c(u)]. \tag{70}$$

A t-conorm satisfies following properties:

1. commutativity, i.e. $\forall (x,y) \in [0,1]^2,\ T(x,y) = T(y,x)$;
2. associativity, i.e. $\forall (x,y,z) \in [0,1]^3,\ T[T(x,y),z] = T[x, T(y,z)]$;
3. 0 is unit element, i.e. $\forall x \in [0,1],\ T(x,0) = T(0,x) = x$;
4. increasingness with respect to the two variables:
$$\forall (x,x',y,y') \in [0,1]^4,\ (x \leq x' \text{ and } y \leq y') \Rightarrow T(x,y) \leq T(x',y');$$
5. limit conditions: $T(0,1) = T(1,1) = T(1,0) = 1$ and $T(0,0) = 0$;
6. 1 is null element ($\forall x \in [0,1], T(x,1) = 1$).

T-conorms generalize union to fuzzy sets, as well as logical "or". Examples of t-conorms are $\max(x,y)$, $x + y - xy$, $\min(1, x+y)$. For any t-conorm T, the following inequality holds:

$$\forall (x,y) \in [0,1]^2, T(x,y) \geq \max(x,y). \tag{71}$$

This shows that the "max" is the smallest t-conorm and that any t-conorm has a disjunctive behavior (an operator is said disjunctive if the result of the combination is greater than each of the combined values).

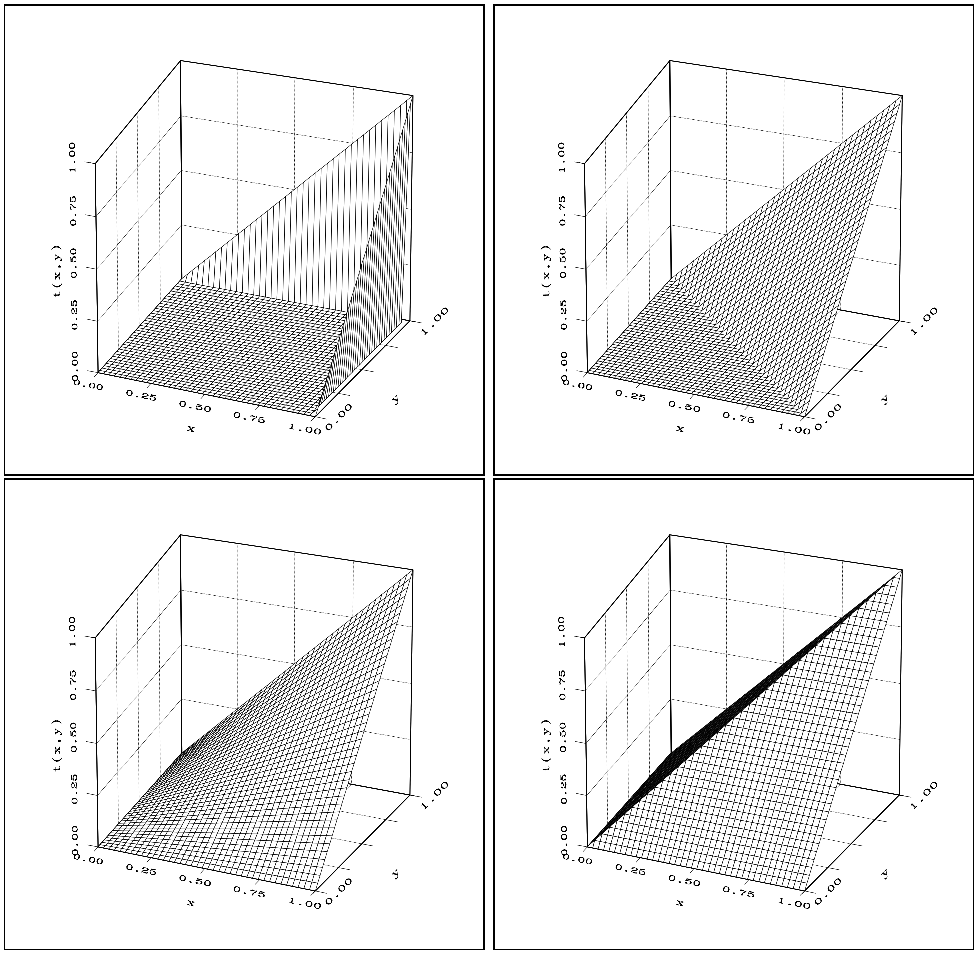


Figure 3: Four examples of t-norms. Upper row: t_0 (smallest t-norm) and Lukasiewicz t-norm. Lower row: product and minimum (largest t-norm).

On the opposite, any t-conorm is always less than T_0, which is the largest t-conorm, defined as:

$$\forall(x,y) \in [0,1]^2,\ T_0(x,y) = \begin{cases} x & if \quad y=0 \\ y & if \quad x=0 \\ 1 & else \end{cases} \tag{72}$$

Moreover, the following inequalities hold between the mentioned t-conorms:

$$\forall(x,y) \in [0,1]^2,\ T_0(x,y) \geq \min(1, x+y) \geq x+y-xy \geq \max(x,y). \tag{73}$$

Examples of t-conorms are shown in Figure 4.

Further useful properties of t-norms and t-conorms are:

- any t-norm or t-conorm is distributive with respect to min and max, i.e. equalities of the following type hold:
$$\forall(x,y,z) \in [0,1]^3,\ t[x, \min(y,z)] = \min[t(x,y), t(x,z)]; \tag{74}$$
- the only t-norms and t-conorms that are mutually distributive are min and max;
- the only t-norm that is idempotent is the min, and the only t-conorm that is idempotent is the max;

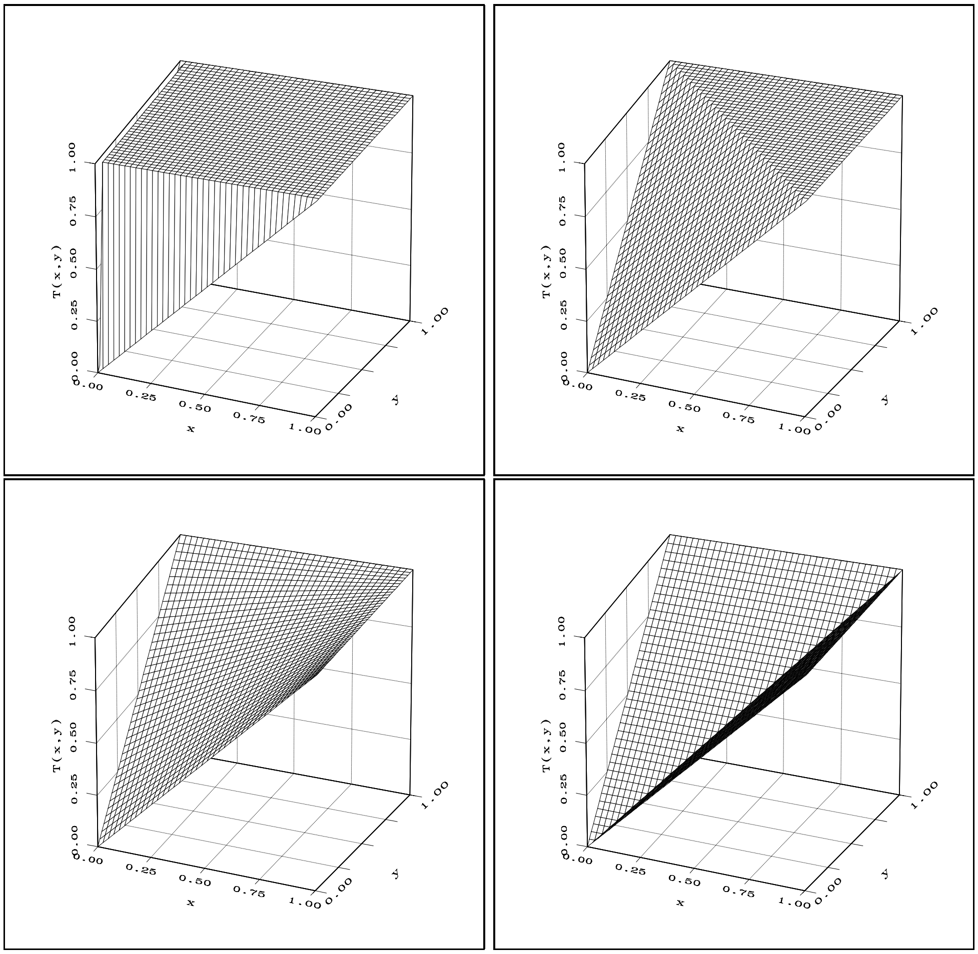


Figure 4: Four examples of t-conorms. Upper row: T_0 (largest t-conorm) and Lukasiewicz t-conorm. Lower row: algebraic sum and maximum (smallest t-norm).

- given any t-norm t and any continuous strictly increasing function h from $[0,1]$ into $[0,1]$ such that $h(0) = 0$ and $h(1) = 1$, the following function t' is a t-norm [155]:

$$\forall (x,y) \in [0,1]^2,\ t'(x,y) = h^{-1}[t(h(x), h(y))]. \tag{75}$$

  This shows how families of t-norms can be generated from a given one.

Several works have been dedicated to the precise description of specific families of t-norms and t-conorms, for which generic definitions could be found [70]. Two of them are particularly useful: Archimedian t-norms and nilpotent t-norms.

An Archimedian strictly monotonous t-norm is a t-norm t such that:

$$\forall x \in [0,1],\ t(x,x) < x, \tag{76}$$

and

$$\forall (x,y,y') \in [0,1]^3,\ y < y' \Rightarrow t(x,y) < t(x,y'). \tag{77}$$

Similarly an Archimedian strictly monotonous T-conorm T satisfies the two following properties:

$$\forall x \in [0,1],\ T(x,x) > x, \tag{78}$$

$$\forall (x,y,y') \in [0,1]^3,\ y < y' \Rightarrow T(x,y) < T(x,y'). \tag{79}$$

Any Archimedian strictly monotonous t-norm t can be expressed in the following form:

$$\forall(x,y) \in [0,1]^2, t(x,y) = f^{-1}[f(x) + f(y)], \tag{80}$$

where f, referred to as "generating function", is a continuous decreasing bijection from $[0,1]$ into $[0,+\infty]$ such that $f(0) = +\infty$ and $f(1) = 0$.

The associated t-conorms take the form:

$$\forall(x,y) \in [0,1]^2, T(x,y) = \varphi^{-1}[\varphi(x) + \varphi(y)], \tag{81}$$

where the generating function φ is a continuous increasing bijection from $[0,1]$ into $[0,+\infty]$ such that $\varphi(0) = 0$ and $\varphi(1) = +\infty$.

Such t-norms and t-conorms never satisfy the non-contradiction and excluded-middle laws. These laws are expressed as:

$$\forall x \in [0,1],\ t[x, c(x)] = 0, \tag{82}$$

and:

$$\forall x \in [0,1],\ T[x, c(x)] = 1. \tag{83}$$

These two statements do not hold for Archimedian strictly monotonous t-norms and t-conorms.

An Archimedian strictly monotonous t-norm (respectively t-conorm) can be defined by a multiplicative generating function as well, both expressions (the additive one and the multiplicative one) being equivalent [54]. It is then expressed as:

$$\forall(x,y) \in [0,1]^2,\ t(x,y) = h^{-1}[h(x)h(y)], \tag{84}$$

where h is a strictly increasing function from $[0,1]$ into $[0,1]$ such that $h(0) = 0$ and $h(1) = 1$. The equivalence with the additive form is simply obtained by setting:

$$h = e^{-f} \tag{85}$$

where f is the additive generating function introduced previously.

The most used t-norms and t-conorms of this class are the product and the algebraic sum:

$$\forall(x,y) \in [0,1]^2,\ t(x,y) = xy, \quad T(x,y) = x + y - xy. \tag{86}$$

The only rational t-norms of this class are the Hamacher's t-norms, defined as [96]:

$$\forall(x,y) \in [0,1]^2,\ \frac{xy}{\gamma + (1-\gamma)(x+y-xy)}, \tag{87}$$

where γ is a positive parameter (for $\gamma = 1$ the operator coincides with the t-norm product). They are illustrated in Figure 5.

Another parametric family of this class is made of Frank's functions, defined as [82]:

$$\forall(x,y) \in [0,1]^2,\ t(x,y) = \log_s[1 + \frac{(s^x - 1)(s^y - 1)}{s - 1}], \tag{88}$$

where s is a strictly positive parameter. These t-norms and their dual t-conorms satisfy the following remarquable equality (and they are the only t-norms and t-conorms satisfying this relation):

$$\forall(x,y) \in [0,1]^2,\ t(x,y) + T(x,y) = x + y. \tag{89}$$

Examples of Frank's t-norms are shown in Figure 6. If s is small and tends towards 0, the t-norm tends towards the minimum. If s tends towards $+\infty$, the t-norm tends towards the Lukasiewicz t-norm. If $s = 1$, the t-norm is equal to the product.

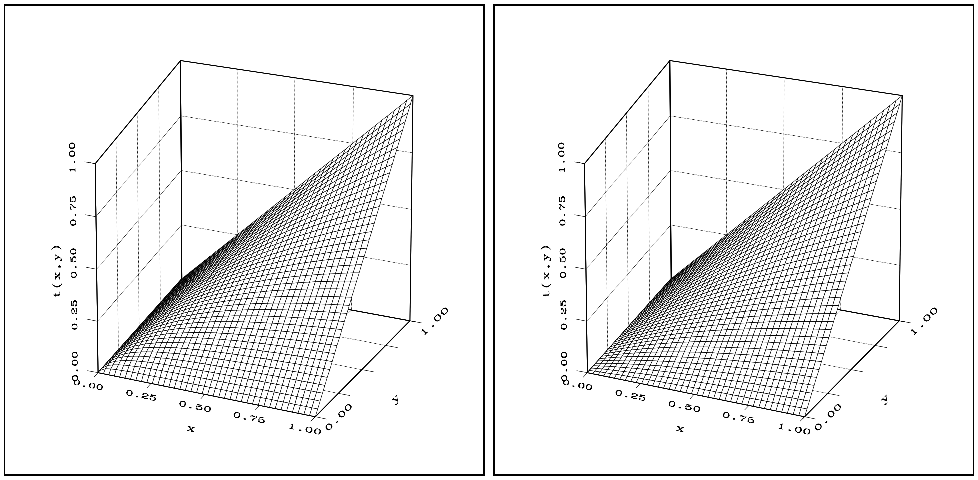


Figure 5: Two examples of Hamacher's t-norms, for $\gamma = 0$ (left) and $\gamma = 0.4$ (right).

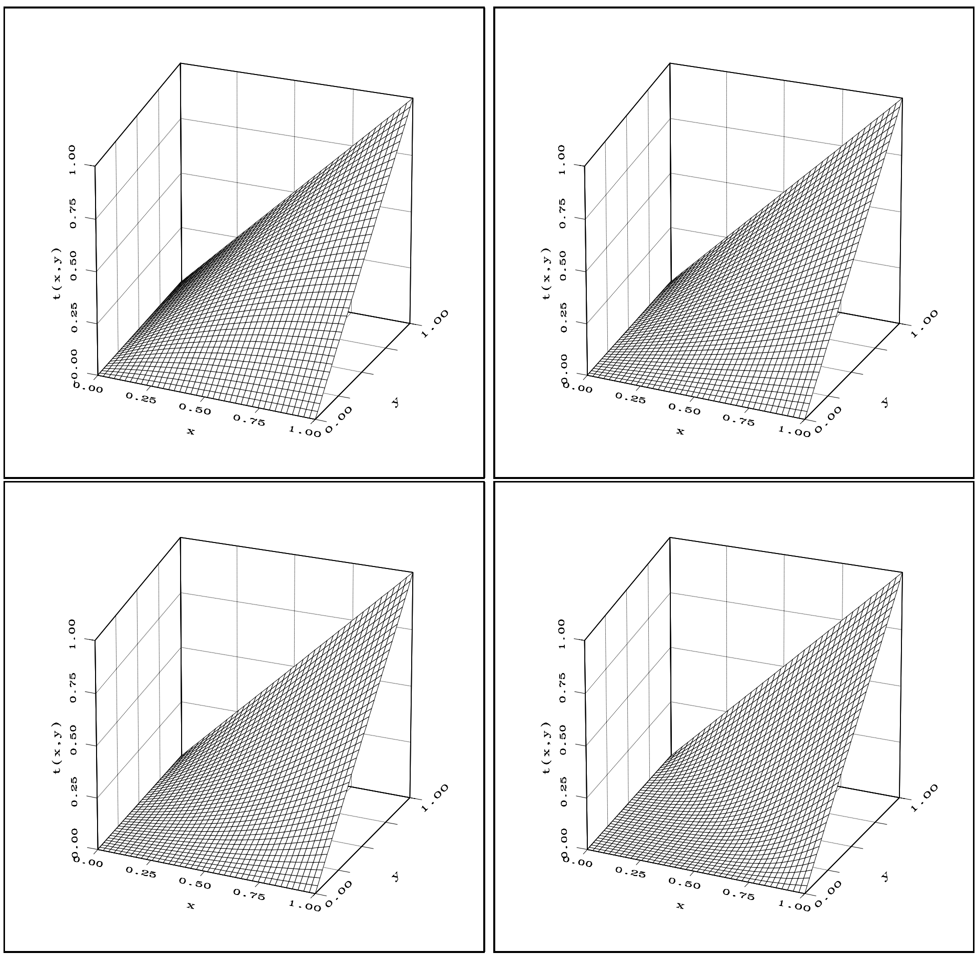


Figure 6: Four examples of Frank's t-norms. Upper row: $s = 0.1$ and $s = 2$. Lower row: $s = 10$ and $s = 1000$.

Another useful family of t-norms and t-conorms si constituted by the nilpotent ones, taking the following general form:

$$\forall (x, y) \in [0,1]^2,\ t(x,y) = f^*[f(x) + f(y)], \tag{90}$$

where f is a decreasing bijection from [0,1] into [0,1], such that $f(0) = 1$, $f(1) = 0$, and $f^*(x) = f^{-1}(x)$ if $x \in [0, 1]$. $f^*(x) = 0$ if $x \geq 1$. The general form of nilpotent t-conorms can be derived by duality. They satisfy the excluded-middle and non-contradiction laws.

The most used t-norm and t-conorm of this class are the Lukasiewicz operators:

$$\forall (x, y) \in [0, 1]^2,\ t(x, y) = \max(0, x + y - 1), \quad T(x, y) = \min(1, x + y). \tag{91}$$

Examples of generating f functions have been proposed e.g. by Schweizer and Sklar [155] or by Yager [179].

Finally, combinations of t-norms and t-conorms are also useful. For instance, compensation operators have been introduced in [192] and take the form:

$$\forall (x, y) \in [0, 1]^2,\ C_\gamma(x, y) = t(x, y)^{1-\gamma} T(x, y)^\gamma, \tag{92}$$

where γ is a parameter in $[0, 1]$.

4.4.3 Mean operators

A mean operator is defined as a function m from $[0, 1] \times [0, 1]$ into $[0, 1]$ such that:

1. $m \neq \min$, $m \neq max$,
2. the result of the combination is a compromise between the smallest and the largest value: $\forall (x, y) \in [0, 1]^2,\ \min(x, y) \leq m(x, y) \leq \max(x, y)$,
3. m is commutative: $\forall (x, y) \in [0, 1]^2,\ m(x, y) = m(y, x)$,
4. m is increasing with respect to both arguments:

$$\forall (x, x', y, y') \in [0, 1]^4,\ (x \leq x' \text{ and } y \leq y') \Rightarrow m(x, y) \leq m(x', y').$$

A consequence of this definition (and more precisely of the second property) is that any mean operator m is idempotent, i.e.:

$$\forall x \in [0, 1],\ m(x, x) = x.$$

Note that associativity is generally not satisfied by mean operators. The only associative mean operators are the median operators defined as:

$$\forall (x, y) \in [0, 1]^2,\ m(x, y) = med(x, y, \alpha) = \begin{cases} x & if \quad y \leq x \leq \alpha \ or \ \alpha \leq x \leq y \\ y & if \quad x \leq y \leq \alpha \ or \ \alpha \leq y \leq x \\ \alpha & if \quad y \leq \alpha \leq x \ or \ x \leq \alpha \leq y \end{cases} \tag{93}$$

where α is a parameter in $[0, 1]$.

Examples of median operators are shown in Figure 7.

Several mean operators satisfy an additional property, called bisymmetry, that can be considered as a counter-part of associativity:

$$\forall (x, y, z, t) \in [0, 1]^4,\ m[m(x, y), m(z, t)] = m[m(x, z), m(y, t)]. \tag{94}$$

The general form of bisymmetrical, continuous and strictly increasing mean operators is as follows:

$$\forall (x, y) \in [0, 1]^2,\ m(x, y) k^{-1}[\frac{k(x) + k(y)}{2}], \tag{95}$$

where k is a continuous and strictly monotonous function from $[0, 1]$ into $[0, 1]$.

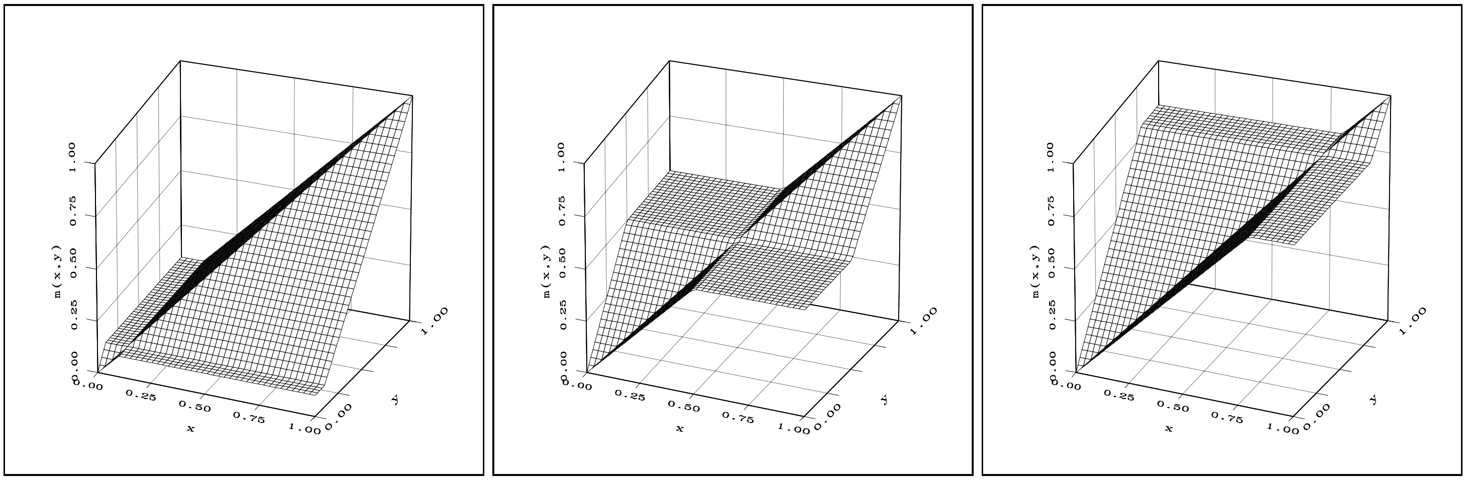


Figure 7: Three examples of median operators for α equal to 0.1, 0.5 and 0.8.

α	$m(x,y)$	comment
$-\infty$	$\min(x,y)$	limit value
-1	$\frac{2xy}{x+y}$	harmonic mean
0	$(xy)^{-1/2}$	geometrical mean
$+1$	$\frac{x+y}{2}$	arithmetical mean
$+2$	$\sqrt{\frac{x^2+y^2}{2}}$	quadratic mean
$+\infty$	$\max(x,y)$	limit value

Table 1: Examples of bisymmetrical, continuous and strictly increasing mean operators. For the harmonic mean, we use the convention $m(0,0) = 0$.

Classical mean operators are found in this class, for k defined as:

$$\forall x \in [0,1], k(x) = x^\alpha,$$

where $\alpha \in \mathbb{R}$. In particular, harmonic, geometrical, arithmetical, and quadratic means are obtained for α equal to -1, 0, 1 and 2 respectively. Table 1 summarizes these results.

Examples of mean operators are shown in Figure 8.

Other mean operators involve weights. This amounts to take the values to be combined into account at different levels. One particularly interesting weighted operator has been proposed by Yager [180], as an ordered weight average operator (OWA). Weights are defined according to the ranking of the values to be combined. If these values are denoted by $a_1, a_2, ...a_n$, they are ordered in a sequence $a_{j_1}, a_{j_2}, ...a_{j_n}$ such that:

$$a_{j_1} \leq a_{j_2} \leq ... \leq a_{j_n}.$$

Then, for a set of weights w_i such that:

$$\sum_{i=1}^{n} w_i = 1, \ \ \forall i, 1 \leq i \leq n, w_i \in [0,1],$$

the OWA operator is defined by the expression:

$$OWA(a_1, a_2, ..., a_n) = \sum_{i=1}^{n} w_i a_{j_i}. \tag{96}$$

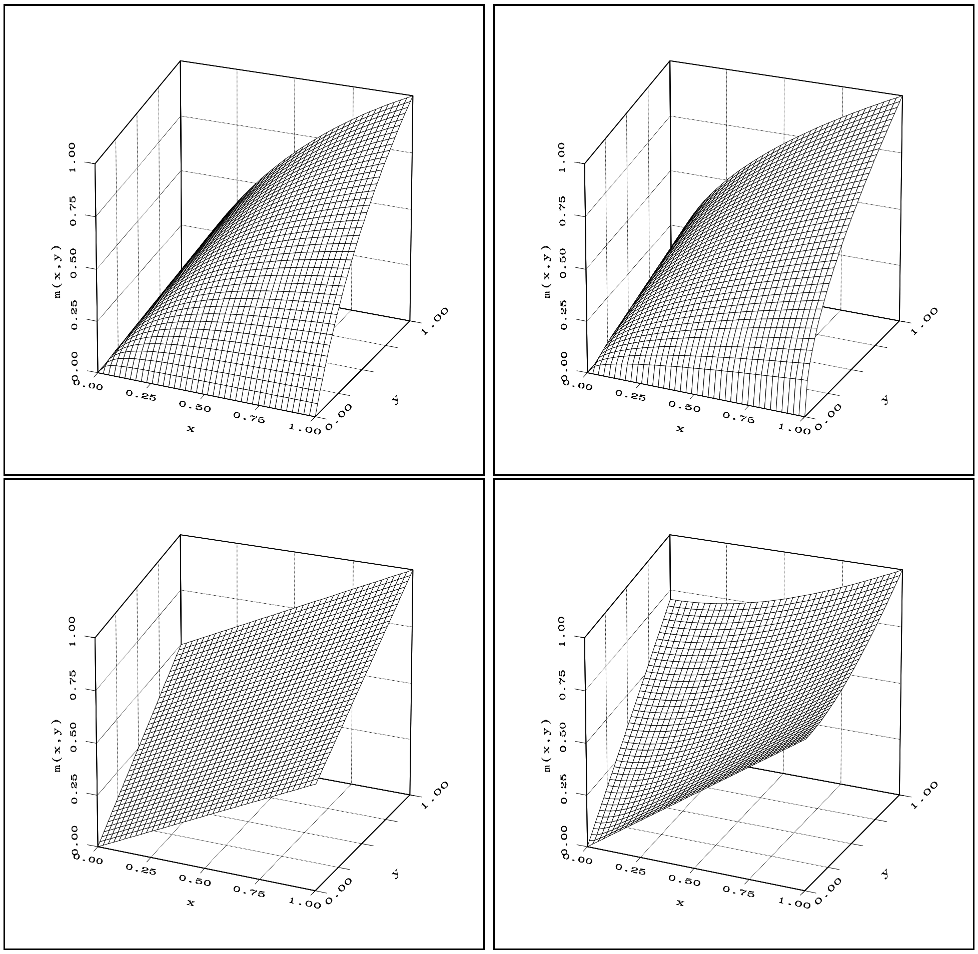


Figure 8: Four examples of mean operators. Upper row: harmonic and geometrical mean operators. Lower row: arithmetical and quadratic mean operators.

Fuzzy integrals also belong to the class of mean operators [94]. Indeed, both Choquet's and Sugeno's integrals are idempotent, continuous, increasing, and between the minimum and the maximum. They include as particular cases order statistics, and therefore minimum, maximum, and median. Choquet's integral defined with respect to an additive measure μ is equivalent to a weighted arithmetical mean, where the weight w_i affected to the value x_i is equal to $\mu(\{x_i\})$.

The OWA operators can also be seen as a particular class of Choquet's fuzzy integrals, where the fuzzy measure is defined as:

$$\forall A, |A| = i, \;\; \mu(A) = \sum_{j=0}^{i-1} w_{n-j}.$$

Conversely, any commutative Choquet's integral is such that $\mu(A)$ only depends on $|A|$ and is equal to an OWA operator the weights of which are:

$$w_1 = 1 - \sum_{i=2}^{n} w_i,$$

$$\forall i \geq 2, \; w_i = \mu(A_{n-i+1}) - \mu(A_{n-i}),$$

where A_i denotes any subset such that $|A_i| = i$.

These properties among other have been extensively studied in [95, 94].

4.4.4 Symmetrical sums

A symmetrical sum is defined as an operator σ from $[0,1] \times [0,1]$ into $[0,1]$ such that:

1. $\sigma(0,0) = 0$,
2. σ is commutative,
3. σ is increasing with respect to both arguments,
4. σ is continuous,
5. $\forall (x,y) \in [0,1]^2, 1 - \sigma(x,y) = \sigma(1-x, 1-y)$.

The last property means that the scale of values can be reversed without changing the conclusions that may be drawn from their combination. This has been used for instance for expert opinion pooling, where saying that 0 means "bad" and 1 means "good" or saying the reverse should not be important. *citer qqch de Dubois ici.* This property can also be expressed using other fuzzy complementations.

Other property hold, that can be deduced from the others:

- $\sigma(1,1) = 1$,
- $\forall x \in]0,1[, \sigma(x, 1-x) = \frac{1}{2}$,
- the only symmetrical sum that is both associative and a mean operator is the median with parameter $\frac{1}{2}$.

The general form of symmetrical sums depends on an increasing, positive, continuous function g such that $g(0,0) = 0$:

$$\forall (x,y) \in [0,1]^2, \sigma(x,y) = \frac{g(x,y)}{g(x,y) + g(1-x, 1-y)}. \tag{97}$$

It is easy to verify that this form satisfies all properties of a symmetrical sum.

If $\forall x \in [0,1], g(0,x) = 0$, $\sigma(0,1)$ is undefined, otherwise $\sigma(0,1) = \frac{1}{2}$.

The general form of associative, strictly increasing symmetrical sums is:

$$\forall (x,y) \in [0,1]^2, \sigma(x,y) = \psi^{-1}[\psi(x) + \psi(y)], \tag{98}$$

where ψ is a strictly monotonous function such that $\psi(0)$ and $\psi(1)$ are non-bounded and $\forall x \in [0,1], \psi(1-x) + \psi(x) = 0$. It follows that the values 0 and 1 are null elements, while $\frac{1}{2}$ is the unit element.

Typical examples of symmetrical sums are given in Table 2. They are obtained by taking t-norms or t-conorms for the generating function g.

$g(x,y)$	$\sigma(x,y)$	property
xy	$\sigma_0(x,y) = \frac{xy}{1-x-y+2xy}$	associative
$x+y-xy$	$\sigma_+(x,y) = \frac{x+y-xy}{1+x+y-2xy}$	not associative
$\min(x,y)$	$\sigma_{\min}(x,y) = \frac{\min(x,y)}{1-\lvert x-y \rvert}$	mean operator
$\max(x,y)$	$\sigma_{\max}(x,y) = \frac{\max(x,y)}{1+\lvert x-y \rvert}$	mean operator

Table 2: Examples of symmetrical sums, defined using t-norms and t-conorms as generating functions. For σ_0, we use the convention $\sigma_0(0,1) = \sigma_0(1,0) = 0$, and for $\sigma_{\min}$, we set $\sigma_{\min}(0,1) = \sigma_{\min}(1,0) = 0$.

These operators satisfy the following order property:

$$\forall (x,y) \in [0,1]^2, x + y \leq 1 \Rightarrow \sigma_0(x,y) \leq \sigma_{\min}(x,y) \leq \sigma_{\max}(x,y) \leq \sigma_+(x,y), \tag{99}$$

$$\forall(x,y)\in[0,1]^2, x+y\geq 1 \Rightarrow \sigma_0(x,y)\geq \sigma_{\min}(x,y)\geq \sigma_{\max}(x,y)\geq \sigma_+(x,y). \tag{100}$$

Examples of symmetrical sums are shown in Figure 9.

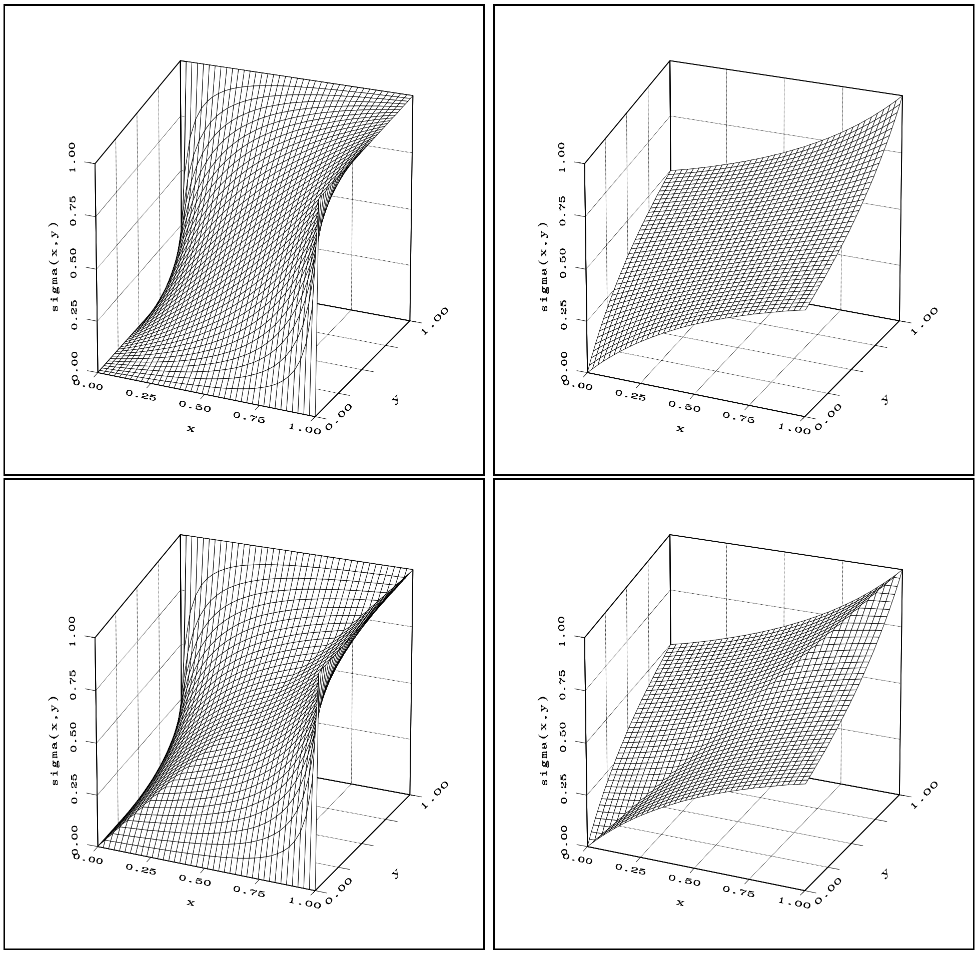


Figure 9: Four examples of symmetrical sums. Upper row: σ_0 and σ_+. Lower row: $\sigma_{\min}$ and $\sigma_{\max}$.

4.4.5 Adaptive operators

A lot of other operators have been proposed in the literature. Among them, operators that are adaptive with respect to some contextual information are of particular interest. This contextual information can be for instance a conflict between possibility distributions (e.g. extracted from different sources of information), or a reliability attached to each of the pieces of information to be combined.

We give some examples of operators depending on conflict between possibility distributions, according to [74]. In the following equations, π_1 and π_2 represent two distributions of possibility to be combined in a global distribution π', $1-h(\pi_1,\pi_2)$ represents a global measure of conflict between these two distributions, and t denotes a t-norm:

$$\pi'(s)=\max\left[\frac{t[\pi_1(s),\pi_2(s)]}{h(\pi_1,\pi_2)}, 1-h(\pi_1,\pi_2)\right], \tag{101}$$

$$\pi'(s)=\min\left[1, \frac{t[\pi_1(s),\pi_2(s)]}{h(\pi_1,\pi_2)}+1-h(\pi_1,\pi_2)\right], \tag{102}$$

$$\pi'(s)=t[\pi_1(s),\pi_2(s)]+1-h(\pi_1,\pi_2), \tag{103}$$

$$\pi'(s) = \max\left[\frac{\min(\pi_1,\pi_2)}{h}, \min[\max(\pi_1,\pi_2), 1-h]\right]. \tag{104}$$

In [74], the proposed measure of conflict is defined as $1-h$ with:

$$h = \sup_s t[\pi_1(s), \pi_2(s)]. \tag{105}$$

This measure, which assures that the resulting distribution π' is normalized, is well adapted for trapeze-shaped possibility distributions.

The last form (Equation 104) is illustrated in Figure 10 for two possibility distributions with increasing conflict.

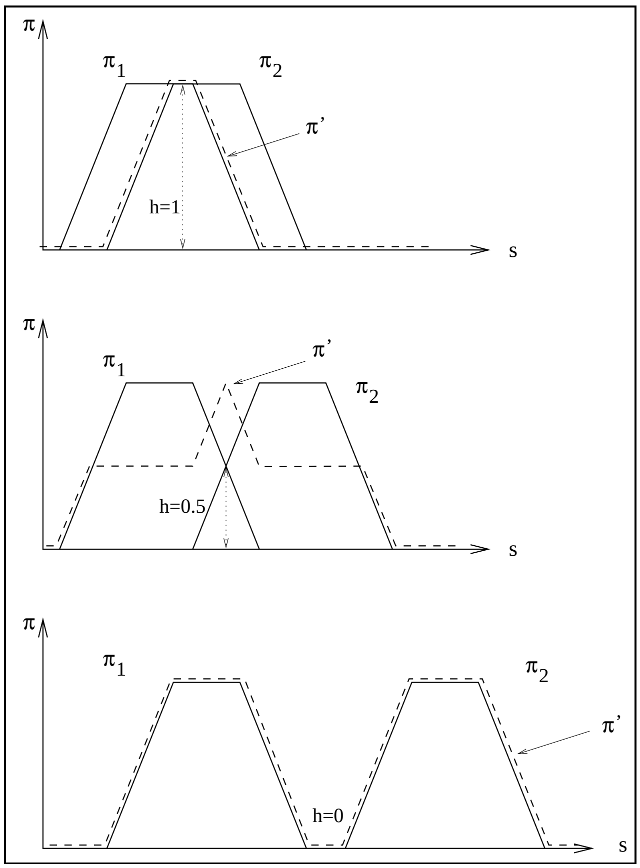


Figure 10: Example of adaptive operator varying from a t-norm for no conflict to a t-conorm for a high conflict between two possibility distributions.

4.5 Linguistic variables

It may happen that numerical representations are not adequate to describe one situation. For instance, if a variable has a large range of variation, we can hardly attach one value to any specific situation and we may prefer to use a qualifier issued from the natural language to group coarsely some typical subsets of interest. For instance, to describe the size of an object, it may be more convenient to use only a few terms having rough frontiers, like large, medium, or small. This corresponds to a granulation of the information. According to [190], the concept of granule is the starting point in "computing with words", and it is defined as "a fuzzy set of points having the form of a clump of elements drawn together by similarity" (quoted from [190]). A word is then a label of a granule. When it comes to compute with these representations, specific tools are needed. Such representations are particularly useful in approximate reasoning.

Such representations are called linguistic variables. They are variables whose values are words or sentences [188].

The advantage of such representations is that linguistic characterizations may be less specific than numerical ones (and therefore need less information) [188].

4.5.1 Definition

The concept of linguistic variable has been introduced by Zadeh [188] and is described by several authors, e.g. [67, 191]. Formally, a linguistic variable is defined as a quintuple $(x, T(x), \mathcal{U}, G, M)$, where x is the name of the variable, $T(x)$ the set of values of x (called "terms"), $\mathcal{U}$ is the universe of discourse on which the values of x are defined (as fuzzy variables), G is a syntactic rule for generating the name X of values of x, and M is a semantic rule, $M(X)$ being a fuzzy set of $\mathcal{U}$ representing the meaning of X.

Such a definition represents a symbolic-numerical conversion, and establishes links between a language and numerical scales.

4.5.2 Example of linguistic variable

Let us consider the example of the size of an object. In numerical terms, this size can be expressed by a value ranging over a domain $\mathcal{U}$ (typically $\mathcal{U}$ is a subset of $\mathbb{R}^+$). In linguistic terms, the size can be expressed with a few terms: very small, small, medium, large, very large. Figure 11 illustrates the linguistic variable "size".

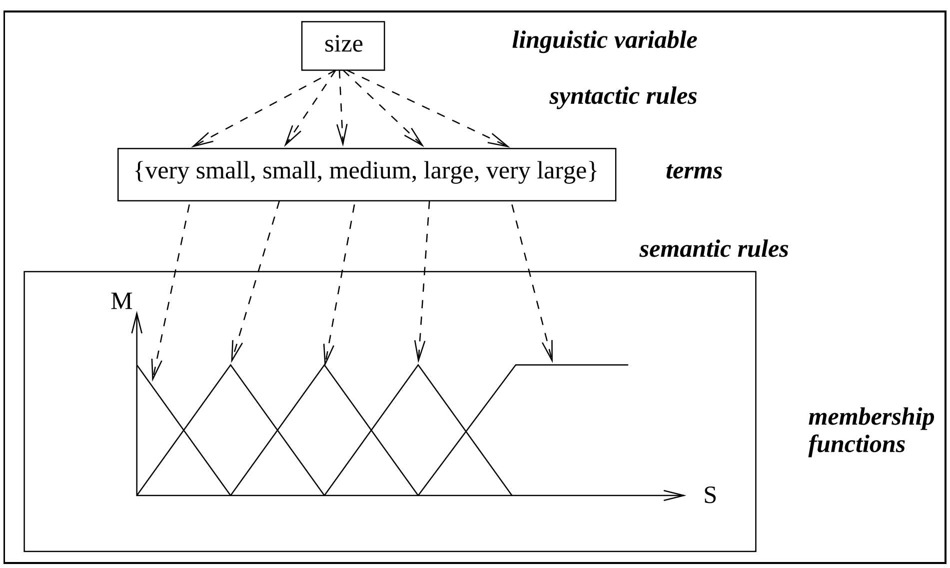


Figure 11: Illustration of the linguistic variable "size", its terms and the associated fuzzy sets. The arrows from the linguistic variable to the set of terms represent the syntactic rules, while the second set of arrows represents the semantic rules, that translate terms in membership functions.

4.5.3 Modifiers

The meaning of a term of a linguistic variable can be modified by operators called "hedges" or "modifiers". If A is a fuzzy set, then the hedge or modifier h generates a composite term $h(A)$ which is a fuzzy set on the same universe of discourse $\mathcal{U}$. The most common operators that are used for defining modifiers are the following:

- normalization:
$$\mu_{norm(A)}(u) = \frac{\mu_A(u)}{\sup_{v \in \mathcal{U}} \mu_A(v)},$$
where μ_A denotes the membership function of A and u any value in $\mathcal{U}$;
- concentration: $\mu_{con(A)}(u) = [\mu_A(u)]^2$;
- dilation[8]: $\mu_{dil(A)}(u) = [\mu_A(u)]^{0.5}$;
- contrast intensification:
$$\mu_{int(A)}(u) = \begin{cases} 2[\mu_A(u)]^2 & if\ \mu_A(u) \in [0, 0.5] \\ 1 - 2[1 - \mu_A(u)]^2 & otherwise. \end{cases}$$

These functions are illustrated in Figure 12, for a simple triangular-shaped membership function.

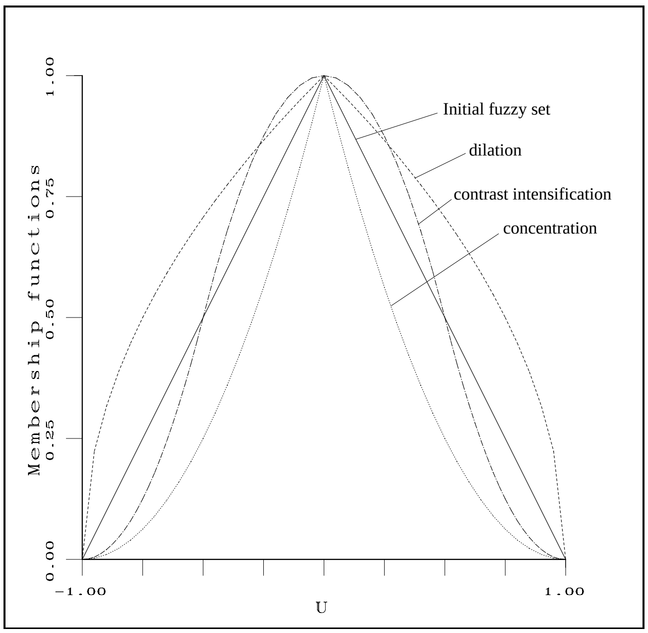


Figure 12: Illustration of some operations used for defining hedges, on a triangular shaped initial fuzzy set. Dotted: concentration, dashed: dilation, dot-dashed: contrast intensification.

Typical hedges derived from these operators are [67]:

- $very\ A = con(A)$,
- $more\ or\ less\ A = dil(A)$,
- $plus\ A = A^{1.25}$,
- $slightly\ A = int[norm(plus\ A\ and\ not(very\ A))]$ where "and" and "not" are defined as a t-norm and a complementation respectively.

[8]Note that it is not a dilation in the morphological sense.

Let us mention, as a nice application of modifiers, the explanation of the Golden Proportion using fuzzy logic, as proposed in [110]. The author that the concept of something being pleasing often depends on an optimum of some variable. For instance, if someone puts a little bit of a nice perfume on, this will increasing the attractiveness. But if the person puts to much perfume, it does not smell good. In this example, and in several others, it appears the the degree x to which something is pleasing follows the following rule: if it increase too much, the opposite effect is obtained. Therefore we have, in linguistic terms:

$$very(x) = not(x)$$

Since $very(x)$ is expressed as x^2 and $not(x)$ as $1 - x$, x happens to be the solution of:

$$x^2 = 1 - x$$

the solution of which being the golden proportion, i.e. $\frac{\sqrt{5}-1}{2}$.

4.6 Fuzzy relations

The concept of relation has been extended to fuzzy sets in [186, 187] to model interactions between some elements. This extension endows fuzzy sets with comparison and structural assessment tools, and is therefore important for image processing, cognitive vision and image interpretation, since it will allow to compare shapes of objects, to establish spatial relationships, etc.

4.6.1 Definitions

Let $\mathcal{U}_1$ and $\mathcal{U}_2$ be two universes. A fuzzy relation R on $\mathcal{U}_1 \times \mathcal{U}_2$ is a fuzzy subset of $\mathcal{U}_1 \times \mathcal{U}_2$:

$$R = \{((u_1, u_2), \mu_R(u_1, u_2)) | (u_1, u_2) \in \mathcal{U}_1 \times \mathcal{U}_2\}. \tag{106}$$

This definition of binary relation extends directly to n-ary relations.

Let μ_1 and μ_2 be two fuzzy sets defined respectively on $\mathcal{U}_1$ and $\mathcal{U}_2$. A fuzzy relation on $\mu_1 \times \mu_2$ is defined as a fuzzy set on $\mathcal{U}_1 \times \mathcal{U}_2$ such that:

$$\forall (u_1, u_2) \in \mathcal{U}_1 \times \mathcal{U}_2, \ \mu_R(u_1, u_2) \leq \min[\mu_1(u_1), \mu_2(u_2)]. \tag{107}$$

The projection of a fuzzy relation R on $\mathcal{U}_1$ is defined as a fuzzy subset of $\mathcal{U}_1$ as:

$$proj[R, \mathcal{U}_1] = \{(u_1, \sup_{u_2 \in \mathcal{U}_2} \mu_R(u_1, u_2)) | u_1 \in \mathcal{U}_1\}. \tag{108}$$

The projection on $\mathcal{U}_2$ is defined in a similar way.

In case of n-ary relation, defined on $\mathcal{U}_1 \times ... \times \mathcal{U}_n$, the projection can be defined on any product space of some of the $\mathcal{U}_i$, as a fuzzy relation on this product space.

The cylindrical extension of a fuzzy relation on a product space $\mathcal{U}_{i_1} \times ... \times \mathcal{U}_{i_k}$, where $\{i_1 ... i_k\} \subset \{1...n\}$, is defined as the "largest" fuzzy relation on $\mathcal{U}_1 \times ... \times \mathcal{U}_n$:

$$c(R) = \{((u_1, ..., u_n), \mu_R(u_{i_1}, ..., u_{i_k})) | (u_1, ..., u_n) \in \mathcal{U}_1 \times ... \times \mathcal{U}_n\}. \tag{109}$$

A fuzzy relation is called fuzzy restriction if it represents an elastic constraint on the values (in $\mathcal{U}_1 \times \mathcal{U}_2$) that can be assigned to a variable v defined on $\mathcal{U}_1 \times \mathcal{U}_2$.

4.6.2 Properties of fuzzy relations

The extension of relation properties to fuzzy relations leads generally to several possible definitions. We mention here the most common ones, for a binary relation in $\mathcal{U} \times \mathcal{U}$.

Reflexivity A fuzzy relation R in $\mathcal{U} \times \mathcal{U}$ is reflexive if [187]:

$$\forall x \in \mathcal{U},\ \mu_R(x,x) = 1. \tag{110}$$

It is ε-reflexive if [183]:

$$\forall x \in \mathcal{U},\ \mu_R(x,x) \geq \varepsilon. \tag{111}$$

It is weakly reflexive if [183]:

$$\forall x \in \mathcal{U}, \forall y \in \mathcal{U},\ \mu_R(x,x) \geq \mu_R(x,y). \tag{112}$$

Symmetry and anti-symmetry A fuzzy relation R in $\mathcal{U} \times \mathcal{U}$ is symmetric if:

$$\forall x \in \mathcal{U}, \forall y \in \mathcal{U},\ \mu_R(x,y) = \mu_R(y,x). \tag{113}$$

A fuzzy relation R in $\mathcal{U} \times \mathcal{U}$ is anti-symmetric [103] if:

$$\forall x \in \mathcal{U}, \forall y \in \mathcal{U},\ x \neq y \Rightarrow \begin{cases} either & \mu_R(x,y) \neq \mu_R(y,x) \\ or & \mu_R(x,y) = \mu_R(y,x) = 0 \end{cases} \tag{114}$$

A fuzzy relation R in $\mathcal{U} \times \mathcal{U}$ is perfectly anti-symmetric [187] if:

$$\forall x \in \mathcal{U}, \forall y \in \mathcal{U},\ x \neq y \ and \ \mu_R(x,y) > 0 \Rightarrow \mu_R(y,x) = 0. \tag{115}$$

The relationships between both definitions of anti-symmetry is as follows: perfect anti-symmetry implies anti-symmetry.

Transitivity Transitivity usually expresses that the relation between two elements should be at least as strong as indirect relations between these elements (i.e. going through other elements).

This idea is translated in fuzzy terms as follows: a fuzzy relation R in $\mathcal{U} \times \mathcal{U}$ is max-min transitive if:

$$\forall (x,y,z) \in \mathcal{U}^3,\ \mu_R(x,z) \geq \min[\mu_R(x,y), \mu_R(y,z)]. \tag{116}$$

This definition has a more general form, called max-$*$ transitivity, where $x * y$ is defined as one of the following expressions:

1. xy,
2. $\max(0, x+y-1)$,
3. $\frac{x+y}{2}$,
4. $\max(x,y)$,
5. $x+y-xy$.

4.6.3 Composition of relations

The most common way to compose two fuzzy relations R and S defined respectively on $\mathcal{U}_1 \times \mathcal{U}_2$ and $\mathcal{U}_2 \times \mathcal{U}_3$ is called the max-min composition, and is defined as:

$$\forall u_1 \in \mathcal{U}_1, \forall u_3 \in \mathcal{U}_3,\ \mu_{R \circ S}(u_1,u_3) = \sup_{u_2 \in \mathcal{U}_2} \min[\mu_R(u_1,u_2), \mu_S(u_2,u_3)]. \tag{117}$$

The composition $R \circ S$ is a fuzzy relation on $\mathcal{U}_1 \times \mathcal{U}_3$.

Instead of the min, another operator can be used, leading to the max-$*$ composition. The operator $*$ can be for instance the product or the arithmetic mean [144].

If the operator $*$ is associative and increasing with respect to both arguments, then the max-$*$ composition is associative, distributive over union and increasing with respect to inclusion.

Since the composition is obtained by considering the "strength" of all paths between u_1 and u_3, there is some link with the max-min transitivity: a fuzzy relation R on $\mathcal{U} \times \mathcal{U}$ is max-min transitive iff $R \circ R \subset R$.

A similar equivalence holds between max-$*$ transitivity and max-$*$ composition.

4.6.4 Similarity relations

Similarity relations constitute a particular class of fuzzy relations, particularly useful for pattern recognition, comparison between objects, between relationships, between elements and models, for case-based reasoning, for recognition based on models or prototypes, etc.

A similarity relation is a fuzzy relation S on $\mathcal{U}$ which is [187]:

1. reflexive,
2. symmetrical,
3. max-min transitive.

Reflexivity can be too restrictive, and can then be replaced by ε-reflexivity, or weak reflexivity. Other variants use max-$*$ transitivity, instead of max-min transitivity.

Since transitivity may be difficult to achieve, a useful fuzzy relation is "tolerance" or "proximity" relation, which is only reflexive and symmetrical.

From a similarity relation S, a dissimilarity relation D can be derived, as:

$$\forall (x, y) \in \mathcal{U}^2, \; \mu_D(x, y) = 1 - \mu_S(x, y). \tag{118}$$

A dissimilarity relation satisfies the following properties:

1. anti-reflexivity:
$$\forall x \in \mathcal{U}, \; \mu_D(x, x) = 0,$$
2. symmetry,
3. min-max transitivity:
$$\forall (x, y, z) \in \mathcal{U}^3, \; \mu_D(x, z) \leq \max[\mu_D(x, y), \mu_D(y, z)].$$

The membership function $\mu_D(x, y)$ represents a distance function on $\mathcal{U}$, which is moreover an ultrametric, thanks to its last property.

Zadeh has shown [187] that a similarity relation S can be represented as a tree, called partition-tree or similarity-tree. Each level of the tree corresponds to the partition of $\mathcal{U}$ induced by an α-cut of S. If Π_α denotes the partition induced by S_α, then Π_β is a refinement of Π_α if $\beta \geq \alpha$. Therefore the partitions for different levels are nested.

Let us take the example provided in [187], where $\mathcal{U} = \{x_1, ...x_6\}$, and the similarity relation is defined by the following matrix:

$$\mu_S = \begin{pmatrix} 1 & 0.2 & 1 & 0.6 & 0.2 & 0.6 \\ 0.2 & 1 & 0.2 & 0.2 & 0.8 & 0.2 \\ 1 & 0.2 & 1 & 0.6 & 0.2 & 0.6 \\ 0.6 & 0.2 & 0.6 & 1 & 0.2 & 0.8 \\ 0.2 & 0.8 & 0.2 & 0.2 & 1 & 0.2 \\ 0.6 & 0.2 & 0.6 & 0.8 & 0.2 & 1 \end{pmatrix}. \quad (119)$$

The partition-tree corresponding to this similarity relation is depicted in Figure 13.

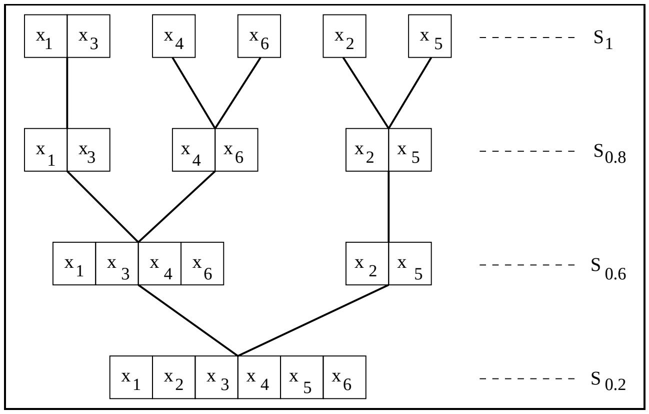


Figure 13: Partition tree corresponding to the similarity relation defined by Equation 119 (from [187]). For $\alpha = .2$, the α-cut $S_{0.2}$ provides only one class, where all x_i are similar up to th degree 0.2. For $\alpha = 0.6$, two classes are obtained, etc.

4.6.5 Order relations

On the contrary to similarity relations, order relations are not symmetrical. Several types of fuzzy orderings can be defined.

A fuzzy pre-order relation is a fuzzy relation that is:

1. reflexive,
2. max-min transitive.

A fuzzy order relation is a fuzzy relation that is:

1. reflexive,
2. anti-symmetric,
3. max-min transitive.

A perfect fuzzy order relation [103] or fuzzy partial ordering [187, 67] is a fuzzy relation that is:

1. reflexive,
2. perfectly anti-symmetric,
3. max-min transitive.

A total fuzzy order relation [103] or fuzzy linear ordering [187, 67] is a fuzzy partial ordering R such that:

$$\forall (x,y) \in \mathcal{U}^2,\ either\ \mu_R(x,y) > 0\ or\ \mu_R(y,x) > 0. \tag{120}$$

A fuzzy partial ordering R on a finite universe $\mathcal{U}$ can be represented equivalently by a triangular matrix and by a Hasse diagram (oriented graph whose arcs are valued by μ_R). Antisymmetry and transitivity of R lead to a graph without cycle. Let us illustrate these representation with the example provided in [187], for $\mathcal{U}$ having 6 elements. The matrix defining the fuzzy partial ordering is:

$$\mu_R = \begin{pmatrix} 1 & 0.8 & 0.2 & 0.6 & 0.6 & 0.4 \\ 0 & 1 & 0 & 0 & 0.6 & 0 \\ 0 & 0 & 1 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 1 & 0.6 & 0.4 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{121}$$

The Hasse diagram corresponding to this partial order relation is depicted in Figure 14.

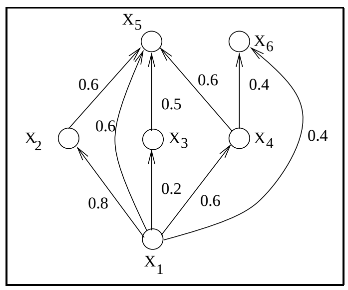


Figure 14: Hasse diagram corresponding to the fuzzy partial order relation defined by Equation 121 (from [187]).

4.7 Fuzzy and possibilistic logics

The development of fuzzy logic is directly linked to the specificities of human reasoning: it is more flexible than classical propositional logic, it can deal with imprecision and allows to derive deductions even from imperfect data and knowledge. It can handle graduate predicates, issued either from continuous referentials or from the notion of typicality (a situation can be more or less typical of known situations and this should be taken into account in similarity-based reasoning).

When reasoning on propositions, uncertainty (to be understood in a large sense) corresponds to the incapacity to decide if a proposition is true or false, either because the information is incomplete, vague, imprecise, or because it is contradicting or changing.

In the first case, a possibility-based modeling allows us to account for this type of uncertainty, while in the second case, a probabilistic modeling is well adapted.

Another important distinction is between degree of certainty and degree of truth. Fuzzy logic handles propositions endowed with truth degrees, while possibilistic logic rather handles propositions endowed with uncertainty degrees. Both perspectives are succinctly described below.

4.7.1 Fuzzy logic

In fuzzy logic [67, 66], the reasoning is performed on fuzzy elementary propositions such as:

$$X\ is\ P$$

where X is a variable taking its value in the reference space $\mathcal{U}$ and P is a fuzzy subset of $\mathcal{U}$, with membership function μ_P.

The truth degrees of such propositions are defined as values in $[0, 1]$ from μ_P.

Logical connectives are defined in a simple way, by using the same operators as their set theoretical equivalents. For instance, the truth degree of a conjunction of the type:

$$X \text{ } is \text{ } A \text{ } and \text{ } Y \text{ } is \text{ } B$$

is defined using a t-norm t by:

$$\mu_{A \wedge B}(x, y) = t[\mu_A(x), \mu_B(y)].$$

Similarly, a disjunction such as:

$$X \text{ } is \text{ } A \text{ } or \text{ } Y \text{ } is \text{ } B$$

has a truth degree defined using a t-conorm T:

$$\mu_{A \vee B}(x, y) = T[\mu_A(x), \mu_B(y)],$$

and a negation has a truth value defined using a fuzzy complementation:

$$\mu_{\neg A}(x) = c[\mu_A(x)].$$

In the case of variables taking values in a product space, X taking values in $\mathcal{U}$, and Y taking values in $\mathcal{V}$, then conjunction is interpreted as a Cartesian product. The corresponding truth degree of:

$$X \text{ } is \text{ } A \text{ } and \text{ } Y \text{ } is \text{ } B$$

can then be written as:

$$\mu_{A \times B}(x, y) = t[\mu_A(x), \mu_B(y)].$$

Let us now consider the implication. In classical logic, we have:

$$A \Rightarrow B \quad \Leftrightarrow \quad (B \text{ } or \text{ } not A),$$

which means that implication can be expressed using a disjunction and a negation. By using the same equivalence in the fuzzy case, we define a fuzzy implication from a t-conorm (disjunction) and a complementation (negation). Let us first consider non-fuzzy A and B. The degree to which A implies B is defined by:

$$Imp(A, B) = T[c(A), B]$$

where T is a t-conorm and c a complementation.

Now if A and B are fuzzy, we have:

$$Imp(A, B) = \inf_x T[c(\mu_A(x)), \mu_B(x)].$$

The following table summarizes the main fuzzy implications found in the literature on approximate 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 all cases, the implication reduces to the classical one in extreme cases of true (1) or false (0) propositions, i.e. taking binary truth values:

A	B	$A \Rightarrow B$
0	0	1
0	1	1
1	0	0
1	1	1

These defintions allow us now to define fuzzy equivalents of the main reasoning modes: modus ponens, modus tollens, syllogism, contraposition. As an example, let us consider modus ponens. In classical logic, it is written as:

$$(A \wedge (A \Rightarrow B)) \Rightarrow B. \tag{126}$$

Its fuzzy equivalent is defined as:

- Let us consider the rule:
$$if\ \ X\ is\ A\ \ then\ \ Y\ is\ B;$$
- and the knowledge:
$$X\ is\ A'$$
where A' is an approximation of A;
- then we derive the conclusion:
$$Y\ is\ B'$$
where B' is an approximation of B, with the degree:
$$\mu_{B'}(y) = \sup_x t[\mu_{A \Rightarrow B}(x, y), \mu_{A'}(x)].$$

We can now model and handle fuzzy rule based systems. Let us for instance consider the following rule:

$$\text{IF}\ \ (x\ is\ A\ \ \text{AND}\ \ y\ is\ B)\ \ \text{THEN}\ \ z\ is\ C$$

and α the truth degree of $x\ is\ A$, β the truth degree of $y\ is\ B$, γ the truth degree of $z\ is\ C$. The truth degree (or satisfaction degree) of the rule is obtained by combining the fuzzy connectives defined earlier:

$$Imp(t(\alpha, \beta), \gamma),$$

or:

$$T[c(T(\alpha, \beta)), \gamma)].$$

Similarly for the rule:

$$\text{IF}\ \ (x\ is\ A\ \ \text{OR}\ \ y\ is\ B)\ \ \text{THEN}\ \ z\ is\ C$$

its satisfaction degree is:

$$Imp(T(\alpha, \beta), \gamma) = T[c(T(\alpha, \beta)), \gamma)].$$

Such rules can be used for describing in a qualitative way the graph of a fuzzy function using a small number of rules. For instance, a function such as the one in Figure 15 can be described, at a somewhat coarse granularity level, by:

$$\text{IF}\ \ X\ is\ small\ \ \text{THEN}\ \ Y\ is\ small$$
$$\text{IF}\ \ X\ is\ medium\ \ \text{THEN}\ \ Y\ is\ large$$
$$\text{IF}\ \ X\ is\ large\ \ \text{THEN}\ \ Y\ is\ small$$

These rules involve the notion of linguistic variable defined earlier and the semantics of the values "small', "medium", "large" are defined by fuzzy sets on the definition domains of X and Y.

Fuzzy rule based systems have been used in many domains, mainly in fuzzy control, but also for approximate reasoning, for modeling flexible criteria in image processing and for their fusion, etc.

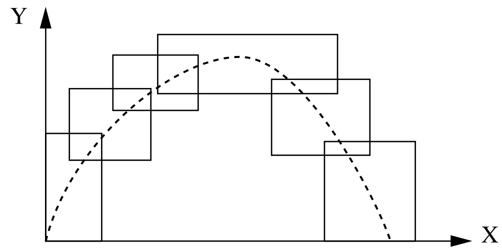


Figure 15: Example of a function graph

4.7.2 Possibilistic logic

Possibilistic logic relies on the definition of a possibility measure Π on a Boolean algebra of formulas B [66]:

$$\Pi : B \to [0,1]$$

such that:

- $\Pi(\bot) = 0$;
- $\Pi(\top) = 1$;
- $\forall \varphi, \phi,\ \Pi(\varphi \vee \psi) = \max(\Pi(\varphi), \Pi(\psi))$;
- $\forall \varphi,\ \Pi(\exists x \varphi) = \sup\{\Pi(\varphi[a|x]), a \in D(x)\}$ (where $D(x)$ is the domain of the variable x, and $\varphi[a|x]$ is obtained by replacing the occurrences of x in φ by a).

Let now Ω be the set of interpretations and π a normalized possibility distribution:

$$\pi : \Omega \to [0,1]$$

such that:

$$\exists \omega \in \Omega,\ \pi(\omega) = 1.$$

The possibility of a formula is then expressed by:

$$\Pi(\varphi) = \sup\{\pi(\omega),\ \omega \models \varphi\}$$

where $\omega \models \varphi$ should be read "ω is a model of φ" and means that φ is satisfied in the world ω.

Similarly as for sets, a necessity measure is defined on formulas by duality as:

$$N(\varphi) = 1 - \Pi(\neg\varphi).$$

The following property holds:

$$\forall \varphi, \phi,\ N(\varphi \wedge \psi) = \min(N(\varphi), N(\psi)).$$

This formalism allows us to deal with numerous situations by modeling them in a simple way. For instance a default rule such as "if A then B" that can have exceptions can be expressed by:

$$\Pi(A \wedge B) \geq \Pi(A \wedge \neg B)$$

Similarly, possibilistic modus ponens reasoning can be modeled by:

- if we have the rule: $N(A \Rightarrow B) = \alpha$

- and 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 a lot of domains, for instance to represent preference or utility models as ordered knowledge bases, and then to reason on these knowledge bases [77].

Let an ordered knowledge based denoted by:

$$KB = \{(\varphi_i, \alpha_i), i = 1...n\}$$

where α_i is a certainty or priority degree associated to formula φ_i (representing a knowledge element).

The satisfaction of this set of formulas in each world is represented by a possibility distribution defined as follows. If the knowledge base contains only one formula, we have:

$$\pi_{(\varphi,\alpha)}(\omega) = \begin{cases} 1 & \text{if } \omega \models \varphi \\ 1 - \alpha & \text{otherwise} \end{cases}$$

More generally, for a set of pieces of knowledges with priorities, we have:

$$\pi_{KB}(\omega) = \min_{i=1...n} \{1 - \alpha_i, \omega \models \neg\varphi_i\} = \min_{i=1...n} \max(1 - \alpha_i, \varphi_i(\omega)).$$

This formula can be interpreted as follows: if a formula is important (α_i close to 1), then the satisfaction degree of this formula in the world ω is taken into account. If, on the contrary, it is not important (α_i close to 0), then it does not play any role in the global evaluation of the knowledge base. The min corresponds to the fact that we are trying to know to which extent the formulas of the knowledge base are simultaneously satisfied in ω.

The degree of inconsistency of the base KB can be measured by the following expression:

$$1 - \max_{\omega} \pi_{KB}(\omega).$$

A base is said to be complete if it allows us to decide for any formula if it is true or if its contrary is true: either $KB \vdash \varphi$ (KB allows us to deduce φ), or $KB \vdash \neg\varphi$ (KB allows us to deduce $\neg\varphi$).

If a base is not complete, it leads to some ignorance about some formulas φ: $KB \nvdash \varphi$ and $KB \nvdash \neg\varphi$. Possibilistic logic leads to a very simple representation of this situation:

$$\Pi(\varphi) = \Pi(\neg\varphi) = 1$$

while there is no such simple model in probability theory for instance.

4.8 General principles for constructing fuzzy operations from binary ones

In this section, we present some common and generic methods that can be used for defining a fuzzy operator or relationship from their equivalent binary ones. These methods can be categorized in three main classes. The first type relies on the “extension principle”, as introduced by Zadeh [188]. The second class relies on computation on α-cuts. These two classes of definitions explicitly involve the operations or relations on crisp sets. The third class of methods consists in providing directly fuzzy definitions of the operations or of the relationships, by substituting all crisp expressions by their fuzzy equivalents.

4.8.1 Extension principle

Definition Originally, the first generic method for extending binary operators to fuzzy ones is due to Zadeh [188] and known as the extension principle.

Let us first consider a function f from $\mathcal{U}$ to $\mathcal{V}$. Let μ be a fuzzy set defined on $\mathcal{U}$. The extension of f to a fuzzy set is a fuzzy set μ' defined on $\mathcal{V}$. It is constructed as follows:

$$\forall y \in \mathcal{V},\ \mu'(y) = \begin{cases} 0 & if\ f^{-1}(y) = \emptyset, \\ \sup_{x \in \mathcal{U} | y = f(x)} \mu(x) & otherwise. \end{cases} \quad (138)$$

For an injective function, this equation reduces to:

$$\forall y \in \mathcal{V},\ \mu'(y) = \begin{cases} 0 & if\ f^{-1}(y) = \emptyset, \\ \mu[f^{-1}(y)] & otherwise. \end{cases} \quad (139)$$

Let us now consider the more general case where f is defined on a product space $\mathcal{U}_1 \times \mathcal{U}_2 \times ... \times \mathcal{U}_n$. Let $\mu_1, ...\mu_n$ be n fuzzy sets defined respectively on $\mathcal{U}_1, ...\mathcal{U}_n$. The extension of f on the μ_i provides a fuzzy set of $\mathcal{V}$ defined as:

$$\forall y \in \mathcal{V},\ \mu'(y) = \begin{cases} 0 & if\ f^{-1}(y) = \emptyset, \\ \sup_{(x_1,...x_n) \in \mathcal{U}_1 \times ... \times \mathcal{U}_n | y = f(x_1,...x_n)} \min[\mu_1(x_1), ...\mu_n(x_n)] & otherwise. \end{cases} \quad (140)$$

If the supremum is attained, i.e. if:

$$\forall y \in \mathcal{V}, \exists (x_1, ...x_n) \in \mathcal{U}_1 \times ... \times \mathcal{U}_n \mid \mu'(y) = \min[\mu_1(x_1), ...\mu_n(x_n)], \quad (141)$$

then the fuzzy extension of a function commutes with the α-cuts. i.e.:

$$\forall \alpha \in [0,1], [f(\mu_1, ...\mu_n)]_\alpha = f[(\mu_1)_\alpha, ...(\mu_n)_\alpha]. \quad (142)$$

Other extensions principles can be defined, using for instance the product instead of the minimum [67].

The extension principle is illustrated in Figure 16 for an injective function defined on a 1D space.

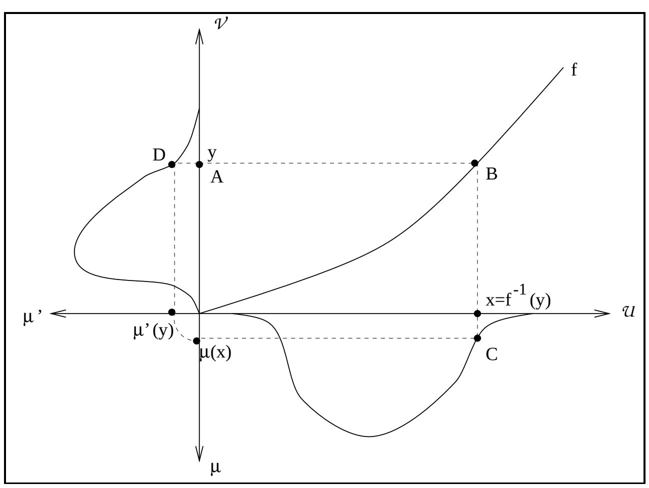


Figure 16: Extension principle. Starting from any point A with y as coordinate, B is obtained on the graph of f. B provides $x = f^{-1}(y)$, the absissa of the point C, on the graph of $\mu(x)$ at position x. With y as absissa and $\mu(x)$ as coordinate, point D of the looked for curve $\mu'(y)$ is obtained.

Application to the compatibility of two fuzzy sets A typical example of application of the extension principle is the compatibility between two fuzzy sets, as defined by Zadeh [185]. Let us consider a fuzzy set μ on $\mathcal{U}$. The value $\mu(x)$ may be interpreted as a degree of compatibility of x with the fuzzy set μ [67] (μ being for instance a fuzzy value, or a linguistic variable). The compatibility of a fuzzy set μ' of $\mathcal{U}$ with μ can be evaluated using the extension principle as a fuzzy set μ_{comp} on $[0,1]$:

$$\forall t \in [0,1],\ \mu_{comp}(t) = \begin{cases} 0 & if\ \mu^{-1}(t) = \emptyset, \\ \sup_{x\in\mathcal{U}|t=\mu(x)} \mu'(x) & otherwise. \end{cases} \quad (143)$$

The construction of the fuzzy compatibility set is illustrated in Figure 17. Note that compatibility is not symmetrical in μ and μ'.

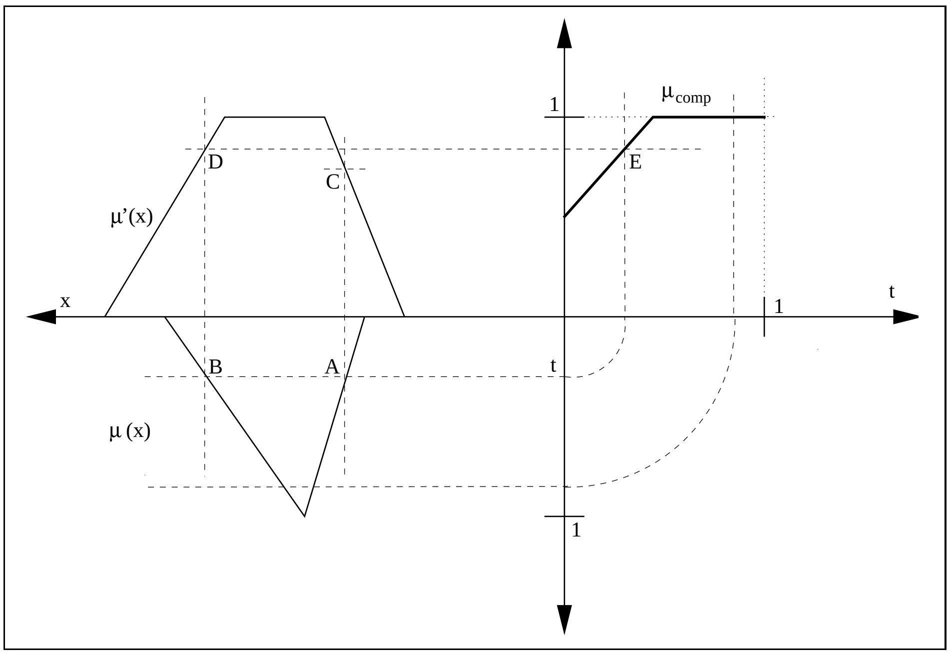


Figure 17: Construction of the compatibilty function between two fuzzy sets μ and μ'. Starting from any level t, we determine on the graph of μ (which has been drawn up-side down for the sake of interpretation), the two points A and B so that $\mu(x) = t$. The two points C and D are obtained on the graph of $\mu'(x)$, and D is chosen as $\mu(D) > \mu(C)$. The final point E with coordinates t and $\mu(D)$ is drawn, belonging to the looked for graph of $\mu_{comp}(t)$.

Application to fuzzy numbers One of the most important applications of the extension principle in the literature concerns the operations on fuzzy numbers, like addition, multiplication, etc. [67]. If μ_A and μ_B are two fuzzy numbers (convex, normalized fuzzy sets on $\mathbb{R}$), and $*$ any operation on numbers, the extension of $*$ is defined as:

$$\forall z \in \mathbb{R},\ \mu_{A*B}(z) = \sup_{(x,y)\in\mathbb{R}^2|x*y=z} \min[\mu_A(x), \mu_B(y)]. \quad (144)$$

The extension of a commutative (respectively associative) operation is commutative (respectively associative).

The computation of such extended operations can be performed using simple algorithms especially in the case of L-R fuzzy numbers, as shown in [67]. For instance let μ_1 and μ_2 two fuzzy numbers defined as:

$$\forall x \in \mathbb{R},\ \mu_1(x) = \begin{cases} L(\frac{m_1-x}{\alpha_1}) & for \quad x \leq m_1, \\ R(\frac{x-m_1}{\beta_1}) & for \quad x \geq m_1. \end{cases} \quad (145)$$

$$\forall x \in \mathbb{R},\ \mu_2(x) = \begin{cases} L(\frac{m_2 - x}{\alpha_2}) & for \quad x \leq m_2, \\ R(\frac{x - m_2}{\beta_2}) & for \quad x \geq m_2. \end{cases} \tag{146}$$

The sum of μ_1 and μ_2 is a L-R fuzzy number with parameters $m_1 + m_2$, $\alpha_1 + \alpha_2$ and $\beta_1 + \beta_2$.

Figure 18 illustrates the difference and the sum of two fuzzy numbers obtained by the extension principle.

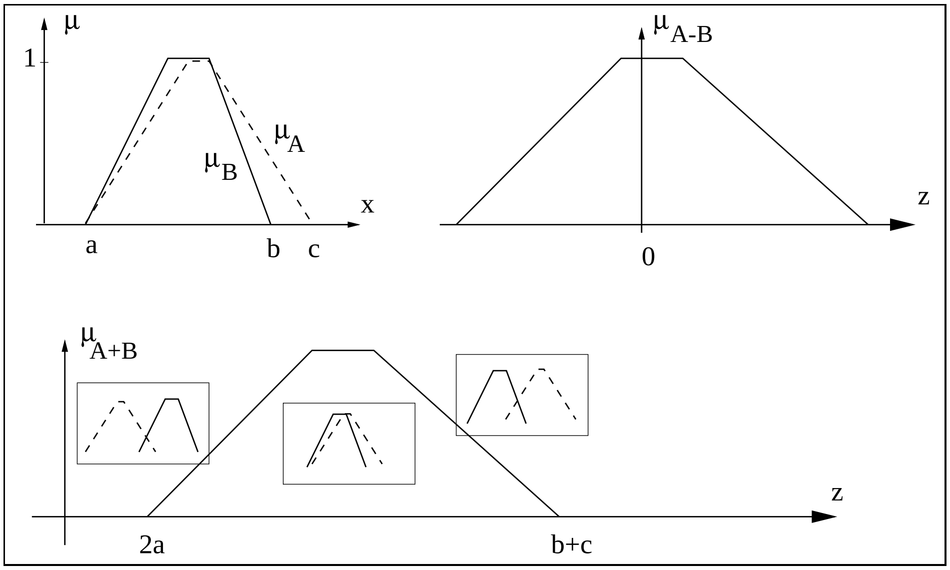


Figure 18: Application of the extension principle to the computation of two fuzzy numbers A and B. In the upper right corner, the difference $A-B$; in the lower row, the addition $A+B$. The three boxes present three different configurations of $\mu_A(x)$ and $\mu_B(z-x)$ for low, average, and high values of z. Note that the shapes of the 2 resulting functions are the same, but translated from a distance $supp(A) + Supp(B)$.

4.8.2 Combination of operations applied on α-coupes

Reconstruction from α-cuts One way to define crisp sets from a fuzzy set consists in taking the α-cuts of this set. Conversely, a fuzzy set can be reconstructed from its α-cuts, as seen in Section 4.1.4. Therefore a class of methods for defining fuzzy operations from crisp ones relies on the application of the crisp operation on each α-cut and then combining the results to reconstruct a fuzzy operation by stacking up the α-cuts.

Unary relations Let us denote by μ the membership functions of a fuzzy set defined on the space $\mathcal{U}$. Let us consider a crisp set function R_B (or operation on sets) taking values in a space $\mathcal{V}$. The fuzzy equivalent R of R_B is then defined as a function from $\mathcal{F}$ in $\mathcal{V}$ (see e.g. [65, 114, 38]):

$$R(\mu) = \int_0^1 R_B(\mu_\alpha) d\alpha. \tag{147}$$

Other fuzzification equations are possible, like [38, 88]:

$$R(\mu) = \sup_{\alpha \in [0,1]} \min(\alpha, R_B(\mu_\alpha)), \tag{148}$$

if the relation takes values in $[0, 1]$, or:

$$R(\mu) = \sup_{\alpha \in [0,1]} (\alpha R_B(\mu_\alpha)). \tag{149}$$

These equations may provide different results. But there also exist some links between them as shown later.

Binary relations Let us now consider a crisp operation R_B having two arguments (typically a relation between sets). The fuzzy equivalent R of R_B, applied to two fuzzy sets μ and ν of $\mathcal{U}$, is defined as a generalization of the previous equations:

$$R(\mu, \nu) = \int_0^1 R_B(\mu_\alpha, \nu_\alpha) d\alpha, \tag{150}$$

or, in this case, by a double integration as:

$$R(\mu, \nu) = \int_0^1 \int_0^1 R_B(\mu_\alpha, \nu_\beta) d\alpha d\beta. \tag{151}$$

The other fuzzification equations (148 and 149) can also be directly extended to operations on more than one fuzzy set.

The extension to n-ary operators is straightforward.

Extension principle based on α-cuts Another method for combining the results on α-cuts is similar to the extension principle [188]. In general it leads to a fuzzy set on $\mathcal{V}$. For instance if $\mathcal{V} = \mathbb{R}$, the crisp relation provides real values, the corresponding fuzzy relation using previous equations provides also numbers, while the following one provides fuzzy numbers. For a binary relation, we have:

$$\forall n \in \mathcal{V}, R(\mu, \nu)(n) = \sup_{R_B(\mu_\alpha, \nu_\alpha) = n} \alpha. \tag{152}$$

We have similar equations for unary or n-ary relations.

If the relationship to be extended only takes binary values (0/1, or true/false), then this equation reduces to:

$$R(\mu, \nu) = \sup_{R_B(\mu_\alpha, \nu_\alpha) = 1} \alpha. \tag{153}$$

The extension principle based on the α-cuts can be interpreted as follows: taking an α-cut of the sets corresponds to taking a binary decision on the boundaries of the objects. Then we look at the obtained value of the crisp relationship for this decision. If the same value of the relation can be obtained from different values of α, then the highest value of α is retained.

4.8.3 Translation of binary expressions into fuzzy ones

A last class of methods consists in translating binary equations into their fuzzy equivalent. This approach completely differs from the two previous ones in the sense that it does not use explicitly the crisp relation or operation. Indeed, in the extension principle as well as in α-cuts based approaches, the definition of a fuzzy operation is a function of the corresponding crisp operation. Here, a fuzzy operation is given directly by an equation involving fuzzy terms, that just mimics crisp equation.

This translation is generally done term by term. For instance, intersection is replaced by a t-norm, union by a t-conorm, sets by fuzzy set membership functions, etc. This translation is particularly straightforward if the binary relationship can be expressed in set theoretical and logical terms. Table 3 summarizes these main crisp concepts involved in set equations, and their fuzzy equivalent.

The many possibilities to translate for instance set union using a t-conorm induce that many definitions are issued from this method, depending on the choice of the fuzzy operators used for translating the crisp corresponding ones.

Let us take a simple example to illustrate this method. As defined before, a fuzzy set μ is said to be included in another fuzzy set ν if:

$$\forall x \in \mathcal{U}, \ \mu(x) \leq \nu(x).$$

crisp concept	equivalent fuzzy concept
set X	fuzzy set / membership function μ
complement of a set	fuzzy complementation c
intersection $\cap$	t-norm t
union $\cup$	t-conorm T
existence $\exists$	supremum
universal symbol $\forall$	infimum

Table 3: Translation of crisp concepts in their fuzzy equivalents.

This is a crisp definition of inclusion of fuzzy sets. We may also consider that if two sets are imprecisely defined, their inclusion relationship may be imprecise too. Therefore inclusion of fuzzy sets becomes a matter of degree. This degree of inclusion can be obtained using the translation principle.

In the crisp case, the set equation expressing inclusion of a set X in a set Y can be written as follows:

$$X \subset Y \Leftrightarrow X^C \cup Y = \mathcal{U} \quad (154)$$

$$\Leftrightarrow \forall x \in \mathcal{U},\ x \in X^C \cup Y. \quad (155)$$

Using the equivalence of Table 3 for each term, we have:

$$\forall x \in \mathcal{U} \leftrightarrow \inf_{x \in \mathcal{U}}, \quad (156)$$

$$x \in X^C \leftrightarrow c[\mu(x)], \quad (157)$$

$$x \in Y \leftrightarrow \nu(x), \quad (158)$$

$$X^C \cup Y \leftrightarrow T[c(\mu), \nu]. \quad (159)$$

Finally, the degree of inclusion of μ in ν is defined as:

$$\mathcal{I}(\mu, \nu) = \inf_{x \in \mathcal{U}} T[c(\mu(x)), \nu(x)], \quad (160)$$

where T is a t-conorm and c a fuzzy complementation.

4.8.4 Comparison

The extension principle has been originally defined for functions. The approaches presented in Section 4.8.2 deal mainly with operators (set operators, relationships between sets, etc.). Links between extension principle and combination of α-cuts using Equation 148 have been established in [88]. Let f be a function from $\mathcal{U}_1 \times ... \times \mathcal{U}_n$ to $\mathcal{V}$, and R_f a set operator defined as:

$$R_f(X_1, X_2, ...X_n) = \{f(x_1, x_2, ...x_n) | x_1 \in X_1, ...x_n \in X_n\}, \quad (161)$$

where $X_1, ...X_n$ are subsets of $\mathcal{U}_1, ...\mathcal{U}_n$. Then the extension of R_f using Equation 148 coincides with Zadeh's extension of f (Equation 140).

Other links exist between definitions of Section 4.8.2. For instance if R_B is a crisp relationship taking values in $\{0, 1\}$, its extension using Equation 153 is a value in $[0, 1]$ and is equivalent to the two fuzzification procedures given by equations 148 and 149.

Let us take the example of extending union between sets, using the methods of Section 4.8.2. Let μ and ν be two fuzzy sets on $\mathcal{U}$. Using the integration over α-cuts, we get:

$$(\mu \cup \nu)(x) = \int_0^1 (\mu_\alpha \cup \nu_\alpha)(x) d\alpha$$

$$= \int_0^{\max[\mu(x), \nu(x)]} 1 d\alpha$$

$$= \max[\mu(x), \nu(x)],$$

since $(\mu_\alpha \cup \nu_\alpha)(x) = 1$ iff $\mu(x) \geq \alpha$ or $\nu(x) \geq \alpha$. This extension leads exactly to the fuzzy union as defined originally in [186]. Exactly the same result is obtained using other fuzzification methods, e.g. with Equation 148 or 153.

Using formal translation of equations as described in Section 4.8.3, we may obtain the same results as using some combination of α-cuts. Examples will be seen later, e.g. for fuzzy connectivity.

The question of which extension should be used does not have a definite answer until now. Extension principle is well adapted for translating analytical expressions, while formal translation is well adapted if the operators to be extended can be expressed using set theoretical and logical terms. The properties of the obtained extended operators have to play an important role in the choice of a method, since they may vary depending on the method. For instance, as will be seen later, using simple integration on the α-cuts (Equation 150) for extending a distance between sets to a distance between fuzzy sets endows the fuzzy distance with the same properties as the crisp distance, while using a double integration (Equation 151), some properties may be lost.

5 Fuzzy sets and possibility theory in image processing and vision: tools for spatial reasoning under imprecision

THIS SECTION COULD PROBABLY BE INCLUDED IN THE REASONING PART OF CCV ONTOLOGY, SECTION METHODS (3.3).

5.1 Introduction

Imprecision is often inherent to images, and its causes can be found at several levels: observed phenomenon (imprecise limits between structures or objects), acquisition process (limited resolution, numerical reconstruction methods), image processing steps (imprecision induced by a filtering for instance). Fuzzy sets have several advantages for representing such imprecision. First, they are able to represent several types of imprecision in images, as for instance imprecision in spatial location of objects, or imprecision in membership of an object to a class. For instance, partial volume effect, which occurs frequently in medical imaging, finds a consistent representation in fuzzy sets (membership degrees of a voxel to tissues or classes directly represent partial membership to the different tissues mixed up in this voxel, using their proportions, leading to a modeling very close to reality). Second, image information can be represented at different levels with fuzzy sets (local, regional, or global), as well as under different forms (numerical, or symbolic). For instance, classification based only on grey levels involves very local information (at the pixel level); introducing spatial coherence in the classification or relationships between features like parallelism involves regional information; and introducing relationships between objects or regions for scene interpretation involves more global information. Third, the fuzzy set framework allows for the representation of very heterogeneous information, and is able to deal with information extracted directly from the images, as well as with information derived from some external knowledge, like expert knowledge for instance. This is exploited in particular in model-based pattern recognition, where fuzzy information extracted from the images is compared and matched to a model representing knowledge expressed in fuzzy terms.

5.2 Representation of spatial information

5.2.1 Fuzzy spatial objects

Here, the space $\mathcal{U}$ is the image space, typically $\mathbb{Z}^2$ or $\mathbb{Z}^3$ for digital 2D or 3D images, or, in the continuous case, $\mathbb{R}^2$ or $\mathbb{R}^3$, and we will denote it $\mathcal{S}$ (to refer to the spatial domain). We are interested in the objects of the image that we may describe as fuzzy sets. Thus we often call them fuzzy image objects.

A fuzzy image object is a fuzzy set defined on $\mathcal{S}$, i.e. a spatial fuzzy set. Its membership function μ represents the imprecision in the spatial extent of the object. For any point x of $\mathcal{S}$ (pixel or voxel), $\mu(x)$ is the degree to which x belongs to the fuzzy object.

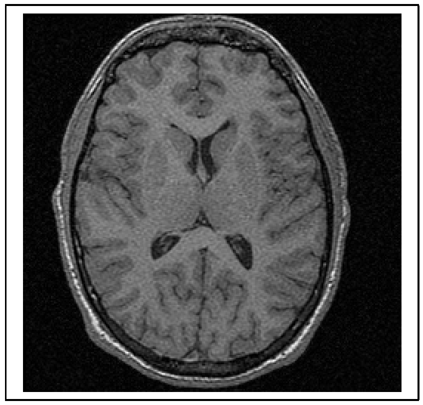


Figure 19: MR image of a brain (one slice of a 3D volume).

As an illustrative example, a slice of a brain image is shown in Figure 19. It is obtained using a T1 weighted acquisition in magnetic resonance imaging (MRI). A few internal structures are represented in Figure 20 as spatial fuzzy sets, where membership degrees are represented using grey levels. The use of fuzzy sets may represent different types of imprecision, either on the boundary of the objects (due for instance to partial volume effect, or to the spatial resolution), or on the individual variability of these structures, etc.

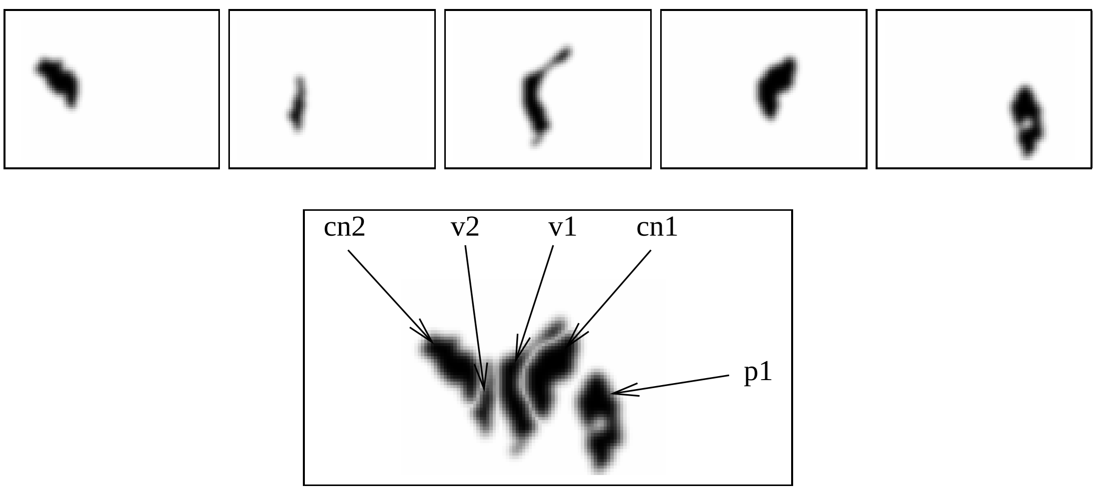


Figure 20: Top: 5 fuzzy objects representing internal brain structures of the image shown in Figure 19 (membership values rank between 0 and 1, from white to black). From left to right: right caudate nucleus (cn2), right lateral ventricle (v2), left lateral ventricle (v1), left caudate nucleus (cn1), left putamen (p1). Bottom: superposition of these fuzzy objects (the maximum membership value is displayed at each point).

One of these structures (v1) is represented in Figure 21, along with its support, its 0.5 α-cut, and its membership function as a 3D graph.

5.2.2 Set operations

Once we have defined spatial fuzzy objects, we can now address the question of defining operations on such objects. Basic operations are set operations. Some of these have been defined in Section 4.1. The

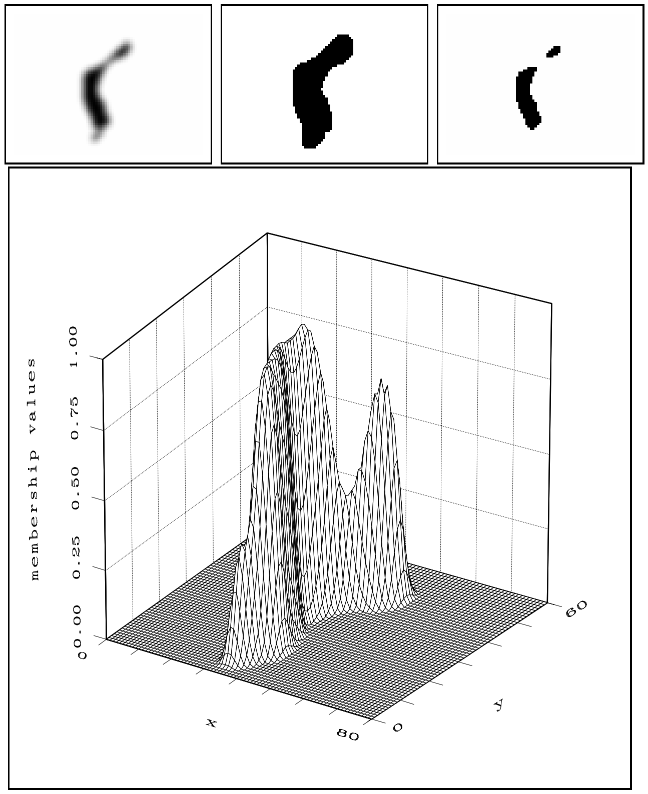


Figure 21: A fuzzy image object (v1), its support, its 0.5 α-cut, and its membership function represented as a 3D graph (two axes represent the spatial coordinates, while the third one represents the membership values).

point of view that was taken was to construct a fuzzy set which is the result of the combination of two fuzzy sets by a set operation. Now, we take another point of view, which addresses the question: are two fuzzy sets satisfying some set relationships? For instance: is μ included in ν? In the crisp case, such questions receive binary answers. When the objects are fuzzy, imprecisely defined, the answer to such questions becomes a matter of degree, and amounts to define a degree to which the relation is satisfied. The example of degree of inclusion was given in Section 4.8, and is further detailed here. We also define degrees of intersection, union, and equality, by making intensive use of the different extension methods presented in Section 4.8.

Degree of intersection

Crisp case Saying that two sets X and Y intersect translates in the fuzzy case as a degree $\mu_{int}(\mu, \nu)$ to which two fuzzy sets μ and ν intersect. The set equation expressing that two crisp sets intersect is:

$$X \cap Y \neq \emptyset. \tag{162}$$

This equation is equivalent to:

$$\exists x \in \mathcal{S},\ x \in X \cap Y. \tag{163}$$

On the other hand, the fact that X and Y do not intersect is expressed by the non-satisfaction of this equation. These two possible states in the crisp case correspond to a binary "degree" of intersection, which is equal to 1 if the equation is satisfied, and to 0 if it is not.

Direct extension In the fuzzy domain, this binary degree becomes a degree in $[0, 1]$, which expresses the degree of satisfaction of this equation. It can be defined for instance using the formal translation method. The simplest fuzzy translation (see Table 3) provides:

$$\mu_{int}(\mu, \nu) = \sup_{x \in \mathcal{S}} t[\mu(x), \nu(x)], \tag{164}$$

where t is any t-norm. This expression may vary from 0, which corresponds to no intersection at all (typically if μ and ν have disjoint supports) to 1 if at least one point x belongs completely to both μ and ν.

Note that in this case, the other fuzzification methods (combination of α-cuts, extension principle, as seen in Section 4.8) provide the same result, or a particular case where $t = \min$. For instance, if we consider the integration over the α-cuts, we have:

$$\mu_{int}(\mu, \nu) = \int_0^1 1_{\mu_\alpha \cap \nu_\alpha \neq \emptyset} d\alpha, \tag{165}$$

where $1_{\mu_\alpha \cap \nu_\alpha \neq \emptyset}$ is the binary function expressing intersection, equals 1 if $\mu_\alpha \cap \nu_\alpha \neq \emptyset$ and 0 else. Since we have:

$$\forall (\alpha, \alpha') \in [0, 1]^2,\ \alpha \leq \alpha' \Rightarrow \mu_{\alpha'} \subset \mu_\alpha \ and\ \nu_{\alpha'} \subset \nu_\alpha, \tag{166}$$

it follows that:

$$\mu_{\alpha'} \cap \nu_{\alpha'} \neq \emptyset \Rightarrow \mu_\alpha \cap \nu_\alpha \neq \emptyset. \tag{167}$$

Therefore, it is enough to consider only the supremum of the values of α such that $\mu_\alpha \cap \nu_\alpha \neq \emptyset$, and the following equalities hold:

$$\begin{aligned} \mu_{int}(\mu, \nu) &= \int_0^1 1_{\mu_\alpha \cap \nu_\alpha \neq \emptyset} d\alpha & (168) \\ &= \sup_{\mu_\alpha \cap \nu_\alpha \neq \emptyset} \alpha & (169) \\ &= \sup_\alpha \min[\alpha, 1_{\mu_\alpha \cap \nu_\alpha \neq \emptyset}] & (170) \\ &= \sup_\alpha [\alpha 1_{\mu_\alpha \cap \nu_\alpha \neq \emptyset}] & (171) \\ &= \sup_{x \in \mathcal{S}} \min[\mu(x), \nu(x)]. & (172) \end{aligned}$$

From the degree of intersection, a degree of empty intersection (or of disjunctness) is then derived as:

$$\mu_{\neg int}(\mu, \nu) = c[\mu_{int}(\mu, \nu)], \tag{173}$$

where c is a fuzzy complementation (for instance defined as $\forall a \in [0, 1], c(a) = 1 - a$).

This form has already been widely used in the fuzzy set literature. In particular, it is often interpreted as a degree of conflict between two fuzzy sets or two possibility distributions [74].

Introducing the volume of the overlapping domain However this form is not always adequate for image processing or vision purposes since it does not include any spatial information. This may even lead to counter-intuitive results. The expression $\sup_{x \in \mathcal{S}} t[\mu(x), \nu(x)]$ only represents the maximum height of the intersection. Although it is generally low for fuzzy sets that have almost disjoint supports, its value does not account for different overlapping situations, as illustrated by Figure 22 (for sake of clarity, $\mathcal{S}$ is represented in 1D only).

The degree of intersection and of non-intersection can therefore be reformulated in order to better represent the notion of spatial overlapping. Another solution, that may be better for some applications, consists in defining a degree of intersection by considering the fuzzy hypervolume V_n (in a space of

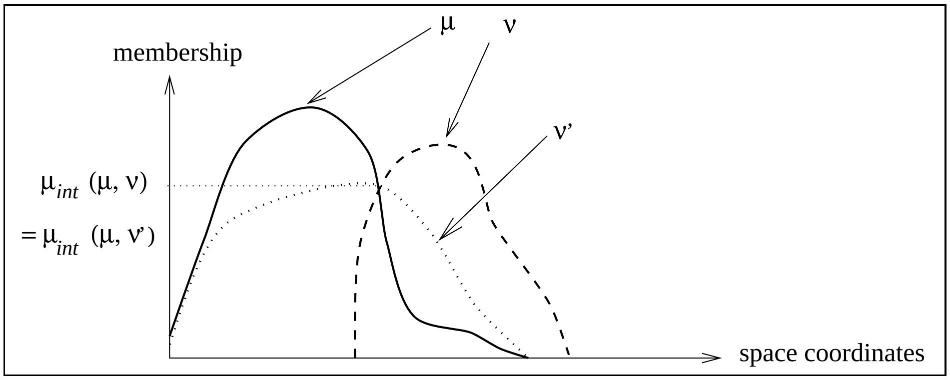


Figure 22: Low discrimination power of the definition of degree of intersection between two fuzzy sets using the maximum of intersection: $\mu_{int}(\mu,\nu') = \mu_{int}(\mu,\nu)$, although μ and ν' strongly overlap and should be considered as more intersecting than μ and ν.

dimension n) of the intersection. This also corresponds to a translation process, in the sense that we have:

$$X \cap Y = \emptyset \;\Leftrightarrow\; V_n(X \cap Y) = 0. \tag{174}$$

For defining the hypervolume of a fuzzy set, we simply use the classical fuzzy cardinality. This provides for a fuzzy set μ (having bounded support) in the discrete case:

$$V_n(\mu) = \sum_{x \in \mathcal{S}} \mu(x), \tag{175}$$

and in the continuous case:

$$V_n(\mu) = \int_{x \in \mathcal{S}} \mu(x). \tag{176}$$

From the hypervolume of $t(\mu,\nu)$, we can derive a degree of intersection in $[0,1]$. It should be equal to 0 if μ and ν have completely disjoint supports, be high if one set is included in the other, and increasing with respect to the hypervolume of the intersection. The following definition satisfies these requirements[9]:

$$\mu_{int}(\mu,\nu) = \frac{V_n[t(\mu,\nu)]}{\min[V_n(\mu), V_n(\nu)]}. \tag{177}$$

Again a degree of non-intersection can be derived from this expression using equation 173.

In the example shown in Figure 23, μ and ν have the same degree of intersection according to equation 164 (maximum of the intersection, equal to 0.63 in this case) than μ and ν'. Using the fuzzy volume as in definition 177, we obtain $\mu_{int}(\mu,\nu) = 0.31$ and $\mu_{int}(\mu,\nu') = 0.66$. This corresponds well to the fact that μ and ν' have a larger overlap than μ and ν.

Properties An important property of these various definitions is that the intersection degrees defined by Equation 164 and Equation 177 are both consistent with the binary definition. Moreover, they satisfy the following properties:

- symmetry: $\forall(\mu,\nu) \in \mathcal{F}^2,\ \mu_{int}(\mu,\nu) = \mu_{int}(\nu,\mu)$;
- reflexivity (equation 164) if the fuzzy sets are normalized: $\exists x \in \mathcal{S}, \mu(x) = 1 \Rightarrow \mu_{int}(\mu,\mu) = 1$; for Equation 177, reflexivity holds if $t = \min$;
- if one of the sets is empty $(\forall x \in \mathcal{S}, \nu(x) = 0)$, then the degree of intersection is always 0;

[9]Other definitions leading to similar properties are possible.

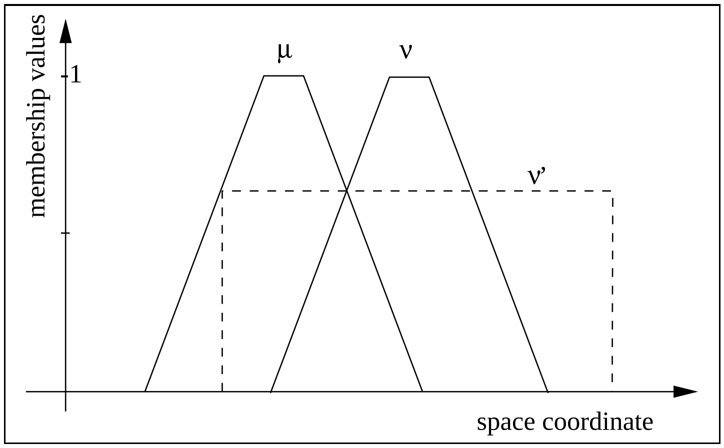


Figure 23: μ and ν have the same degree of intersection than μ and ν' using the maximum of the intersection, while they have a lower one using the fuzzy hypervolume.

- if one of the sets is equal to $\mathcal{S}$ ($\forall x \in \mathcal{S}, \nu(x) = 1$), the degree of intersection is always equal to 1 using Equation 177, and to 1 for normalized fuzzy sets using Equation 164;
- invariance with respect to geometrical transformations (translation, rotation).

Application to the non-contradiction principle In the crisp case, we have for any set X:

$$X \cap X^C = \emptyset \tag{178}$$

which is a set equation equivalent to the logical principle of non-contradiction.

In the fuzzy case, the degree of intersection between any fuzzy set μ and its complement is:

$$\mu_{int}(\mu, c(\mu)) = \sup_{x \in \mathcal{S}} t[\mu(x), 1 - \mu(x)], \tag{179}$$

if we take $c(a) = 1 - a$ for the fuzzy complementation.

In general this degree is non zero. However, for the Lukasiewicz t-norm (see 4.4), we have:

$$\mu_{int}(\mu, c(\mu)) = \sup_{x \in \mathcal{S}} \max[0, \mu(x) + 1 - \mu(x) - 1] = 0. \tag{180}$$

This extends to sets the conclusion drawn for values in Section 4.1: in general a t-norm does not lead to the non-contradiction principle, but the Lukasiewicz t-norm does.

Degree of union When union is concerned, the question is raised in different terms from what is done for intersection. Here, it is not interesting to look to which degree two sets have a non-empty union: it is always satisfied. More interesting is the degree to which the union of two sets covers the whole space. This will be useful for instance for looking at the excluded-middle law.

In the crisp case, we have for two sets X and Y:

$$X \cup Y = \mathcal{S} \Leftrightarrow \forall x \in \mathcal{S},\ x \in X \cup Y. \tag{181}$$

As for intersection, the binary satisfaction of this equation is replaced by a degree in $[0, 1]$ in the fuzzy case. Using the formal translation principle, we get:

$$\mu_{union}(\mu, \nu) = \inf_{x \in \mathcal{S}} T[\mu(x), \nu(x)], \tag{182}$$

where T is a t-conorm. The same result (and in particular with $T = \max$) is obtained using the other fuzzification principles.

The properties of this definition are:

- consistency with the binary definition;
- symmetry: $\forall(\mu, \nu) \in \mathcal{F}^2,\ \mu_{union}(\mu, \nu) = \mu_{union}(\nu, \mu)$;
- if one of the sets is empty ($\forall x \in \mathcal{S}, \nu(x) = 0$), then μ_{union} is always 0 for bounded support fuzzy sets;
- if one of the sets is equal to $\mathcal{S}$ ($\forall x \in \mathcal{S}, \nu(x) = 1$), μ_{union} is always equal to 1;
- invariance with respect to geometrical transformations (translation, rotation).

Let us now look at the excluded-middle law. In the crisp case, we have for any set X:

$$X \cup X^C = \mathcal{S} \tag{183}$$

which is a set equation equivalent to the logical principle of excluded-middle.

In the fuzzy case, the degree of union between any fuzzy set μ and its complement is:

$$\mu_{union}(\mu, c(\mu)) = \inf_{x \in \mathcal{S}} T[\mu(x), 1 - \mu(x)], \tag{184}$$

if we take $c(a) = 1 - a$ for the fuzzy complementation.

In general this degree is not equal to 1. However, for the Lukasiewicz t-conorm, we have:

$$\mu_{union}(\mu, c(\mu)) = \inf_{x \in \mathcal{S}} \min[1, \mu(x) + 1 - \mu(x)] = 1. \tag{185}$$

This extends to sets the conclusion already obtained for values in Section 4.1: in general a t-conorm does not lead to the excluded-middle law, but the Lukasiewicz t-conorm does.

Degree of inclusion

Inclusion from other set operations In the crisp case, we have for two sets X and Y;

$$X \subset Y \quad \Leftrightarrow \quad X \cap Y^C = \emptyset \tag{186}$$

$$\Leftrightarrow \quad X^C \cup Y = \mathcal{S}. \tag{187}$$

These two equivalences show that inclusion is simply expressed either by an intersection or by a union. The extension to fuzzy sets can therefore directly use the degrees of intersection and union as defined in this Section.

Using the degree of intersection, we obtain for the degree of inclusion of μ in ν:

$$\mathcal{I}(\mu, \nu) = c[\mu_{int}(\mu, c(\nu))] = c[\sup_{x \in \mathcal{S}} t[\mu(x), c(\nu(x))]], \tag{188}$$

where t is a t-norm and c a fuzzy complementation.

Using the union, we obtain:

$$\mathcal{I}(\mu, \nu) = \mu_{union}(c(\mu), \nu) = \inf_{x \in \mathcal{S}} T[c(\mu(x)), \nu(x)], \tag{189}$$

where T is a t-conorm.

Actually, due to the duality between t-norms and t-conorms, these two definitions are equivalent. Let t and T be a pair of dual t-norm and t-conorm according to the complementation c. Then we have:

$$\inf_{x\in\mathcal{S}} T[c(\mu(x)), \nu(x)] = \inf_{x\in\mathcal{S}} c[t[\mu(x), c(\nu(x))]] = c[\sup_{x\in\mathcal{S}} t[\mu(x), c(\nu(x))]], \tag{190}$$

which proves the equivalence between both formulas.

This definition has the same drawback as degrees of intersection and union: they basically may depend on one point only. Here again, the coverage between both fuzzy sets could be taken into account.

The properties of the degree of inclusion are directly derived from those of intersection and union:

- consistency with the binary definition;
- if μ is empty ($\forall x \in \mathcal{S}, \mu(x) = 0$), then $\mathcal{I}(\mu, \nu)$ always equals 1;
- if ν is empty, then $\mathcal{I}(\mu, \nu)$ is equal to 0 for normalized fuzzy sets;
- if μ is equal to $\mathcal{S}$ ($\forall x \in \mathcal{S}, \mu(x) = 1$), $\mathcal{I}(\mu, \nu)$ is equal to 0 for bounded support fuzzy sets;
- if ν is equal to $\mathcal{S}$, $\mathcal{I}(\mu, \nu)$ is always equal to 1;
- invariance with respect to geometrical transformations (translation, rotation).

Axiomatization of fuzzy inclusion Other methods for defining a degree of inclusion rely on a set of axioms and on the determination of functions satisfying these axioms. This is the method adopted for instance in [161] and in [184].

The axioms of [161] are the following:

1. $\mathcal{I}(\mu, \nu) = 1$ iff $\mu \subset \nu$ in Zadeh's sense, i.e. $\forall x \in \mathcal{S},\ \mu(x) \leq \nu(x)$.
2. $\mathcal{I}(\mu, \nu) = 0$ iff $Core(\mu) \cap [Supp(\mu)]^C \neq \emptyset$.
3. $\mathcal{I}$ is increasing in ν: if $\nu_1 \subset \nu_2$, then $\mathcal{I}(\mu, \nu_1) \leq \mathcal{I}(\mu, \nu_2)$.
4. $\mathcal{I}$ is decreasing in μ: if $\mu_1 \subset \mu_2$, then $\mathcal{I}(\mu_1, \nu) \geq \mathcal{I}(\mu_2, \nu)$.
5. $\mathcal{I}$ is invariant under geometric transformations like translation, rotation, etc.
6. $\mathcal{I}(\mu, \nu) = \mathcal{I}(c(\nu), c(\mu))$.
7. $\mathcal{I}(\mu \cup \mu', \nu) = \min[\mathcal{I}(\mu, \nu), \mathcal{I}(\mu', \nu)]$.
8. $\mathcal{I}(\mu, \nu \cap \nu') = \min[\mathcal{I}(\mu, \nu), \mathcal{I}(\mu, \nu')]$.
9. $\mathcal{I}(\mu, \nu \cup \nu') \geq \max[\mathcal{I}(\mu, \nu), \mathcal{I}(\mu, \nu')]$.

Some additional properties are proposed in [161], but not as mandatory ones:

10. $\mathcal{I}(\mu, \nu) + \mathcal{I}(\mu, c(\nu)) \geq 1$.
11. $\mathcal{I}(\mu \cup c(\mu), \mu \cap c(\mu)) \leq \mathcal{I}(\nu \cup c(\nu), \nu \cap c(\nu))$ if μ is a refinement of ν.
12. $\mathcal{I}(\mu, \nu) \geq 0.5$ if μ is weakly included in ν (i.e. $\forall x \in \mathcal{S}$ *either* $\mu(x) \leq 0.5$ *or* $\nu(x) > 0,5$).

The degree of inclusion proposed in [161] according to their axioms is defined as:

$$\forall (\mu, \nu) \in \mathcal{F}^2,\ \mathcal{I}(\mu, \nu) = \inf_{x\in\mathcal{S}} \min[1, \lambda(\mu(x)) + \lambda(1 - \nu(x))] \tag{191}$$

where λ is a function from $[0, 1]$ into $[0, 1]$ such that:

- λ is non-increasing,
- $\lambda(0) = 1$,
- $\lambda(1) = 0$,
- the equation $\lambda(x) = 0$ has a single solution,
- $\forall \alpha \in [0.5, 1]$, the equation $\lambda(x) = \alpha$ has a single solution,
- $\forall a \in [0,1],\ \lambda(a) + \lambda(1-a) \geq 1$.

A typical example for λ is: $\lambda(a) = 1 - a^n$, with $n \geq 1$. In particular, for $n = 1$, the degree of inclusion becomes:

$$\mathcal{I}(\mu, \nu) = \inf_{x \in \mathcal{S}} \min[1, 1 - \mu(x) + \nu(x)] \tag{192}$$

which is exactly the inclusion obtained from intersection or union (Equations 188 and 189) for the complementation $c(a) = 1 - a$ and for the Lukasiewicz t-norm and t-conorm.

Despite the apparent similarity between Equation 191 and 189, they are not equivalent. Indeed, the function defined as:

$$\min[1, \lambda(1-a) + \lambda(1-b)], \tag{193}$$

which plays in Equation 191 the same role as the t-conorm T in Equation 189, is actually not a t-conorm, since it it not associative and it does not admit 0 as unit element, except for $\lambda(a) = 1 - a$ [38]. This induces a loss of properties of the inclusion degree in comparison to those of inclusion derived from a true t-conorm, as will be seen for instance for fuzzy mathematical morphology (Section 5.3).

Another axiomatization has been proposed in [184]. The axioms for degree of inclusion are the following:

1. $\mathcal{I}(\mu, \nu) = 1$ iff $\mu \subset \nu$ in Zadeh's sense, i.e. $\forall x \in \mathcal{S},\ \mu(x) \leq \nu(x)$; this is the same as the first axiom of [161].
2. Let ν be such that $\forall x \in \mathcal{S},\ \nu(x) = 0.5$. If $\nu \subset \mu$, then $\mathcal{I}(\mu, c(\mu)) = 0$ iff $\mu = \mathcal{S}$ (i.e. $\forall x \in \mathcal{S},\ \mu(x) = 1$); this contrasts with the second axiom of [161].
3. if $\nu \subset \mu_1 \subset \mu_2$, then $\mathcal{I}(\mu_1, \nu) \geq \mathcal{I}(\mu_2, \nu)$, which is weaker than axiom 4 of [161]; if $\nu_1 \subset \nu_2$, then $\mathcal{I}(\mu, \nu_1) \leq \mathcal{I}(\mu, \nu_2)$, which is axiom 3 of [161].

These axioms are weaker than the ones of [161]. They lead to weaker properties of the degree of inclusion, and also to weaker properties than the degree of inclusion derived from t-norms and t-conorms.

Inclusion and fuzzy entropy Links between degree of inclusion and fuzzy entropy have been studied by several authors, including [111, 184]. These links are expressed as:

$$\forall \mu \in \mathcal{F},\ \mathcal{I}(\mu \cup c(\mu), \mu \cap c(\mu)) = E(\mu), \tag{194}$$

where $E(\mu)$ denotes the fuzzy entropy of μ [122].

The definition of degree of inclusion of [111], defined for finite $\mathcal{S}$, is:

$$\mathcal{I}(\mu, \nu) = \frac{|\mu \cap \nu|}{|\mu|} = \frac{\sum_{x \in \mathcal{S}} \min[\mu(x), \nu(x)]}{\sum_{x \in \mathcal{S}} \mu(x)}, \tag{195}$$

with the convention $\mathcal{I}(\mu, \nu) = 1$ if $\forall x \in \mathcal{S},\ \mu(x) = 0$.

The corresponding fuzzy entropy is then:

$$E(\mu) = \frac{\sum_{x \in \mathcal{S}} \min[\mu(x), 1 - \mu(x)]}{\sum_{x \in \mathcal{S}} \max[\mu(x), 1 - \mu(x)]}. \tag{196}$$

Another example is the degree of inclusion of [93], defined for finite $\mathcal{S}$, as:

$$\mathcal{I}(\mu,\nu) = \frac{1}{|\mathcal{S}|}\sum_{x\in\mathcal{S}} \min[1, 1-\mu(x)+\nu(x)]. \tag{197}$$

The corresponding entropy measure is:

$$E(\mu) = \frac{1}{|\mathcal{S}|}\sum_{x\in\mathcal{S}} \min[\mu(x), 1-\mu(x)], \tag{198}$$

which is the fuzzy entropy of [103].

This degree of inclusion is similar to the one of [98], that relies on the same combination operator but using a different type of normalization:

$$\mathcal{I}(\mu,\nu) = \frac{\inf_{x\in\mathcal{S}} \min[1, 1-\mu(x)+\nu(x)]}{\sup_{x\in\mathcal{S}} \mu(x)}. \tag{199}$$

In [184], it is proved more generally that if $\mathcal{I}$ is an inclusion degree that satisfies the 3 axioms of [184], then the measure defined by:

$$\forall\mu\in\mathcal{F},\ E(\mu) = \mathcal{I}(\mu\cup c(\mu), \mu\cap c(\mu)) \tag{200}$$

is a fuzzy entropy measure on $\mathcal{S}$.

Inclusion from fuzzy implication Finally, inclusion can be defined from implication [6, 177, 184], as:

$$\mathcal{I}(\mu,\nu) = \inf_{x\in\mathcal{S}} Imp[\mu(x),\nu(x)], \tag{201}$$

where $Imp[\mu(x),\nu(x)]$ denotes the degree to which $\mu(x)$ implies $\nu(x)$.

The degrees of implication defined earlier can be used for this purpose. For instance, if we use:

$$Imp[\mu(x),\nu(x)] = T[c(\mu(x)),\nu(x)], \tag{202}$$

we recover exactly the same definition of degree of inclusion as the one obtained from a t-conorm (Equation 189).

Degree of equality The degree of equality between two fuzzy sets μ and ν can be simply defined from the inclusion degree. In the binary case, we have:

$$X = Y \Leftrightarrow X\subset Y\ and\ Y\subset X. \tag{203}$$

Therefore, in the fuzzy case, the degree of equality between μ and ν is defined as:

$$\mathcal{E}(\mu,\nu) = \min[\mathcal{I}(\mu,\nu), \mathcal{I}(\nu,\mu)], \tag{204}$$

or more generally, for any t-norm t:

$$\mathcal{E}(\mu,\nu) = t[\mathcal{I}(\mu,\nu), \mathcal{I}(\nu,\mu)]. \tag{205}$$

For a degree of inclusion defined from a t-norm or equivalently from a t-conorm (Equations 188 and 189), we obtain:

$$\mathcal{E}(\mu,\nu) = t[\inf_{x\in\mathcal{S}} T[c(\mu(x)),\nu(x)], \inf_{x\in\mathcal{S}} T[c(\nu(x)),\mu(x)]]. \tag{206}$$

The properties of the degree of equality are the following:

- consistency with the binary definition;
- reflexivity if T is such that the excluded-middle law holds (e.g. for the Lukasiewicz t-conorm);
- symmetry: $\mathcal{E}(\mu,\nu) = \mathcal{E}(\nu,\mu)$;
- if μ is empty ($\forall x \in \mathcal{S}, \mu(x) = 0$), then $\mathcal{E}(\mu,\nu) = \inf_{x\in\mathcal{S}} c(\nu(x))$, which is equal to 0 for normalized fuzzy sets;
- if μ is equal to $\mathcal{S}$ ($\forall x \in \mathcal{S}, \mu(x) = 1$), $\mathcal{I}(\mu,\nu)$ is equal to 0 for bounded support fuzzy sets;
- invariance with respect to geometrical transformations (translation, rotation).

5.2.3 Geometrical fuzzy sets

Fuzzy points and lines Fuzzy points are an extension in the n-dimensional space of fuzzy numbers. Indeed, a fuzzy number can be considered as a fuzzy point of the real line [191]. Two methods can be used for defining a fuzzy point from a fuzzy number [48]. The first one consists in defining a fuzzy number on each axis, let say $\mu_1, ...\mu_n$. A fuzzy point is then the set $(\mu_1, ...\mu_n)$, where μ_i is a fuzzy number on the i-th axis. The second method consists in defining a fuzzy point as a fuzzy set on $\mathcal{S}$, through its membership function μ, that has to satisfy a set of properties extending the fuzzy number ones:

1. μ is upper semi-continuous;
2. $\exists ! x_0 \in \mathcal{S},\ \forall x \in \mathcal{S},\ \mu(x) = 1 \Leftrightarrow x = x_0$, which expresses the unicity of the modal value of μ; the fuzzy number is said to be centered at x_0;
3. $\forall \alpha \in [0,1],\ \mu_\alpha$ is a compact and convex subset of $\mathcal{S}$; note that this is equivalent to say that μ is a convex fuzzy set.

This second definition is more convenient since a fuzzy point is a fuzzy object in the image space, which is in accordance with the notion of fuzzy image object. An example of fuzzy point is shown in Figure 24 according to this definition.

Several definitions of fuzzy lines have been proposed in [48]. They are summarized below for a 2D space (extensions to higher dimension is straightforward):

1. Let μ_A, μ_B and μ_C be the membership functions of three fuzzy numbers. A fuzzy line is defined as the set of all fuzzy numbers μ_1 and μ_2 that are solutions of the following equation:
$$\mu_A \mu_1 + \mu_B \mu_2 = \mu_C. \tag{207}$$
The resolution of such equations calls for fuzzy arithmetic [154], and does not always have solutions. Therefore, this definition is not further considered.
2. Let μ_A and μ_B be the membership functions of two fuzzy numbers. A fuzzy line is the set (μ_1, μ_2) where μ_1 is any fuzzy number and μ_2 is obtained by the following equation:
$$\mu_2 = \mu_A \mu_1 + \mu_B, \tag{208}$$
where the sum and product of fuzzy numbers is computed using the extension principle as shown in Section 4.8. However, this definition does not lead to a fuzzy object in $\mathcal{S}$. One of the consequences mentioned in [48] is that it cannot be visualized (as pictures or graphs as in Figure 24).
3. Let μ_A, μ_B and μ_C be the membership functions of three fuzzy numbers, such that the modal values of μ_A and μ_B are not both zero. A fuzzy line is defined as a fuzzy set on $\mathcal{S}$ through its membership function μ_L as:
$$\forall (x,y) \in \mathcal{S},\ \mu_L(x,y) = \sup\{\alpha \in [0,1] \mid ax + by = c,\ a \in \mu_{A_\alpha}, b \in \mu_{B_\alpha}, c \in \mu_{C_\alpha}\}. \tag{209}$$
This definition corresponds to the extension principle applied on the definition of a crisp line, and extending it by considering that the coefficient defining the line are fuzzy.

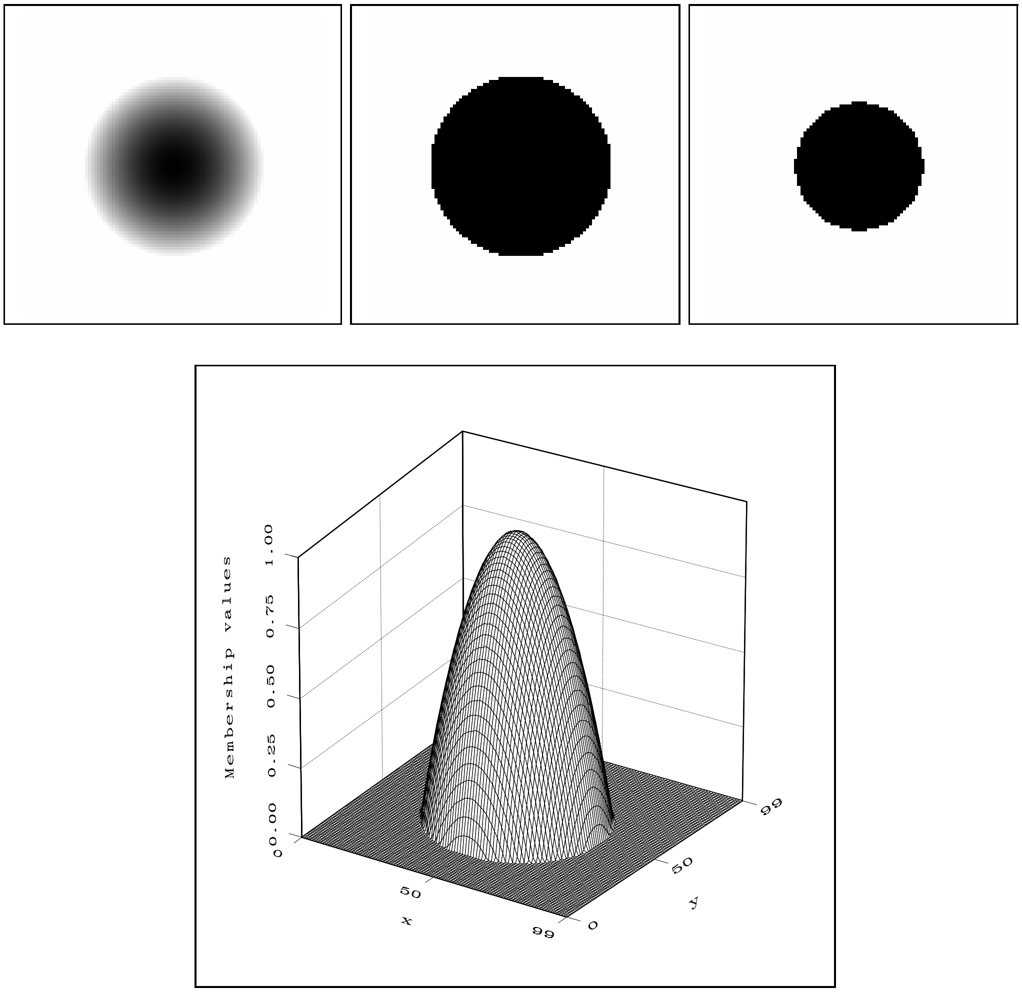


Figure 24: Example of fuzzy point in the 2D space represented using grey levels (white for a zero membership value), its support and its α-cut for $\alpha = 0.5$. Second line: 3D representation of the membership values.

4. Another definition is obtained using the same method as for the third definition, but using the equation $y = ax + b$, which is another form of the equation of a crisp line.
5. Similarly, another definition is obtained using the slope form of a line: $y - b = c(x - a)$.
6. The last definition uses two fuzzy points μ_A and μ_B of $\mathcal{S}$ and uses again the extension principle:

$$\forall (x, y) \in \mathcal{S},\ \mu_L(x, y) = \sup\{\alpha \in [0, 1] \mid \frac{y - v_1}{x - u_1} = \frac{v_2 - v_1}{u_2 - u_1},\ (u_1, v_1) \in \mu_{A_\alpha}, (u_2, v_2) \in \mu_{B_\alpha}\}. \quad (210)$$

Definitions 3-6 are in agreement with the concept of fuzzy image object. The two other ones are not further considered here. Note that definitions 3-6 all rely on the same principle, and just use different analytical expressions of a crisp line. Any other expression could be used, and would lead to a definition of fuzzy line.

It is shown in [48], that under certain conditions (such that 0 does not belong to the support of μ_B in definition 3), definitions 3 to 6 may provide the same fuzzy line.

The properties of μ_L according to definitions 3-6 are:

- the α-cuts of μ_L are closed, connected and arc-wise connected, but not necessarily convex;

- μ_L is normalized, and there is at least one crisp line in μ_{L1};
- μ_L is upper semi-continuous.

Let us take an example, where we use the equation $y = ax + b$ and where only b is extended to a fuzzy number μ_B, a remaining crisp. Then the α-cuts of the fuzzy line are bands in the direction of a, and the width of which is exactly the width of μ_B at level α, as illustrated in Figure 25.

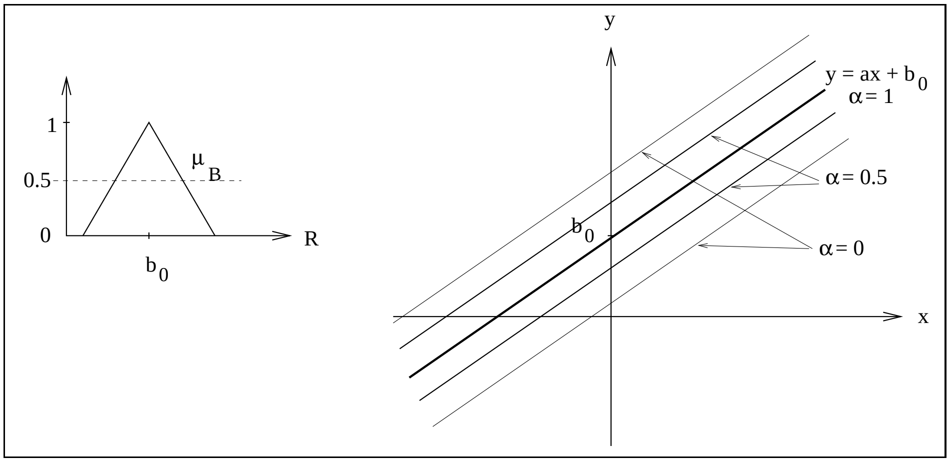


Figure 25: Example of fuzzy line in the 2D space, where only b is fuzzy in the equation $y = ax + b$. Three α-cuts are shown. The fuzzy number μ_B is represented on the left.

Fuzzy rectangles and fuzzy convex polygons Let us give another example of fuzzy geometrical structure, that extends the notion of rectangle, and, more generally, of convex polygon, as defined in [146].

We consider here again a 2D space $\mathcal{S}$. A fuzzy set μ on $\mathcal{S}$ is separable if there exists a coordinate frame (x, y) such that μ can be expressed as the conjunction of two fuzzy sets μ_x and μ_y defined on the x-axis and y-axis respectively:

$$\forall (x, y) \in \mathcal{S}^2, \; \mu(x, y) = \min[\mu_x(x), \mu_y(y)]. \tag{211}$$

Note that instead of the minimum, any t-norm could be used, which makes the definition more general. Moreover, using the t-norm product, the classical definition of a separable function, as defined in functional analysis, is obtained.

From this notion of separable fuzzy set, the one of fuzzy rectangle is derived. A fuzzy set μ on $\mathcal{S}$ is a fuzzy rectangle if it is separable and fuzzy convex. It is proved in [146] that for a separable fuzzy set, the three following properties are equivalent:

1. μ is fuzzy connected, i.e. $\forall (P, Q) \in \mathcal{S}^2$, there exists an arc from P to Q such that for each point R in the arc, we have:
$$\mu(R) \geq \min[\mu(P), \mu(Q)].$$
We will elaborate on this notion of fuzzy connectivity later on.
2. μ is fuzzy convex, i.e. for each R on the segment joining P to Q, we have:
$$\mu(R) \geq \min[\mu(P), \mu(Q)].$$
3. μ is fuzzy orthoconvex, i.e. we have the same equation as for convexity whenever P and Q have the same x-coordinate or the same y-coordinate (when they form a vertical or horizontal segment).

A fuzzy rectangle can be characterized as follows: μ is a fuzzy rectangle if and only if there exist two fuzzy convex sets in 1D, μ_x and μ_y, such that:

$$\forall (x,y) \in \mathcal{S}^2,\ \mu(x,y) = \min[\mu_x(x), \mu_y(y)]. \tag{212}$$

Another characterization is that μ is a fuzzy rectangle if and only if all α-cuts of μ are crisp rectangles. Therefore a fuzzy rectangle appears as a "stack" of crisp rectangles, as illustrated in Figure 26.

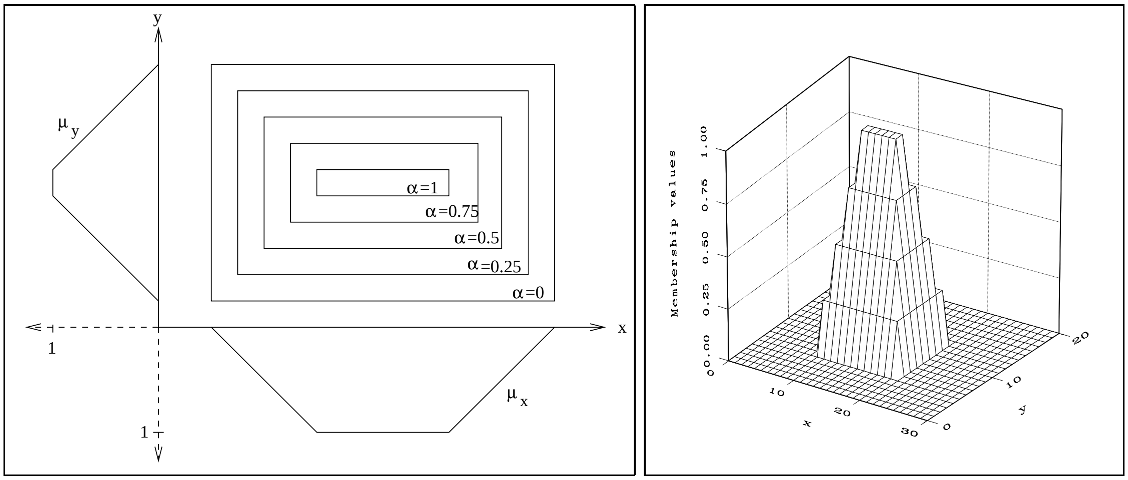


Figure 26: Example of fuzzy rectangle. Left: μ_x and μ_y and a few α-cuts of μ. Right: 3D representation of μ.

In a more general way, fuzzy convex polygons are defined as intersections of fuzzy half-planes. A fuzzy set μ is a fuzzy half-plane if there exists a direction x and a 1D fuzzy set ν such that:

$$\forall (x,y) \in \mathcal{S}^2,\ \mu(x,y) = \nu(x), \tag{213}$$

and ν is monotonically non-increasing:

$$x_1 \leq x_2 \Rightarrow \nu(x_1) \geq \nu(x_2). \tag{214}$$

A fuzzy set μ of $\mathcal{S}$ is a fuzzy halfplane if and only if the α-cuts of μ are (possibly degenerate) nested half-planes. An important property is that a fuzzy half-plane is fuzzy convex.

Let us now consider k fuzzy half-planes $\mu_1, ...\mu_k$ of $\mathcal{S}$, whose associated directions are denoted by $x_1, ...x_k$ and are given in cyclic order modulo 2π. If every pair of successive directions (for $i < k$, (x_i, x_{i+1}) and (x_k, x_1)) differs by less than π, then the fuzzy set μ defined on $\mathcal{S}$ as:

$$\mu = \min(\mu_1, \mu_2, ...\mu_k) \tag{215}$$

is a fuzzy convex polygon. It is indeed fuzzy convex, as the intersection of fuzzy convex sets. Again a characteristic property is that μ is a fuzzy convex polygon if and only if its α-cuts are nested convex polygons.

It can be proved that a fuzzy rectangle is a fuzzy convex polygon.

Fuzzy disks A fuzzy set μ of $\mathcal{S}$ is a fuzzy disk if there exists a point P of $\mathcal{S}$ such that the membership of any point only depends on its distance to P. The point P is called the center of the fuzzy disk. In polar coordinates in a 2D space, say (r, θ), a fuzzy disk has a membership function that is a function of r alone.

A fuzzy disk is convex if and only if $\mu(r)$ is non-increasing in r. Moreover, we generally assume that μ has a bounded support. The α-cuts of a convex fuzzy disk are nested crisp disks. If the fuzzy disk has a unique modal value (i.e. a unique point x such that $\mu(x) = 1$), then it is a fuzzy point.

Two examples of fuzzy disks are shown in Figure 27, one being convex and the other not.

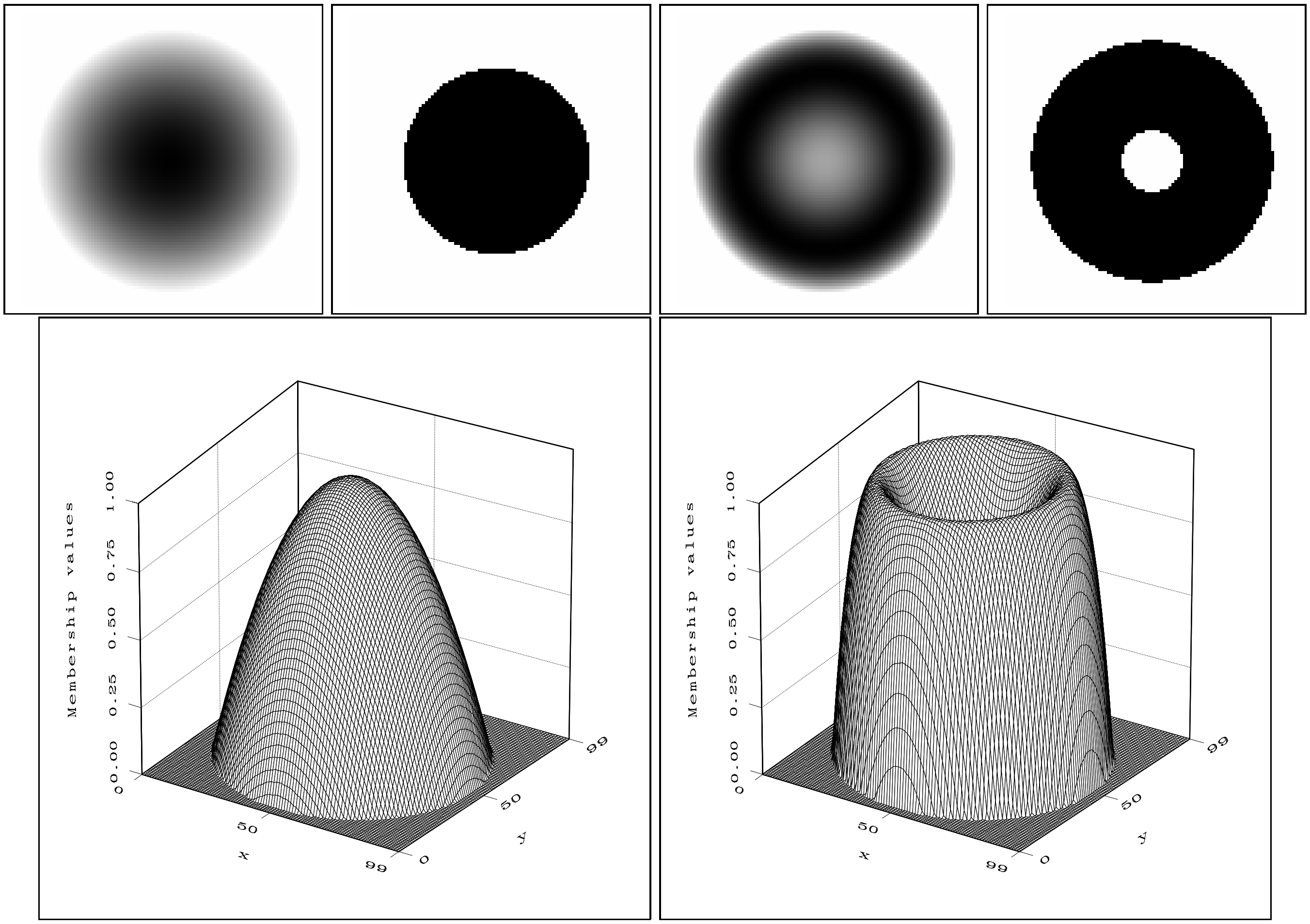


Figure 27: Two examples of fuzzy disks. The left one is convex, the right one is not. Their α-cuts for $\alpha = 0.5$ are also shown. The second line shows 3D representations of these two fuzzy disks (the third axis represents the membership values).

5.2.4 Geometrical measures of fuzzy objects

Several geometrical measures are useful for describing, characterizing and recognizing shapes, as perimeter, area, compactness, as well as more complex measures like symmetry, parallelism, etc. In this Section, we present how to compute such measures on fuzzy objects.

Area of a fuzzy set The area of a fuzzy set μ defined on the space $\mathcal{S}$ is defined as [146]:

$$a(\mu) = \int_{\mathcal{S}} \mu(x) dx, \tag{216}$$

and in the case where $\mathcal{S}$ is finite:

$$a(\mu) = \sum_{x \in \mathcal{S}} \mu(x). \tag{217}$$

Note that this is nothing else but the fuzzy cardinality of μ.

For the two examples of fuzzy disks shown in Figure 27, we obtain respectively $a(\mu) = 3168.62$ for the convex one, and it has 6325 points in its support, and $a(\mu) = 4567.99$ for the non-convex one (which has 6349 points in its support, i.e. approximately the same as for the first one).

More generally for a fuzzy disk, where μ is a function of r only in polar coordinates, the area is equal to:

$$a(\mu) = \int_{(x,y)\in\mathcal{S}} \mu(x,y)dxdy = \int_0^{2\pi}\int_0^{+\infty} r\mu(r)drd\theta = 2\pi\int_0^{+\infty} r\mu(r)dr. \quad (218)$$

Further examples are shown in Table 4, for the brain structures shown in Figure 20.

Fuzzy object	size of support	fuzzy area
nc1	512	239.40
nc2	451	192.55
p1	535	244.10
v1	636	224.86
v2	295	77.98

Table 4: Some examples of fuzzy areas, for the objects shown in Figure 20. For each fuzzy object, the cardinality of its support (i.e. the number of points having a strictly positive membership value) and its fuzzy area are given.

Perimeter of a fuzzy set The perimeter of a fuzzy set has been defined in [148], again in the 2D case. Let us first consider the case where $\mathcal{S}$ is finite, and where the fuzzy set μ is piecewise constant. Note that this is the case that is useful in image processing, and corresponds to a quantization of the membership values, that can take a finite number of values μ_i $(i = 1...n)$ in $[0,1]$ (i.e. only a finite number of α-cuts are different). In such cases, there exists a partition of $\mathcal{S}$ in regions U_i on which μ is constant and equal to μ_i. Let us denote by U_{ij} the common boundary of U_i and U_j, that is written as:

$$U_{ij} = \overline{U_i} \cap \overline{U_j}, \quad (219)$$

where $\overline{U_i}$ denotes the closure of U_i. It is composed of a set of arcs, denoted by $A_{ijk}, k = 1...n_{ij}$, that are rectifiable, and the length of which is denoted by $l(A_{ijk})$. The perimeter of μ is then defined as:

$$p(\mu) = \sum_{i=1}^{n}\sum_{j=i+1}^{n}\sum_{k=1}^{n_{ij}} |\mu_i - \mu_j| l(A_{ijk}). \quad (220)$$

If μ is crisp, this definition is equivalent to the classical perimeter of a crisp set.

Let us now consider the continuous case, and assume that the fuzzy object has a membership function μ that is differentiable. The gradient of μ is then defined, and its magnitude is written as:

$$|\nabla\mu(x,y)| = \sqrt{(\frac{\partial\mu}{\partial x})^2 + (\frac{\partial\mu}{\partial y})^2}. \quad (221)$$

Then the perimeter of μ is defined as:

$$p(\mu) = \int_{(x,y)\in\mathcal{S}} |\nabla\mu(x,y)|dxdy, \quad (222)$$

if this integral exists.

It is shown in [148] that both definitions, in the continuous case and in the piecewise constant case, coincide in the limit. Moreover, they are both special cases of a more general formulation in the framework of generalized functions.

In the convex case, then the α-cuts of μ are nested convex crisp sets, and therefore simply connected. The boundaries U_{ij} are then just curves. The fuzzy perimeter is then a weighted sum of the length of these curves (see Figure 28).

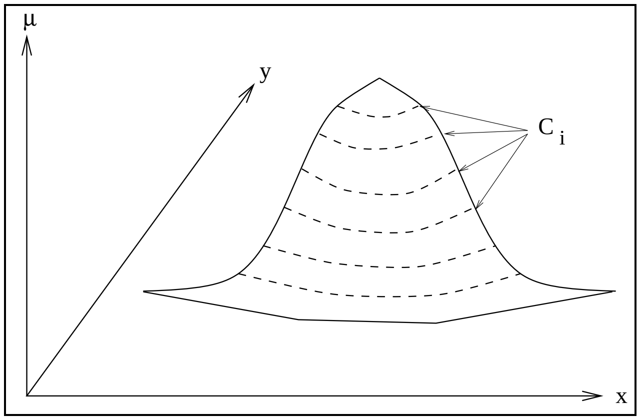


Figure 28: Computation of the perimeter of a convex fuzzy set: it is a weighted sum of the lengths of the curves C_i delimitating the α-cuts of μ.

An important property is that if μ and ν are piecewise constant fuzzy convex subsets of $\mathcal{S}$ such that $\mu \leq \nu$, then $p(\mu) \leq p(\nu)$.

In the case where μ is a fuzzy disk, then, with $\mu'(r)$ denoting the derivative of μ with respect to r, we have:

$$p(\mu) = 2\pi \int_0^{+\infty} r|\mu'(r)|dr. \tag{223}$$

It comes out that a fuzzy disk may have an arbitrarily large perimeter (typically if μ oscillates rapidly), while having a small area. It is interesting to note that this phenomenon, that is not surprising for crisp sets, can also be observed for disks in the fuzzy case.

In the case of a convex fuzzy disk, we have:

$$|\mu'(r)| = -\mu'(r), \tag{224}$$

since μ is decreasing in r. Then the perimeter becomes:

$$p(\mu) = 2\pi \int_0^{+\infty} -r\mu'(r)dr = 2\pi \left[\int_0^{+\infty} [-r\mu(r)]'dr + \int_0^{+\infty} \mu(r)dr \right]. \tag{225}$$

If we further assume that μ decreases fast enough so that:

$$\lim_{r \to +\infty} r\mu(r) = 0, \tag{226}$$

then we have:

$$p(\mu) = 2\pi \int_0^{+\infty} \mu(r)dr. \tag{227}$$

Table 5 provides the perimeter of a few fuzzy 2D objects. The computation is performed on a discrete grid, and the neighbors are determined using both 4- and 8-connectivities (the standard types of discrete connectivity used in image processing on a square grid). These results show for instance that for the two fuzzy disks of Figure 27, the non-convex one has a much larger perimeter, although the size of its support is about the same as for the convex one, due to the non-convexity of its membership function.

Compactness of a fuzzy set The compactness of a fuzzy set is defined as [146]:

$$c(\mu) = \frac{a(\mu)}{p(\mu)^2}. \tag{228}$$

Fuzzy object	fuzzy perimeter (4c)	fuzzy perimeter (8c)
convex fuzzy disk	144.35	168.93
non-convex fuzzy disk	243.86	285.39
nc1	54.25	60.72
nc2	44.20	51.29
p1	60.54	67.46
v1	70.35	78.42
v2	32.48	35.06

Table 5: Some examples of fuzzy perimeters, for the two fuzzy disks of Figure 27 and for the objects shown in Figure 20. For each fuzzy object, the perimeter in 4-connectivity and in 8-connectivity are given.

In the crisp case, the compactness is the largest for disks, where it is equal to $\frac{1}{4\pi}$ (isoperimetric inequality), which means that the perimeter cannot be small while the area is large. In the fuzzy case, we do not have the same inequality. However, it can be shown that if μ is a convex fuzzy disk, then we have:

$$\frac{a(\mu)}{p(\mu)^2} \geq \frac{1}{4\pi}. \tag{229}$$

It means that among all possible convex fuzzy disks, the compactness is smallest for a crisp disk.

Table 6 shows the values of compactness obtained for the fuzzy objects of Figures 27 and 20.

Fuzzy object	fuzzy compactness (4c)	fuzzy compactness (8c)
convex fuzzy disk	0.15	0.11
non-convex fuzzy disk	0.08	0.06
nc1	0.08	0.06
nc2	0.10	0.07
p1	0.07	0.05
v1	0.05	0.04
v2	0.07	0.06

Table 6: Some examples of fuzzy compactness, for the two fuzzy disks of Figure 27 and for the objects shown in Figure 20. For each fuzzy object, the perimeter in 4-connectivity and in 8-connectivity is used for computing the compactness. The obtained values should be compared to $\frac{1}{4\pi} \sim 0.08$.

Height, width and diameter of a fuzzy set The height and width of a fuzzy object μ in a 2D space are defined respectively by [146]:

$$h(\mu) = \int_y [\sup_x \mu(x, y)] dy, \tag{230}$$

$$w(\mu) = \int_x [\sup_y \mu(x, y)] dx. \tag{231}$$

These equations correspond to the integral of the maximum membership value on horizontal (respectively vertical) lines. In the finite case, these equations become:

$$h(\mu) = \sum_y \max_x \mu(x, y), \tag{232}$$

$$w(\mu) = \sum_x \max_y \mu(x, y). \tag{233}$$

The following relationship holds with the area:

$$a(\mu^2) \leq h(\mu)w(\mu). \tag{234}$$

The extrinsic diameter of a fuzzy object μ is defined as the supremum of the integrals of its projections [146]:

$$e(\mu) = \sup_x \int_u [\sup_v \mu(u,v)]du, \tag{235}$$

where u and v are the coordinates in a frame formed by any pair of two orthogonal directions.

The extrinsic diameter is related to the area by the following inequality:

$$a(\mu^2) \leq [e(\mu)]^2. \tag{236}$$

Let us now consider the case of a connected fuzzy object μ, i.e. such that all its α-cuts are connected in the crisp sense. Then the intrinsic diameter of μ (in the finite case) is defined as:

$$i(\mu) = \max_{P,Q}[\min_{\mathcal{P}_{PQ}} \int_{\mathcal{P}_{PQ}} \mu(x,y)dxdy], \tag{237}$$

where the max is taken over all points P and Q in $\mathcal{S}$, and the min is taken over over all paths $\mathcal{P}_{PQ}$ between P and Q such that:

$$\forall (x,y) \in \mathcal{P}_{PQ},\ \mu(x,y) \geq \min[\mu(P), \mu(Q)].$$

The connection of μ guarantees that such paths always exist. If μ is crisp, $i(\mu)$ is exactly the intrinsic crisp diameter, i.e. the greatest possible distance between two points in μ, computed along paths that are completely included in μ. In the crisp case, we have $e(\mu) \leq i(\mu)$. In the convex crisp case, we have $e(\mu) = i(\mu)$. In the fuzzy case, we can have $e(\mu) > i(\mu)$, even in the convex case. However, if μ is convex, the two following inequalities hold:

$$e(\mu) \geq i(\mu), \tag{238}$$

$$i(\mu) \leq \frac{p(\mu)}{2}. \tag{239}$$

Intersection and parallelism between fuzzy lines These notions have been introduced in [48]. Let μ_{L_1} and μ_{L_2} be two fuzzy lines. A measure of parallelism between these two fuzzy lines if defined by $\rho(\mu_{L_1}, \mu_{L_2}) = 1 - \lambda(\mu_{L_1}, \mu_{L_2})$, where:

$$\lambda(\mu_{L_1}, \mu_{L_2}) = \sup_{\mathcal{S}}[\min(\mu_{L_1}(x,y), \mu_{L_2}(x,y))]. \tag{240}$$

This is nothing but the height of the intersection between μ_{L_1} and μ_{L_2}. If the supports of the two fuzzy lines do not intersect, then the lines are completely parallel. In the crisp case, the degree of parallelism is equal to 1 iff the lines are parallel in the standard sense.

5.2.5 Fuzzy geometrical measures of fuzzy objects

In the previous definitions, the geometrical measures of fuzzy objects were defined as numbers. If the objects are imprecisely defined, we can also expect their geometrical measures to be imprecisely defined too, i.e. to be fuzzy numbers instead of numbers. The extension principle can be used for this purpose (see Section 4.8).

Let M be any geometrical measure (area, perimeter, etc.). The extension principle leads to:

$$\forall \lambda \in \mathbb{R}^+,\ M(\mu)(\lambda) = \sup_{M(\mu_\alpha)=\lambda} \alpha. \tag{241}$$

For instance for the area, the obtained definition is equivalent to the cardinality of a fuzzy set defined as a fuzzy number [67].

Since the α-cuts of μ are nested, their area decreases when α increases. It follows that the fuzzy area has a decreasing membership function.

The definition of fuzzy area as a fuzzy number is illustrated in Figure 29 for the two fuzzy disks shown in Figure 27, and in Figure 30 for the brain structures shown in Figure 20.

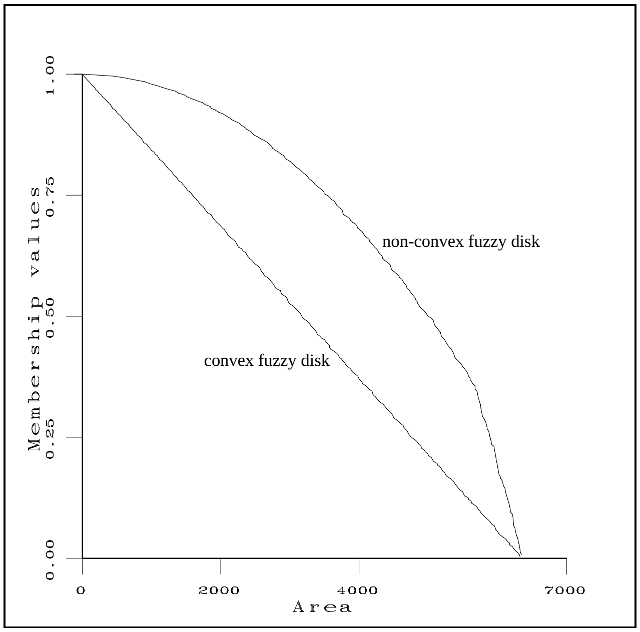


Figure 29: Area of a fuzzy set as a fuzzy number, illustrated on the two fuzzy disks shown in Figure 27.

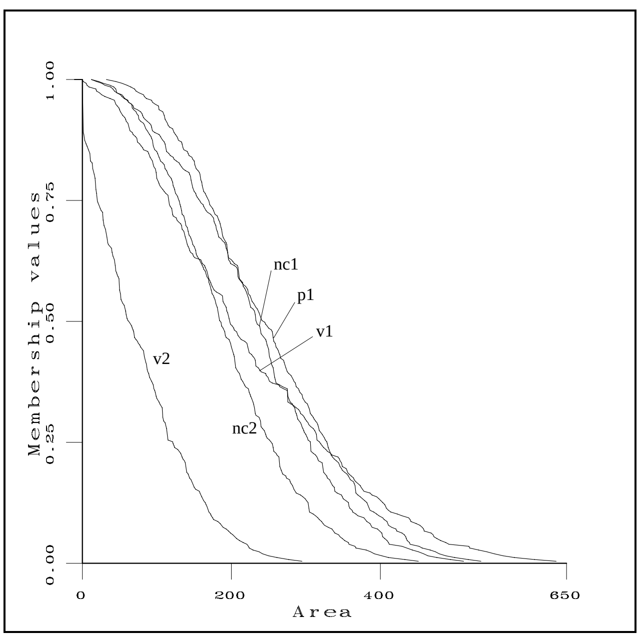


Figure 30: Area of fuzzy sets as fuzzy numbers, illustrated on the brain structures shown in Figure 20.

This definition of fuzzy geometrical measure applies for any measure. It is illustrated for the perimeter in Figures 31 and 32. Note that in this case, the variation of the perimeter of the α-cuts with respect to

α is no longer monotonous, and therefore the resulting fuzzy set is not necessarily convex.

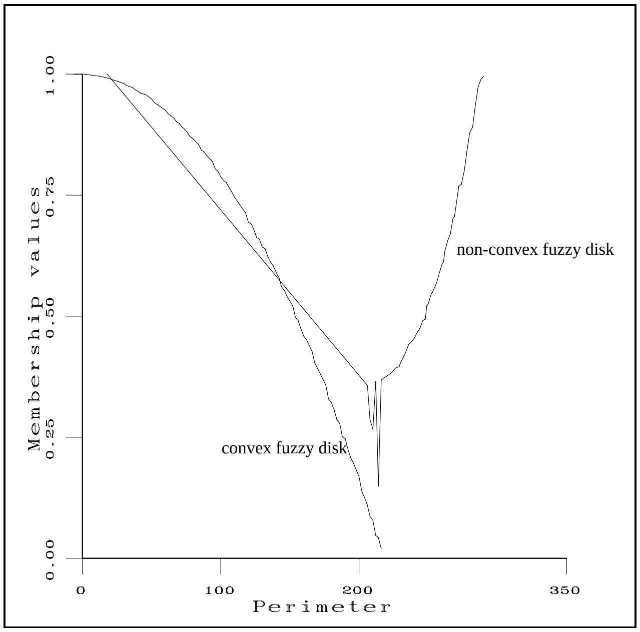


Figure 31: Perimeter of fuzzy sets as fuzzy numbers, illustrated on the two fuzzy disks shown in Figure 27.

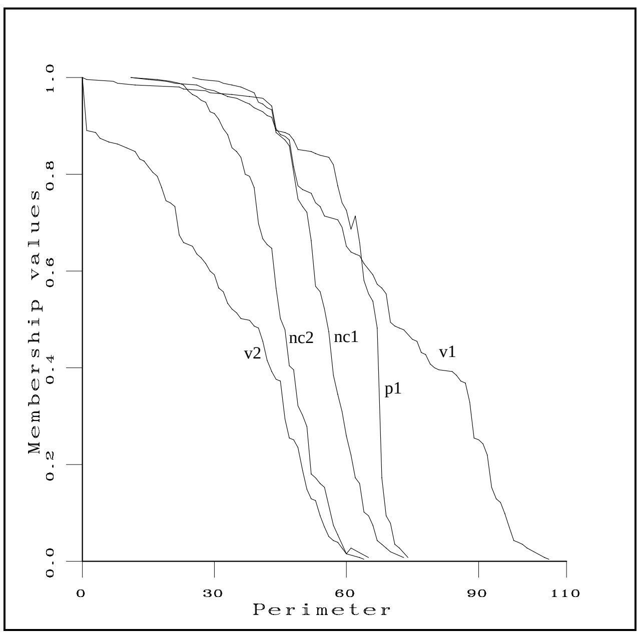


Figure 32: Perimeter of several fuzzy sets as fuzzy numbers, illustrated on the brain structures shown in Figure 20.

This form can be useful for further processing, like evaluating propositions such as "the measure is large, small, etc.", where "large" or "small" are values of linguistic variables. Such an evaluation can be performed by comparing the membership functions of the geometrical measure and of the linguistic variable, using pattern matching, or compatibility fuzzy set.

5.2.6 Crisp and fuzzy geometrical transforms

In this Section, we deal with geometrical transformations of fuzzy sets, like translation and rotation. We distinguish two cases, for well defined transformations, and for fuzzy transformations respectively.

Transformation of a fuzzy set by a crisp operation Let $\mathcal{T}$ be any geometrical transformation (translation, rotation, symmetry, scaling). Here we assume that $\mathcal{T}$ is well defined, i.e. the parameters that define the transformation are crisp. Let μ be a fuzzy object defined in $\mathcal{S}$ (in any dimension). The transformation of μ by $\mathcal{T}$ is then simply defined as:

$$\forall x \in \mathcal{S},\ \mathcal{T}(\mu)[\mathcal{T}(x)] = \mu(x), \tag{242}$$

or equivalently:

$$\forall x \in \mathcal{S},\ \mathcal{T}(\mu)(x) = \mu[\mathcal{T}^{-1}(x)], \tag{243}$$

In the discrete case, $\mathcal{T}(x)$ may not belong to $\mathcal{S}$ (i.e. may not coincide with one of the discrete points). The problem is similar to the one encountered in classical image processing, and similar interpolation methods can be used [140]. The transformed fuzzy set is then computed as:

$$\forall x \in \mathcal{S},\ \mathcal{T}(\mu)(x) = \mathit{Interpol}\{\mu(y), y \in V(\mathcal{T}^{-1}(x))\}, \tag{244}$$

where *Interpol* denotes any interpolation function, and $V(\mathcal{T}^{-1}(x))$ denotes a neighborhood of $\mathcal{T}^{-1}(x)$, composed of points of $\mathcal{S}$ (i.e. discrete points that are closed to the, possibly not discrete, point $\mathcal{T}^{-1}(x)$). Typically, $V(\mathcal{T}^{-1}(x))$ can be the nearest neighbor of $\mathcal{T}^{-1}(x)$, or the 4 closest points in 2D (or 8 closest points in 3D), etc..

Transformation of a fuzzy set by a fuzzy operation Now we assume that the transformation is imprecisely defined. Let $\mathcal{T}$ be the fuzzy transformation, depending on a set of parameters $p_1, ...p_n$, each one being fuzzy. Let μ_{p_i} be the membership function associated to the parameter p_i. We denote by $\mathcal{T}_{p_1,...p_n}$ the crisp transformation obtained for precise values of the parameters.

Let us first consider a point y in $\mathcal{S}$. Its transformation by $\mathcal{T}$ is a fuzzy set, the membership function of which is defined using the extension principle [42]:

$$\forall z \in \mathcal{S},\ \mu_{\mathcal{T}(y)}(z) = \sup_{p_1,...p_n \mid \mathcal{T}_{p_1,...p_n}(y)=z} t[\mu_{p_1}(p_1), ...\mu_{p_n}(p_n)], \tag{245}$$

where t is a t-norm[10].

This is interpreted as follows: for a given set of parameter values, the transformation of y by $\mathcal{T}_{p_1,...p_n}$ gives rise to a point $\mathcal{T}_{p_1,...p_n}(y)$ with membership values $t[\mu_{p_1}(p_1), ...\mu_{p_n}(p_n)]$. If several parameter values provide the same point, the supremum of the membership values is taken.

Let us now extend this definition for one point to a fuzzy set μ. Then the membership values of $\mu(y)$ have to be combined to the ones of $\mu_{\mathcal{T}(y)}$. This leads to the following definition of the transformation of μ [42]:

$$\forall x \in \mathcal{S},\ \mathcal{T}(\mu)(x) = \sup_{y \in \mathcal{S}} t[\mu(y), \mu_{\mathcal{T}(y)}(x)]. \tag{246}$$

In practice, the computation of such transformations is not always straightforward. One solution consists in applying directly the extension principle, which might be computationally heavy. Another solution could be to derive some analytical expressions using parametric representations of the fuzzy parameters defining the transformations. An example of such computation has been given in Section 4.8 for operations on L-R fuzzy numbers. However, such derivation is not straightforward for any type of operations, and for chained operations where the same variable can be involved several times.

[10] In such expressions $t(a, b, c)$ stands for $t[t(a, b), c]$. This notation is adopted for sake of simplicity and justified since any t-norm is commutative and associative. This extends to the combination of n values.

In the discrete case, another solution can be developed [42]: the definition domain of each parameter can be discretized, and for each possible value of the parameters, the corresponding point can be computed according to Equation 246, along with its membership degree.

5.3 Fuzzy mathematical morphology

Several operations have already been defined on spatial fuzzy sets. Among the operations often used in classical image processing, mathematical morphology operators gained a large importance. Therefore its extension to fuzzy sets is interesting. Moreover, mathematical morphology serves as a basis for defining several spatial relationships between objects, as will be seen later. Objects can be recognized in a scene using their own characteristics, but also, and this is very important, using their relationships to other objects of the scene. Extending mathematical morphology to fuzzy sets also leads to extensions of such relationships between fuzzy objects, and to applications in structural pattern recognition under imprecision.

Fuzzy mathematical morphology appears as a logical consequence of the research efforts that have been made to process imprecise spatial information. Indeed, since the introduction of fuzzy sets in 1965 by Zadeh [186], elementary set operations have been defined (intersection, union, complementation, inclusion, etc.) [186, 70, 181, 21], followed by several concepts useful in image processing and cognitive vision such as topological [145] or geometrical [146] operations, etc. These developments motivated further extensions of image processing operators to fuzzy sets, in particular mathematical morphology operators. Several definitions have been proposed since a few years. Some of them just consider grey level as membership functions [90, 89, 116, 62, 63, 132], or use binary structuring elements [146]. Here we restrict the presentation to really fuzzy approaches, where fuzzy sets have to be transformed according to fuzzy structuring elements. Initial developments can be found in the definition of fuzzy Minkowski addition [68, 104]. Then this problem has been addressed by several authors independently [34, 37, 38, 160, 4, 3, 5, 139, 162]

Attention will be paid here only on the 4 basic operations of mathematical morphology (erosion, dilation, opening, closing), but it should be clear for the reader that for every definition, a complete set of morphological operations could be derived.

For the sake of clarity, we call binary mathematical morphology (denoted by BMM) the morphology which uses binary sets and binary structuring elements; grey level mathematical morphology (GMM) will stand for classical morphology with a function (or a grey level image) and a binary structuring element, whereas functional mathematical morphology (FMM) will represent classical mathematical morphology on functions with functional structuring element. Links between FMM and BMM are obtained by considering the umbra or subgraph of the functions. Links between GMM and BMM comes from the fact that the binary dilations (for instance) of the thresholds of a function by a binary structuring element are equal to the thresholds of the dilation using GMM. We are interested in developing a fuzzy morphology (ΦMM). As said before, considering a fuzzy set as a grey level function does not allow to use FMM as a ΦMM (see Table 7) since, for instance, the dilation of a set defined in $[0,1]$ by a set defined in $[0,1]$ provides a new set defined in $[0,2]$.

Set	Structuring element	Classical MM	ΦMM
Binary	Binary	BMM	possibly compatible
Fuzzy	Binary	GMM	possibly compatible
Fuzzy	Fuzzy	FMM	*not compatible*

Table 7: As binary sets are special fuzzy sets, we may expect ΦMM to be compatible with classical mathematical morphology, when at least one of the two sets is binary. On the contrary, FMM cannot guarantee this internal property (except in very particular cases).

5.3.1 Definitions

Fuzzy morphology using the α-cuts Let us first consider the fuzzification method based on the integration of the equivalent crisp operation over all α-cuts [37, 35].

For two fuzzy sets μ and ν, the dilation of μ by ν is obtained by fuzzification over μ, followed by fuzzification over ν, or, equivalently, by the converse (see Equation 151):

$$D_\nu(\mu)(x) = \int_0^1 D_\nu(\mu_\beta)(x)d\beta = \int_0^1 \int_0^1 D_{\nu_\alpha}(\mu_\beta)(x)d\alpha d\beta \quad (247)$$

$$= \int_0^1 D_{\nu_\alpha}(\mu)(x)d\alpha = \int_0^1 \int_0^1 D_{\nu_\alpha}(\mu_\beta)(x)d\beta d\alpha. \quad (248)$$

A straightforward derivation provides:

$$\forall x \in \mathcal{S},\ D_\nu(\mu)(x) = \int_0^1 \sup_{y\in(\nu_\alpha)_x} \mu(y)d\alpha. \quad (249)$$

In the same way, the erosion of μ by ν is obtained by:

$$\forall x \in \mathcal{S},\ E_\nu(\mu)(x) = \int_0^1 \inf_{y\in(\nu_\alpha)_x} \mu(y)d\alpha, \quad (250)$$

where $(\nu_\alpha)_x$ denotes the translation of ν_α at x.

These definitions guarantee that $D_\nu(\mu)$ and $E_\nu(\mu)$ are the membership functions of fuzzy sets (i.e. taking values in $[0, 1]$). By construction, if ν is crisp, the definitions coincide with the classical ones (in the GMM framework).

Using the other fuzzification principles (Equations 148 and 149), we obtain respectively:

$$\forall x \in \mathcal{S},\ D_\nu(\mu)(x) = \sup_{y\in\mathcal{S}} \min[\mu(y), \nu(y-x)], \quad (251)$$

$$\forall x \in \mathcal{S},\ E_\nu(\mu)(x) = \inf_{y\in\mathcal{S}} \max[\mu(y), 1-\nu(y-x)]. \quad (252)$$

and:

$$\forall x \in \mathcal{S},\ D_\nu(\mu)(x) = \sup_{y\in\mathcal{S}}[\mu(y)\nu(y-x)], \quad (253)$$

$$\forall x \in \mathcal{S},\ E_\nu(\mu)(x) = \inf_{y\in\mathcal{S}}[\mu(y)\nu(y-x) + 1 - \nu(y-x)]. \quad (254)$$

The proofs of these derivations can be found in [14].

The extension principle applied to the α-cuts (Equation 153) leads also to:

$$\forall x \in \mathcal{S},\ D_\nu(\mu)(x) = \sup_{y\in\mathcal{S}} \min[\mu(y), \nu(y-x)], \quad (255)$$

$$\forall x \in \mathcal{S},\ E_\nu(\mu)(x) = \inf_{y\in\mathcal{S}} \max[\mu(y), 1-\nu(y-x)]. \quad (256)$$

These equations happen to be particular cases of the definitions obtained by formal translation, as will be seen in the next paragraph.

Fuzzy morphology by formal translation based on t-norms and t-conorms Another way to build ΦMM is to base the initial definitions on the translation of set operations (inclusion, intersection, union) into functional terms, i.e. to exhibit, for each operation, functions from $[0,1]\times[0,1]$ to $[0,1]$ which satisfy some given limit conditions expressing the compatibility with binary set operations. Several definitions exist, that may receive a unified presentation by means of t-norms (and associated t-conorms) as they can be interpreted respectively as fuzzy intersection and fuzzy union (from which a fuzzy inclusion can be derived). This follows exactly the formal translation principle described above. The corresponding definitions have been proposed in [15, 38], and coincide with the definitions independently developed in [5].

The construction calls for formal translation of intersection and inclusion. From the following set equivalence (where $E_B(X)$ denotes the erosion of the set X by B):

$$x \in E_B(X) \Leftrightarrow B_x \subset X \Leftrightarrow X \cup B_x^C = \mathcal{S} \Leftrightarrow \forall y \in \mathcal{S},\ y \in X \cup B_x^C, \tag{257}$$

a natural way to define the erosion of a fuzzy set μ by a fuzzy structuring element ν is:

$$\forall x \in \mathcal{S},\ E_\nu(\mu)(x) = \inf_{y\in\mathcal{S}} T[c(\nu(y-x)), \mu(y)]. \tag{258}$$

In this equation, the union $\cup$ has been translated in terms of a t-conorm T and the universal symbol $\forall$ by an infimum. This corresponds to the extension of inclusion described before.

By duality with respect to the complementation c, fuzzy dilation is then defined by:

$$\forall x \in \mathcal{S},\ D_\nu(\mu)(x) = \sup_{y\in\mathcal{S}} t[\nu(y-x), \mu(y)], \tag{259}$$

where t is the t-norm associated to the t-conorm T with respect to the complementation c .

This definition of dilation corresponds to the following set equivalence:

$$x \in D_B(x) \Leftrightarrow B_x \cap X \neq \emptyset \Leftrightarrow \exists y \in \mathcal{S},\ y \in B_x \cap X. \tag{260}$$

Here, intersection $\cap$ has been translated in terms of a t-norm t and the existential symbol by a supremum.

This form of fuzzy dilation and fuzzy erosion are very general, and several definitions found in the literature appear as particular cases: for instance, the definitions proposed in [34, 15, 5, 4], correspond to Equations 258 and 259, for $t = \min$ and $T = \max$:

$$\forall x \in \mathcal{S},\ D_\nu(\mu)(x) = \sup_{y\in\mathcal{S}} \min[\mu(y), \nu(y-x)], \tag{261}$$

$$\forall x \in \mathcal{S},\ E_\nu(\mu)(x) = \inf_{y\in\mathcal{S}} \max[\mu(y), 1-\nu(y-x)]. \tag{262}$$

The definitions introduced in [34] correspond to Equations 258 and 259, for the product as t-norm and the algebraic sum as t-conorm:

$$\forall x \in \mathcal{S},\ D_\nu(\mu)(x) = \sup_{y\in\mathcal{S}} [\mu(y)\nu(y-x)], \tag{263}$$

$$\forall x \in \mathcal{S},\ E_\nu(\mu)(x) = \inf_{y\in\mathcal{S}} [\mu(y)\nu(y-x) + 1 - \nu(y-x)]. \tag{264}$$

The definitions introduced in [160] correspond to Equations 258 and 259, for Lukasiewicz t-norm and t-conorm:

$$\forall x \in \mathcal{S},\ D_\nu(\mu)(x) = \sup_{y\in\mathcal{S}} \max[0, \mu(y) + \nu(y-x) - 1], \tag{265}$$

$$\forall x \in \mathcal{S},\ E_\nu(\mu)(x) = \inf_{y\in\mathcal{S}} \min[1, 1 + \mu(y) - \nu(y-x)]. \tag{266}$$

It is straightforward to see that the definitions proposed in previous works for particular cases [146, 104, 90] fall within the general definitions using t-norm and t-conorm.

Finally, fuzzy opening (respectively fuzzy closing) is simply defined as the combination of a fuzzy erosion followed by a fuzzy dilation (respectively a fuzzy dilation followed by a fuzzy erosion), by using dual t-norms and t-conorms.

Using weak t-norms and t-conorms If we replace, in the previous construction, the t-norm by a weak t-norm, and the t-conorm by a weak t-conorm (which are not associative and do not admit 1 (respectively 0) as unit element, in general), Equations 258 and 259 appear as a generalization of the definitions proposed in [161]:

$$\forall x \in \mathcal{S},\ D_\nu(\mu)(x) = \sup_{y \in \mathcal{S}} \max[0, 1 - \lambda(\mu(y)) - \lambda(\nu(y-x))], \tag{267}$$

$$\forall x \in \mathcal{S},\ E_\nu(\mu)(x) = \inf_{y \in \mathcal{S}} \min[1, \lambda(1 - \mu(y)) + \lambda(\nu(y-x))], \tag{268}$$

with λ a function from $[0,1]$ to $[0,1]$ satisfying the 6 following conditions:

1. $\lambda(z)$ is a non-increasing function of z,
2. $\lambda(0) = 1$,
3. $\lambda(1) = 0$,
4. the equation $\lambda(z) = 0$ has a single solution,
5. $\forall \alpha \in [0.5, 1]$, the equation $\lambda(z) = \alpha$ has a single solution,
6. $\forall z \in [0,1], \lambda(z) + \lambda(1-z) \geq 1$.

The weak t-norm and t-conorm involved in these definitions are:

$$\forall (a,b) \in [0,1]^2, t(a,b) = \max[0, 1 - \lambda(a) - \lambda(b)], \tag{269}$$

$$\forall (a,b) \in [0,1]^2, T(a,b) = \min[1, \lambda(1-a) + \lambda(1-b)]. \tag{270}$$

The case where $\lambda(z) = 1 - z = \lambda_0(z)$ leads to a true t-norm and t-conorm. The corresponding definitions for fuzzy dilation and erosion exactly correspond to Equations 259 and 258 for the Lukasiewicz operators (i.e. Equations 265 and 266).

We will see in Section 5.3.2 that the definitions using weak t-norm and t-conorm have weaker properties than the ones using true t-norm and t-conorm.

Fuzzy morphology using residual implications Fuzzy implication is often defined as [72]:

$$Imp(a,b) = T[c(a), b)]. \tag{271}$$

Fuzzy inclusion is related to implication by the following equation:

$$\mathcal{I}(\nu, \mu) = \inf_{x \in \mathcal{S}} Imp[\nu(x), \mu(x)], \tag{272}$$

which allows to relate directly fuzzy erosion to fuzzy implication, leading to the general definition using t-conorm, and by duality also fuzzy dilation.

This suggests another way to define fuzzy erosion (and dilation), by using other forms of fuzzy implication. One interesting approach is to use residual implications:

$$Imp(a,b) = \sup\{\varepsilon \in [0,1], t(a,\varepsilon) \leq b\}. \tag{273}$$

This provides the following expression for the degree of inclusion:

$$\mathcal{I}(\nu, \mu) = inf_{x \in \mathcal{S}} \sup\{\varepsilon \in [0,1], t(\nu(x), \varepsilon) \leq \mu(x)\}. \tag{274}$$

This definition coincides with the previous one if t is an Archimedian t-norm with nilpotent elements, typically Lukasiewicz t-norm: $t(a,b) = \max(0, a+b-1)$ (the corresponding t-conorm being $T(a,b) = \min(1, a+b)$).

The derivation of fuzzy morphological operators from residual implication has been proposed in [3]. One of its main advantages is that it leads to idempotent fuzzy closing and opening.

5.3.2 Properties

The detail of properties for various definitions can be found in [38], with the main proofs in [14]. We summarize here the main properties. The general definition using t-norms and t-conorms satisfies most properties:

- erosion and dilation (respectively opening and closing) are dual with respect to the complementation c;
- if the structuring element is binary, the same definitions as in GMM are obtained;
- compatibility with translations;
- local knowledge property;
- continuity if the t-norm is continuous (which is most often the case);
- increasingness of all operations with respect to inclusion;
- extensivity of dilation and anti-extensivity of erosion iff $\nu(0) = 1$ (this corresponds to the condition that the origin should belong to the structuring element in the crisp case);
- extensivity of closing, anti-extensivity of opening and idempotence of these two operations iff $t[b, u(c(b), a)] \leq a$, which is satisfied for Lukasiewicz t-norm and t-conorm;
- commutation with union of dilation (with intersection for erosion);
- iteration property of dilation.

For the definitions obtained by integration over the α-cuts, extensivity of closing, anti-extensivity of opening, idempotence of opening and closing, commutation with union of dilation (with intersection for erosion), and iteration property of dilation are lost, the other properties being kept.

For the definitions derived from weak t-norm and t-conorm, several properties hold only for the case where $\lambda = \lambda_0$.

For the definitions derived from a residual implication, opening and closing are idempotent and anti-extensive (respectively extensive).

5.3.3 Examples

A few examples of fuzzy dilations and erosions are illustratred in Figure 33, on synthetical bidimensionnal fuzzy objects.

Let us now briefly describe a real application, where fuzzy dilation is used to model spatial imprecisions. The problem concerns 3D reconstruction of blood vessels by fusion of angiographic and echographic data [42, 136]. The aim of the fusion is to avoid the limits inherent to each modality: 3D reconstruction from two angiographic projections is imprecise because of the hypotheses that have to be done, and reconstruction from endovascular echographic slices is limited by physical factors and also calls for simplifying hypotheses that induce imprecisions. Figure 34 provides an example of data.

The fusion of both modalities requires to register all data in the same coordinate frame. For instance, the position of each echographic slice is defined by three translation parameters and three rotation parameters. Their estimation is performed with the help of the control radiographies. The reconstruction of both modalities in the common reference frame leads to contradictions between them, due to the imprecision on the estimated parameters. Instead of considering them as numbers, they can be modeled as fuzzy numbers. Taking all imprecisions into account for the six parameters leads to the definition of fuzzy structuring elements representing the possible locations of each point of the vessel surface. Fuzzy dilation appears then as an appropriate tool to introduce these imprecisions in a controlled way, by

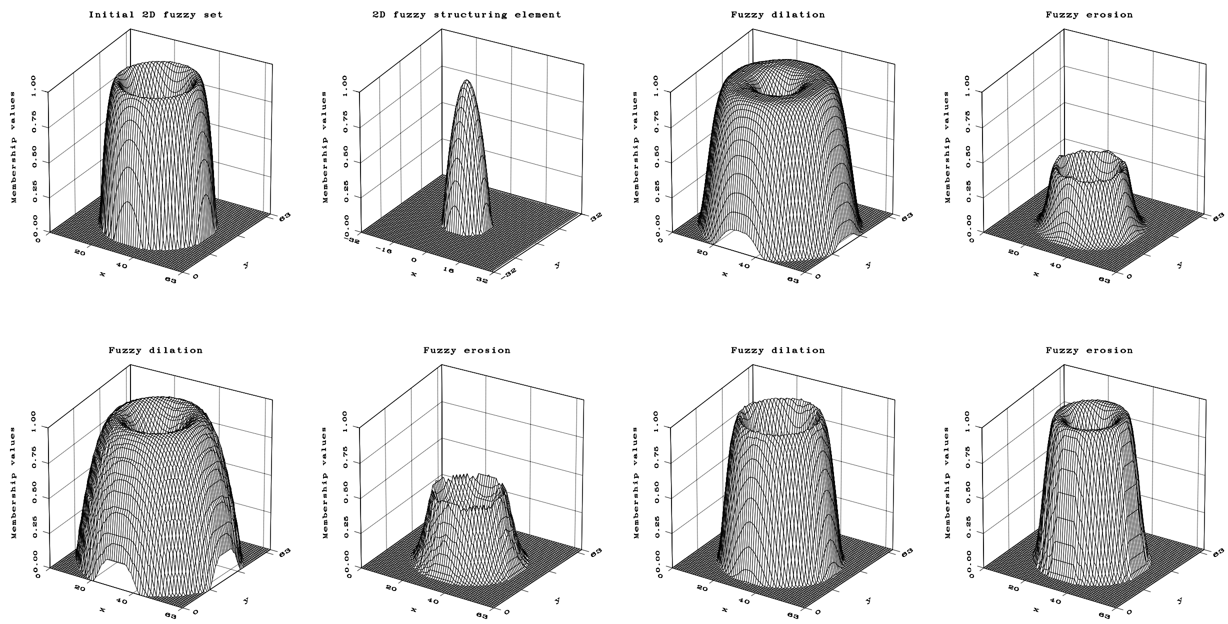


Figure 33: Illustrations of some definitions of fuzzy dilations and erosions on a 2D example. First line: fuzzy set and structuring element defined on $\mathbb{R}^2$, dilation and erosion by integration over the α-cuts; second line: dilation and erosion with the t-norm min and the t-conorm max, and with the Lukasiewicz operators.

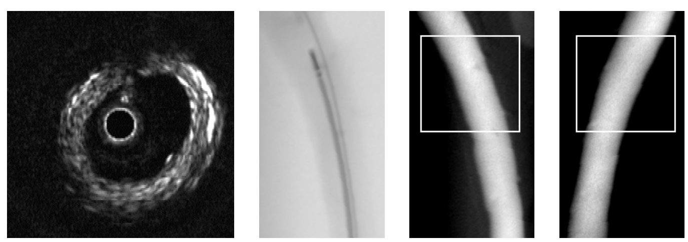


Figure 34: Images used for the reconstruction. Echographic slice, control radiography (used to estimate the position and orientation of the echographic slices), left and right angiographies. The frame indicates the area in which reconstruction will be performed.

keeping good properties. The fact that dilation commutes with union allows us to process the surface globally as a set of points, and the iteration property allows us to introduce successively imprecisions in rotation and in translation. Let V_{bin} be the binary surface (without imprecisions), and ν_1^x the imprecision in rotation at point x (indeed these imprecisions are not spatially invariant). Then a first fuzzy vlume is obtained by:

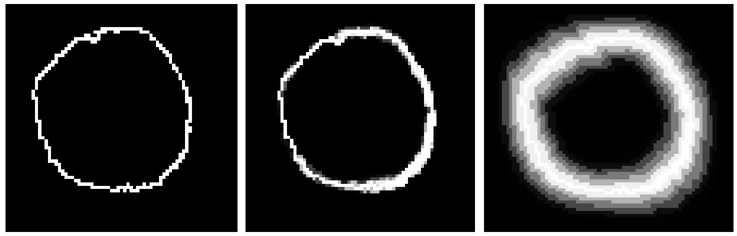
$$\mu_{V_f'}(x) = \sup\{\nu_1^y(x) \mid y \in V_{bin}\}.$$

In a second step, a dilation by the fuzzy structuring element ν_2 representing the imprecisions in translation (which are spatially invariant) is performed and provides:

$$V_f = \bigcup \{D_{D_{\nu_2}(\nu_1^x)}(\{x\}) \mid x \in V_{bin}\} = D_{\nu_2}(V_f').$$

Figure 35 shows an axial slice of the reconstructed vessel at different steps (the method is actually applied directly in 3D).

Once this procedure is applied on both modalities, the two reconstructions contain explicitly all imprecisions and are no more conflicting. A conjunctive fusion, typically using the min operator, can



(a) (b) (c)

Figure 35: Axial echographic slice of the reconstructed vessel. (a) Binary reconstruction. (b) Fuzzy reconstruction including rotation imprecisions. (c) Fuzzy reconstruction after dilation by ν_2 (this reconstruction includes all imprecisions).

then be performed, in order to find a consensus between both modalities and to reduce the imprecisions. The final binary decision is obtained by a watershed surface (3D extension of the classical watershed), which provides a surface with the required topology (here a cylindrical one) and going through the points with the highest membership values to the combined volume. The result is illustrated in Figure 36. Contradictions between both modalities are solved in a satisfactory manner: in areas where there was no conflict, the same result is obtained, while in areas where the two binary reconstructions were conflicting, an intermediate position is obtained.

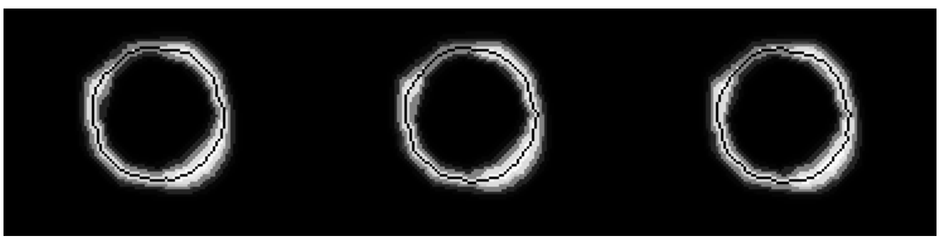


Figure 36: Superimposition of the fuzzy volume after fusion and of the watershed surface (in black) on a few slices of the vascular segment.

5.3.4 Conclusion and extensions

From the basic transformations presented above, a lot of other ones can be built, such as geodesic operations [27], filters [162], operations by reconstruction, convexity measures [138], etc. Various applications appear in the literature, either in vision, or in other domains. For instance, applications in data analysis have been developed in [169]. Duality properties of the morphological operations allows us to use them to define dual operators in other domains, such as in belief function theory [22], in rough sets theory [28], ot in logics [33, 30]. In vision and structural pattern recognition, fuzzy morphology also provides powerful tools for defining spatial relations in imprecise situations [26], as will be seen in the next sections.

5.4 Fuzzy topology

This Section is dedicated to simple fuzzy topological notions, such as connectivity or adjacency, which are very important in vision and image processing, both locally in the neighborhood of a point and globally when spatial relationships between objects have to be assessed.

5.4.1 Fuzzy connectivity and neighborhood

The notion of connectivity is the core of several transformations, of measures, of analysis and recognition methods. Their fuzzy extension therefore calls for a fuzzy extension of connectivity. Moreover, it is useful for travelling from one point to the other in a fuzzy set, which can be exploited for defining fuzzy geodesic distances [17] and geodesic morphological operations [27].

Several definitions of fuzzy connectivity have been proposed, which mainly differ by their behavior in the case of plateaus [146, 171, 172]. We present here the definition of [146], and assume that the space $\mathcal{S}$ is a digital bounded space, on which a digital connectivity is defined.

The degree of connectivity between two points x and y in a fuzzy object μ in a finite discrete space is defined as [146]:

$$c_\mu(x,y) = \max_{L_{xy}} \min_{t_i \in L_{xy}} \mu(t_i) \tag{275}$$

where L_{xy} is any path from x to y. It takes values in $[0,1]$. By considering the membership degrees as a relief on $\mathcal{S}$, $c_\mu(x,y)$ corresponds to the minimal altitude on the path from x to y by descending at least as possible.

The notion of fuzzy connected component associated to a point x is defined as the following fuzzy set:

$$\forall y \in \mathcal{S},\ \Gamma^x_\mu(y) = c_\mu(x,y). \tag{276}$$

The degree of connectivity has the following properties:

1. c_μ is weakly reflexive: $\forall (x,y) \in \mathcal{S}^2,\ c_\mu(x,x) \geq c_\mu(x,y)$ and $c_\mu(x,x) = \mu(x)$,
2. c_μ is symmetrical,
3. c_μ is transitive, in the sense of the max-min transitivity: $\forall (x,z) \in \mathcal{S}^2,\ c_\mu(x,z) \geq \max_{y \in E}[\min(c_\mu(x,y), c_\mu(y,z))]$,
4. $c_\mu(c,y)$ is obtained by descending as least as possible in the membership values when going from x to y, and is linked to the membership values of x and y, in particular: $c_\mu(x,y) \leq \min[\mu(x), \mu(y)]$.

In the above definitions, the neighborhood is the one of the digital topology defined on $\mathcal{S}$. Fuzzy neighborhoods can also be defined, as well as a degree n_{xy} to which two points x and y are neighbors. Several definitions have been proposed [60, 40], which are typically decreasing functions of the distance between both points:

$$n_{xy} = \frac{1}{1+d(x,y)}, \text{ or } n_{xy} = \frac{1+\exp(-b)}{1+\exp b(\frac{d(x,y)-1}{S} - 1)}, \tag{277}$$

where b and S are two positive parameters which control the shape of the curve. Other functions can be used, like S-functions for instance.

5.4.2 Boundary of a fuzzy object

If an object is fuzzy, it is natural to consider that its boundary is fuzzy too. The definition can be simply obtained by translating the expression "a point belongs to the (internal) boundary of a set if and only if it belongs to the set and it has at least one neighbor outside the set". The internal fuzzy boundary of a fuzzy set μ is then defined by the membership function b^i_μ as:

$$\forall x \in \mathcal{S}, b^i_\mu(x) = t[\mu(x), \sup_{z \in \mathcal{S}} t[c(\mu)(z), n_{xz}]], \tag{278}$$

where t is a t-norm, c a fuzzy complementation, and n_{xz} the degree to which x and z are neighbors.

The fuzzy boundary of a fuzzy set μ can also be expressed in morphological terms, since the points of the boundary are the one that are not in its erosion (or equivalently which are in the dilation of its complement). It is then defined from fuzzy dilation as:

$$b^i_\mu(x) = t[\mu(x), D_{B_c}(c(\mu))(x)] \tag{279}$$

where B_c is either the elementary structuring element or a fuzzy structuring element.

External boundary can be defined in a similar way.

5.4.3 Adjacency between two fuzzy objects

Rosenfeld and Klette [149] define a degree of adjacency between two crisp sets, using a geometrical approach based on the notion of "visibility" of a set from another one. This definition is then extended to degree of adjacency between two fuzzy sets. However, this definition is not symmetrical, and probably not easy to transpose to higher dimensions. Another approach consists in using the notion of contours, frontiers, and neighborhood [60, 40]. We present here this second approach. The space $\mathcal{S}$ is endowed with a discrete connectivity c.

In the crisp discrete case, two image regions X and Y are adjacent if:

$$X \cap Y = \emptyset \ \textit{and} \ \exists x \in X, \exists y \in Y : n_c(x,y), \tag{280}$$

where $n_c(x,y)$ is the Boolean variable stating that x and y are neighbors in the sense of the discrete c-connectivity.

A consequence of this definition is that, if X and Y are adjacent, then any $x \in X$ and $y \in Y$ that satisfy $n_c(x,y)$ belong to the boundary of X and Y respectively.

Therefore the fuzzy extension of definition 280 can be obtained either by considering only the constraint on the neighborhood, or by considering also the constraint on the boundary

Definition 280 can also be expressed equivalently in terms of morphological dilation, as:

$$X \cap Y = \emptyset \ \textit{and} \ D_{B_c}(X) \cap Y \neq \emptyset, \ \ D_{B_c}(Y) \cap X \neq \emptyset, \tag{281}$$

where $D_{B_c}(X)$ denotes the dilation of X by the structuring element B_c, since the following equivalence holds (for $x \neq y$):

$$n_c(x,y) \Leftrightarrow x \in B_{cy} \Leftrightarrow y \in B_{cx}, \tag{282}$$

where B_{cx} denotes the structuring element B_c translated at point x.

This will provide a third way to extend the definition to fuzzy sets, either directly from fuzzy dilation, or by means of distance computation, which is closely related to dilation (equation 281 means that the minimum (nearest point) distance between X and Y is equal to 1).

Using neighborhood constraint The extension of this definition, as detailed in [40], involves the definitions of a degree of intersection $\mu_{int}(\mu,\nu)$ between two fuzzy sets μ and ν defined on $\mathcal{S}$, as well as a degree of non-intersection $\mu_{\neg int}(\mu,\nu)$, and a degree of neighborhood n_{xy} between two points x and y of $\mathcal{S}$. All these notions have been introduced previously. This leads to the following definition for fuzzy adjacency between μ and ν:

$$\mu_{adj}(\mu,\nu) = t[\mu_{\neg int}(\mu,\nu), \sup_{x\in\mathcal{S}} \sup_{y\in\mathcal{S}} t[\mu(x), \nu(y), n_{xy}]]. \tag{283}$$

This definition is symmetrical, consistent with the discrete binary definition (i.e. in the case where μ and ν are crisp and $n_{xy} = n_c(x,y)$), and decreasing with respect to the distance between the two fuzzy sets. It is invariant with respect to geometrical transformations (for scaling, only if n_{xy} is itself invariant).

Figure 37 illustrates the results obtained with definition 283 with the t-norm minimum. Using the maximum of the intersection we obtain $\mu_{adj}(\mu,\nu) = 0.36$ and $\mu_{adj}(\mu,\nu') = 0.35$, which are very similar values. On the contrary, using the fuzzy hypervolume to define the degree of intersection, definition 283 accounts for the differences in intersection and provides $\mu_{adj}(\mu,\nu) = 0.67$ and $\mu_{adj}(\mu,\nu') = 0.34$, which are this time very different.

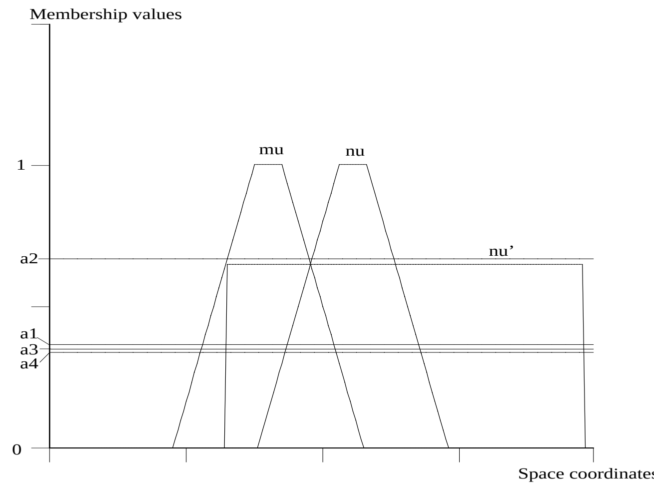


Figure 37: Illustration of definition 283 when using different definitions for the degree of intersection. Using the maximum of the intersection we obtain $\mu_{adj}(\mu,\nu) = \alpha_1(=0.36)$ and $\mu_{adj}(\mu,\nu') = \alpha_3(=0.35)$, and using the fuzzy hypervolume $\mu_{adj}(\mu,\nu) = \alpha_2(=0.67)$ and $\mu_{adj}(\mu,\nu') = \alpha_4(=0.34)$.

Adding boundary constraints The translation of definition 280 along with the property on boundary leads now to the following definition of the degree of adjacency between μ and ν:

$$\mu_{adj}(\mu,\nu) = t\left[\mu_{\neg int}(\mu,\nu), \sup_{x\in\mathcal{S}}\sup_{y\in\mathcal{S}} t[b_\mu(x), b_\nu(y), n_{xy}]\right]. \tag{284}$$

This definition has the same properties as the previous one.

Using fuzzy morphological operators The degree of adjacency between μ and ν involving fuzzy dilation is defined as:

$$\mu_{adj}(\mu,\nu) = t\left[\mu_{\neg int}(\mu,\nu), \mu_{int}[D_{B_c}(\mu),\nu], \mu_{int}[D_{B_c}(\nu),\mu]\right]. \tag{285}$$

This definition represents a conjunctive combination of a degree of non-intersection between μ and ν and a degree of intersection between one fuzzy set and the dilation of the other. B_c can be taken as the elementary structuring element related to the considered connectivity, or as a fuzzy structuring element, representing for instance spatial imprecision (i.e. the possibility distribution of the location of each point). Again the same properties are satisfied.

The adjacency degrees between some of the fuzzy objects shown in Figure 20 are given in Table 8. The results are in agreement with what can be expected from the model (crisp adjacency between atlas objects). In this case, crisp adjacency would provide completely different results in the model and in the image, preventing its use for recognition. This suggests that fuzzy adjacency degree can indeed be used for pattern recognition purposes, of course combined with other spatial relationships.

Fuzzy object 1	Fuzzy object 2	degree of adjacency	adjacency in the model (crisp)
v1	v2	0.368	1
v1	cn1	0.463	1
v1	p1	0.000	0
v1	cn2	0.035	0
v2	cn2	0.427	1
cn1	p1	0.035	0

Table 8: Results obtained for fuzzy adjacency. Labels of structures are given in Figure 20. High degrees are obtained between structures where adjacency is expected, while very low degrees are obtained in the opposite case.

5.5 Distances

The importance of distances in image processing and cognitive vision is well established. Their extensions to fuzzy sets can be useful in several parts of image processing under imprecision. Let us mention a few possible applications of these distances for problems where imprecision has to be taken into account. Distance from a point to a fuzzy set can be used for classification purposes, where a point has to be attributed to the nearest fuzzy class. When considering distance from a point to the complement of a fuzzy set μ, we obtain the basic information for computing a fuzzy skeleton of μ. Mean distance is useful for registration: if we want to register a fuzzy set with respect to another one, we may use this distance as a minimization criterion, that can be optimized over all possible positions (typically translation and rotation) of the one fuzzy set with respect to the other.

5.5.1 Representations

The most used representation of a distance between two fuzzy sets is as a number d, taking values in $\mathbb{R}^+$ (or more specifically in $[0, 1]$ for some of them). However, since we consider fuzzy sets, i.e. objects that are imprecisely defined, we may expect that the distance between them is imprecise too. This argument is advocated in particular in [69] and [147]. Then the distance is better represented as a fuzzy set, and more precisely as a fuzzy number (a convex upper semi-continuous fuzzy set on $\mathbb{R}^+$ having a bounded support).

In [147], Rosenfeld defines two concepts that will be used in the sequel. One is distance density, denoted by $\delta(\mu, \nu)$, and the other distance distribution, denoted by $\Delta(\mu, \nu)$, both being fuzzy sets on $\mathbb{R}^+$. They are linked together by the following relation:

$$\Delta(\mu, \nu)(n) = \int_0^n \delta(\mu, \nu)(n')dn'. \quad (286)$$

While the distance distribution value $\Delta(\mu, \nu)(n)$ represents the degree to which the distance between μ and ν is less than n, the distance density value $\delta(\mu, \nu)(n)$ represents the degree to which the distance is equal to n.

Finally, the concept of distance can be represented as a linguistic variable. This assumes a granulation [185] of the set of possible distance values into symbolic classes such as "near", "far", etc., each of these classes being defined as a fuzzy set. This approach has been drawn e.g. in [8, 114, 13].

Although we may speak about distances between image objects in a very general way, this expression does not make necessarily the assumption that we are dealing with true metrics. For several applications in image processing, it is not sure that all properties are needed. An important use of distances is related to the comparison of shapes, which reinforces the interest of deriving distances from similarities. The concept of similarities between objects, in particular image objects, contains some subjective aspects. As already stated by Poincaré at the beginning of the century, and underlined by several authors in the fuzzy sets domain (see e.g. [101, 73]), subjective similarities does not require to be transitive. This induces a loss of triangular inequality in the derived distance. In cognitive vision, typically for applications where image objects have to be compared to models, the triangular inequality is of no use, since the two arguments of the distance function belong to two different sets of objects. For such applications, semi-metrics or even semi-pseudometrics may be sufficient.

5.5.2 Distance from a point to a fuzzy object

As a number Distances from a point to a fuzzy set can be defined using a weighting approach or using a fuzzification from α-cuts. In this way, they are defined as numbers. The idea in the weighting approach is that a point that has a low membership value to μ should have less influence in the computation of the infimium (or minimum). Therefore the distance between x and μ may be defined as:

$$d(x, \mu) = \inf_{y \in \mathcal{S}} [d_{\mathcal{S}}(x, y) f(\mu(y))], \tag{287}$$

where f is a decreasing function of μ (e.g. $f(\mu(y)) = \frac{1}{\mu(y)}$) such that $f(1) < +\infty$ (in order to guarantee that if x belongs completely to μ, i.e. if $\mu(x) = 1$, the distance is attained for $y = x$), and with the convention $0 f(0) = +\infty$. If $\mu(x) = 0$, i.e. if x is completely outside of μ, this definition leads to satisfactory results. However, if $\mu(x) > 0$, it leads always to 0, on the whole support of μ. This can be seen as a strong drawback of this definition, since we would intuitively rather expect that $d(x, \mu)$ depends on the membership degree of x to μ.

Generally speaking, it is required that $d(x, \mu)$ be a strictly decreasing function of $\mu(x)$, with $d(x, \mu) = 0$ if $\mu(x) = 1$.

Defining a fuzzy function from its crisp equivalent applied on the α-cuts is a very common way to proceed, which has already been used for defining several operations on fuzzy sets [65]. The two following equations express different combinations of the α-cuts for defining $d(x, \mu)$:

$$d(x, \mu) = \int_0^1 d(x, \mu_\alpha) d\alpha, \tag{288}$$

$$d(x, \mu) = \sup_{\alpha \in]0,1]} [\alpha d(x, \mu_\alpha)]. \tag{289}$$

The first one consists in "stacking" the results obtained on each α-cut, while the second one consists in weighting these results by the level of the cut, $d(x, \mu_\alpha)$ being the classical distance from a point to a crisp set.

Equation 289 does not lead to convenient results, since the obtained distance is always the distance from x to the core of μ, i.e. $d(x, \mu) = d(x, \mu_1)$, and therefore does not depend on $\mu(x)$ if $\mu(x) \neq 1$.

Equation 288 does not share the same disadvantage, since all α-cuts are explicitely involved in the result. For instance for μ and ν having the same core and $\mu(x) > \nu(x)$, we have $d(x, \mu) < d(x, \nu)$ which is intuitively desired, while equation 289 provides the same distance value.

As a fuzzy number an original approach for defining the distance $d(x, \mu)$ from a point x of $\mathcal{S}$ to a fuzzy object μ as a fuzzy number has been proposed in [16], by translating crisp equations into their fuzzy equivalent. Standard expressions for the distance involve concepts that are not set theoretical ones, and are therefore not trivial to translate. Therefore, the easiest way to perform this translation is to find a formalism where distances are expressed in set theoretical terms. This formalism is provided by mathematical morphology, since the distance from a point to a set can be expressed in terms of morphological dilation, as well as several distances between two sets. The translation of dilation in fuzzy terms can be achieved with good properties using the framework of fuzzy mathematical morphology we developed in [38].

In the crisp case, and in a finite discrete space, we have respectively for $n = 0$ and for $n > 0$:

$$d(x, X) = 0 \quad \Leftrightarrow \quad x \in X \tag{290}$$

$$d(x, X) = n \quad \Leftrightarrow \quad x \in D^n(X) \text{ and } x \notin D^{n-1}(X) \tag{291}$$

where D^n denotes the dilation by a ball of radius n centered at the origin of $\mathcal{S}$ (and $D^0(X) = X$) (see e.g. [44] for a study of discrete balls and discrete distances in the crisp case). In this case, the extensivity property of the dilation holds [157], and $x \notin D^{n-1}(X)$ is equivalent to $\forall n' < n, x \notin D^{n'}(X)$. Equation 291 is equivalent to:

$$x \in D^n(X) \cap [D^{n-1}(X)]^C, \tag{292}$$

where A^C denotes the complement set of A in $\mathcal{S}$. This is a pure set theoretical expression, that we can now translate into fuzzy terms. This leads to the following definition of the degree to which $d(x, \mu)$ is equal to n:

$$\delta_{(x,\mu)}(0) = \mu(x), \tag{293}$$

$$\delta_{(x,\mu)}(n) = t[D^n_\nu(\mu)(x), c[D^{n-1}_\nu(\mu)(x)]], \tag{294}$$

where t is a t-norm (fuzzy intersection), c a fuzzy complementation (typically $c(a) = 1 - a$ for $a \in [0, 1]$), and ν a fuzzy structuring element used for performing the dilation. Several choices of ν are possible. It can be simply the unit ball, or a fuzzy set representing for instance the smallest sensitive unit in the image, along with the imprecision attached to it. In this case, ν has to be equal to 1 at the origin of $\mathcal{S}$, such that the extensivity of the dilation still holds [38].

The properties of this definition are the following [16]:

- if $\mu(x) = 1$, $\delta_{(x,\mu)}(0) = 1$ and $\forall n > 0, \delta_{(x,\mu)}(n) = 0$, i.e. the distance is a crisp number in this case;
- if μ and ν are binary, the proposed definition coincides with the binary one;
- the fuzzy set $\delta_{(x,\mu)}$ can be interpreted as a density distance, from which a distance distribution can be deduced by integration;
- finally, $\delta_{(x,\mu)}$ is a non normalized fuzzy number (in the discrete finite case).

Figure 38 presents an example of fuzzy numbers $\delta_{(x,\mu)}(n)$ obtained for different points, the spatial domain being reduced to a one dimensional space in this example. The point x_1 is outside the support of μ and at a larger distance from it than x_2. The results correspond to the intuition, since the fuzzy number $\delta_{(x_2,\mu)}(n)$ is more concentrated around very small values of n than $\delta_{(x_1,\mu)}(n)$.

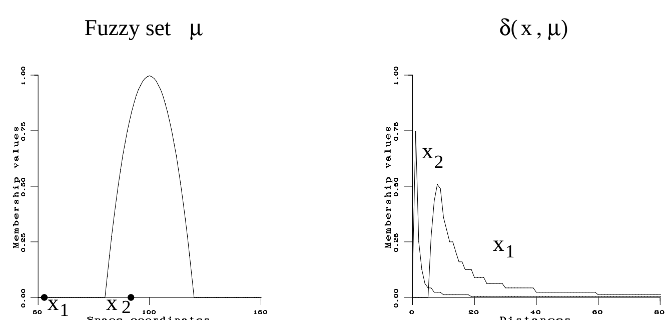


Figure 38: Fuzzy numbers representing $\delta_{(x,\mu)}$ (μ being shown on the left) for two different x.

From this definition of a point to a fuzzy set, distances between two fuzzy sets can be derived using supremum or infimum computation of fuzzy numbers using the extension principle [67]. The details are given in [16]. Since the nearest point distance for instance is simply a minimum over distances from a point to a fuzzy set, the fuzzy minimum taken over points in a fuzzy set leads directly to a fuzzy nearest distance between two fuzzy sets (as a fuzzy number). Similarly the Hausdorff distance can be directly derived from the distance from a point to a fuzzy set using the maximum of fuzzy numbers.

5.5.3 Distance between two fuzzy objects

Several definitions can be found in the literature for distances between fuzzy sets (which is the main addressed problem). They can be roughly divided in two classes: distances that take only membership functions into account and that compare them pointwise, and distances that additionally include spatial distances.

The presentation given below is directly inspired by the classification proposed in [193], but adapted to image processing and vision purposes, by underlining for each definition the type of image information on which it relies. A complete review can be found in [31].

Comparison of membership functions In this Section, we review the main distances proposed in the literature that aim at comparing membership functions. They have generally been proposed in a general fuzzy set framework, and not specifically in the context of image processing. They do not really include information about spatial distances. The classification chosen here is inspired from the one found in [193]. Similar classifications can be found in [53, 135, 99].

The **functional approach** is probably the most popular. It relies on a L_p norm between μ and ν, leading to the following generic definition [102, 69, 121]:

$$d_p(\mu,\nu) = [\int_{x\in\mathcal{S}} |\mu(x)-\nu(x)|^p]^{1/p}, \tag{295}$$

$$d_\infty(\mu,\nu) = \sup_{x\in\mathcal{S}} |\mu(x)-\nu(x)|. \tag{296}$$

In the discrete finite case, these definitions become:

$$d_p(\mu,\nu) = [\sum_{x\in\mathcal{S}} |\mu(x)-\nu(x)|^p]^{1/p}, \tag{297}$$

$$d_\infty(\mu,\nu) = \max_{x\in\mathcal{S}} |\mu(x)-\nu(x)|. \tag{298}$$

A noticeable property of d_p is that it takes a constant value if the supports of μ and ν are disjoint, irrespectively of how far the supports are from each other in $\mathcal{S}$.

Among the **information theoretic approaches**, based on their definition of fuzzy entropy $E(\mu)$, de Luca and Termini define a pseudo-metric as [122]:

$$d(\mu,\nu) = |E(\mu)-E(\nu)|. \tag{299}$$

This distance does not satisfy the separability condition. This can be overcome by considering the quotient space obtained through the equivalence relation $\mu \sim \nu \Leftrightarrow E(\mu) = E(\nu)$. However this is not suitable for image processing. Indeed, since the entropy of a crisp set is zero, two crisp structures in an image belong to the same equivalence class, even if they are completely different. One main drawback of this approach is that the distance is based on the comparison of two global measures performed on μ an ν separately: there is nothing linking points of μ to points of ν, which is of reduced interest for computing distances.

Entropy functions under similarity [182, 46] combine this approach with the membership comparison approach. It has been applied in decision problems (in particular for questionnaires) but to our knowledge not in image processing.

Based on a similar approach, a notion of fuzzy divergence (which can be interpreted as a distance) has been introduced in [12], by mimicking Kullback's approach [115]:

$$d(\mu,\nu) = \frac{1}{|\mathcal{S}|}\sum_{x\in\mathcal{S}}[D_x(\mu,\nu)+D_x(\nu,\mu)] \tag{300}$$

with:

$$D_x(\mu,\nu) = \mu(x)\log\frac{\mu(x)}{\nu(x)} + (1-\mu(x))\log\frac{1-\mu(x)}{1-\nu(x)},$$

and the convention $\frac{0}{0}=1$. This distance is positive, symmetrical, but does not satisfy the triangular inequality. Moreover, it is always equal to 0 for crisp sets.

In the **set theoretic approach**, distance between two fuzzy sets is seen as a set dissimilarity function, based on fuzzy union and intersection. Examples are given in [193]. The basic idea is that the distance should be larger if the two fuzzy sets weakly intersect. Most of the proposed measures are inspired from the work by Tversky [170] that proposes two parametric similarity measures between two sets A and B:

$$\theta f(A\cap B) - \alpha f(A-B) - \beta f(B-A), \tag{301}$$

and in a rational form:

$$\frac{f(A\cap B)}{f(A\cap B)+\alpha f(A\cap\bar{B})+\beta f(B\cap\bar{A})}, \tag{302}$$

where $f(X)$ is typically the cardinality of X, α, β and θ are parameters leading to different kinds of measures, and $\bar{B}$ denotes the complement of B.

Let us mention a few examples (they are given in the finite discrete case). A measure being derived from the second Tversky measure by setting $\alpha=\beta=1$ has been used by several authors [67, 135, 99, 53, 57, 176, 193]:

$$d(\mu,\nu) = 1 - \frac{\sum_{x\in\mathcal{S}}\min[\mu(x),\nu(x)]}{\sum_{x\in\mathcal{S}}\max[\mu(x),\nu(x)]}. \tag{303}$$

This distance is a semi-metric, and always takes the constant value 1 as soon as the two fuzzy sets have disjoint supports. It also corresponds to the Jaccard index [57]. With respect to the typology presented in [45], this distance is a comparison measure, and more precisely a dissimilarity measure. Moreover, $1-d$ is a resemblance measure. Applications in image processing can be found e.g. in [174], where it is used on fuzzy sets representing objects features (and not directly spatial image objects) for structural pattern recognition on polygonal 2D objects.

A slightly different formula has been proposed in [175], which however translates a similar idea:

$$d(\mu,\nu) = 1 - \frac{1}{|\mathcal{S}|}\sum_{x\in\mathcal{S}}\frac{\min[\mu(x),\nu(x)]}{\max[\mu(x),\nu(x)]} \tag{304}$$

with the convention $\frac{0}{0}=1$. It is a semi-metric. It takes the constant value 1 if the two fuzzy sets have disjoint supports, without any other condition on their relative position in the space.

Another measure takes into account only the intersection of the two fuzzy sets [99, 53, 193]:

$$d(\mu,\nu) = 1 - \max_{x\in\mathcal{S}}\min[\mu(x),\nu(x)]. \tag{305}$$

It is a semi-pseudometric if the fuzzy sets are normalized. Again it is a dissimilarity measure, and $1-d$ is a resemblance measure. It is always equal to 1 if the supports of μ and ν are disjoint.

If we set $(\mu\Box\nu)(x) = \max[\min(\mu(x),1-\nu(x)),\min(1-\mu(x),\nu(x))]$, two other distances can be derived, as [99, 193]:

$$d(\mu,\nu) = \sup_{x\in\mathcal{S}}(\mu\Box\nu)(x), \tag{306}$$

$$d(\mu,\nu) = \sum_{x\in\mathcal{S}}(\mu\Box\nu)(x). \tag{307}$$

These two distances are symmetrical measures. They are separable only for binary sets. Also we have $d(\mu,\mu)=0$ only for binary sets. They are dissimilarity measures. The first one is equal to 1 if μ and ν have disjoint supports and are normalized (if they are not normalized, then this constant value is equal to the maximum membership value of μ and ν). The second measure is always equal to $|\mu|+|\nu|$ is μ and ν have disjoint supports.

These measures actually rely on measures of inclusion of each fuzzy set in the other. Indeed, an inclusion index can be defined as [161, 38] as:

$$\mathcal{I}(\mu,\nu) = \inf_{x\in\mathcal{S}} T[\mu(x), 1-\nu(x)], \tag{308}$$

where T is a t-conorm. Since the distance should be small if the two sets have a small degree of equality (the equality between μ and ν can be expressed by "μ included in ν and ν included in μ", which leads to an easy transposition to fuzzy equality), a distance may be defined from an inclusion degree as:

$$d(\mu,\nu) = 1 - \min[\mathcal{I}(\mu,\nu), \mathcal{I}(\nu,\mu)]. \tag{309}$$

By taking $T = \max$, we recover the definition derived from $(\mu\Box\nu)$. This approach has been used in [6, 176]. Other choices of T may lead to different properties of d. For instance, if T is taken as the Lukasiewicz t-conorm (bounded sum), then $(\mu\Box\nu)(x) = |\mu(x)-\nu(x)|$. Therefore we have:

$$\sup_{x\in\mathcal{S}}(\mu\Box\nu)(x) = d_\infty(\mu,\nu), \tag{310}$$

and:

$$\sum_{x\in\mathcal{S}}(\mu\Box\nu)(x) = d_1(\mu,\nu). \tag{311}$$

In this case, both distances are metrics in the discrete finite case.

These measures have been applied in image processing for image databases applications in [99].

Other inclusion indexes can be defined, e.g. from Tversky measure by setting $\alpha = 1$ and $\beta = 0$, leading to $\frac{f(A\cap B)}{f(A)}$ [57].

The last definitions given by equations 305 and 306 are respectively equivalent to $1-\Pi(\mu;\nu)$ and $1-\max[N(\mu;\nu), N(\nu;\mu)]$ (where Π and N are possibility and necessity functions) used in fuzzy pattern matching [49], [76], which has a large application domain, including image processing and vision (see e.g. [100]). It is interesting to note that they are related to fuzzy mathematical morphology, since $\Pi(\mu;\nu)$ corresponds to the dilation of μ by ν at origin, while $N(\mu;\nu)$ corresponds to the erosion of μ by ν at origin. These definitions can be straightforwardly generalized to fuzzy union and intersection derived from t-norms and t-conorms, leading to a correspondence with other forms of fuzzy mathematical morphology [38].

The **pattern recognition approach** consists in first expressing each fuzzy set in a feature space (for instance cardinality, moments, skewness) and to compute the Euclidean distance between two feature vectors [193] or attribute vectors [163]. This approach may take advantage of some of the previous approaches, for instance by using entropy or similarity in the set of features. It has been applied for instance for database applications [163].

A similar approach, called signal detection theory, has been proposed in [99]. It is based on counting the number of similar and different features.

Combination of spatial and membership comparisons The second class of methods tries to include the spatial distance $d_\mathcal{S}$ in the distance between μ and ν. In contrary to the definitions given in above, in this second class the membership values at different points of $\mathcal{S}$ are linked using some formal computation, making the introduction of $d_\mathcal{S}$ possible. This leads to definitions that do not share the drawbacks of previous approaches, for instance when the supports of the two fuzzy sets are disjoint.

The **geometrical approach** consists in generalizing one of the distances between crisp sets. This has been done for instance for nearest point distance [69, 147], mean distance [147], Hausdorff distance [69], and could easily be extended to other distances (see e.g. [41] for a review of crisp set distances). These generalizations follow four main principles.

The first one consists in considering fuzzy sets in a n dimensional space as $n+1$ dimensional crisp sets and then in using classical distances [91]. However, this is often not satisfactory in image processing because the n dimensions of $\mathcal{S}$ and the membership dimension (values in [0,1]) have completely different interpretations, and treating them in a unique way is questionable.

The second principle is a fuzzification principle (see Section 4.8): let D be a distance between crisp sets, then its fuzzy equivalent is defined by:

$$d(\mu,\nu) = \int_0^1 D(\mu_\alpha, \nu_\alpha) d\alpha. \tag{312}$$

or by a discrete sum if the fuzzy membership functions are piecewise constant [65, 193] (μ_α denotes the α-cut of μ). In this way, $d(\mu,\nu)$ inherits the properties of the chosen crisp distance. Another way to consider the fuzzification principle consists in using a double integration (see Section 4.8). However using this double fuzzification, some properties of the underlying distance may be lost.

The third principle consists in weighting distances by membership values. For the mean distance this leads for instance to [147]:

$$d(\mu,\nu) = \frac{\sum_{x\in\mathcal{S}}\sum_{y\in\mathcal{S}} d_{\mathcal{S}}(x,y)\min[\mu(x),\nu(y)]}{\sum_{x\in\mathcal{S}}\sum_{y\in\mathcal{S}}\min[\mu(x),\nu(y)]}. \tag{313}$$

The last approach consists in defining a fuzzy distance as a fuzzy set on $\mathbb{R}^+$ instead of as a crisp number using the extension principle (see Section 4.8). For the nearest point distance this leads to [147]:

$$d(\mu,\nu)(r) = \sup_{x,y,d_{\mathcal{S}}(x,y)\leq r} \min[\mu(x),\nu(y)]. \tag{314}$$

The Hausdorff distance is probably the distance between sets, the fuzzy extension of which has been the most widely studied. One reason for this may be that it is a true metric in the crisp case, while other set distances like minimum or average distances have weaker properties. Another reason is that it has been used to determine a degree of similarity between two objects, or between an object and a model [97]. Extensions of this distance have been defined using fuzzification over the α-cuts and using the extension principle [141, 142, 193, 59, 51, 47]. Other authors use the Hausdorff distance between the endographs of the two membership functions [59]. Several generalizations of Hausdorff distance have also been proposed under the form of fuzzy numbers [69]. Extensions of the Hausdorff distance based on fuzzy mathematical morphology have also been developed [16] and are presented in the next Section.

Extensions of these definitions may be obtained by using other weighting functions, for instance by using t-norms instead of min.

These distances share most of the advantages and drawbacks of the underlying crisp distance [41]: computation cost can be high (it is already high for several crisp distances); moreover, interpretation and robustness strongly depend on the chosen distance (for instance, Hausdorff distance is noise sensitive, whereas mean distance is not).

A **morphological approach** has been proposed in [16, 25]. We just give the examples of nearest point distance and Hausdorff distance.

In the binary case, for $n > 0$, the nearest point distance can be expressed in morphological terms as:

$$d_N(X,Y) = n \Leftrightarrow D^n(X)\cap Y \neq \emptyset \text{ and } D^{n-1}(X)\cap Y = \emptyset \tag{315}$$

and the symmetrical expression. For $n = 0$ we have:

$$d_N(X,Y) = 0 \Leftrightarrow X \cap Y \neq \emptyset. \tag{316}$$

The translation of these equivalences provides, for $n > 0$, the following distance density:

$$\delta_N(\mu,\mu')(n) = t[\sup_{x\in\mathcal{S}} t[\mu'(x), D^n_\nu(\mu)(x)], c[\sup_{x\in\mathcal{S}} t[\mu'(x), D^{n-1}_\nu(\mu)(x)]]] \tag{317}$$

or a symmetrical expression derived from this one, and:

$$\delta_N(\mu,\mu')(0) = \sup_{x\in\mathcal{S}} t[\mu(x),\mu'(x)]. \tag{318}$$

This expression shows how the membership values to μ' are included, without involving the extension principle.

Like for the nearest point distance, we can extend the Hausdorff distance by translating directly the binary equation defining the Hausdorff distance:

$$d_H(X,Y) = \max[\sup_{x\in X} d_B(x,Y), \sup_{y\in Y} d_B(y,X)]. \tag{319}$$

This distance can be expressed in morphological terms as:

$$d_H(X,Y) = \inf\{n, X\subset D^n(Y) \text{ and } Y\subset D^n(X)\}. \tag{320}$$

From equation 320, a distance distribution can be defined, by introducing fuzzy dilation:

$$\Delta_H(\mu,\mu')(n) = t[\inf_{x\in\mathcal{S}} T[D^n_\nu(\mu)(x), c(\mu'(x))], \inf_{x\in\mathcal{S}} T[D^n_\nu(\mu')(x), c(\mu(x))]], \tag{321}$$

where c is a complementation, t a t-norm and T a t-conorm. A distance density can be derived implicitly from this distance distribution.

A direct definition of a distance density can be obtained from:

$$d_H(X,Y) = 0 \Leftrightarrow X = Y, \tag{322}$$

and for $n > 0$:

$$\begin{aligned} d_H(X,Y) = n \Leftrightarrow X\subset D^n(Y) &\text{ and } Y\subset D^n(X) \\ \text{and } \left(X \not\subset D^{n-1}(Y) \right.&\left.\text{ or } Y\not\subset D^{n-1}(X)\right). \end{aligned} \tag{323}$$

Translating these equations leads to a definition of the Hausdorff distance between two fuzzy sets μ and μ' as a fuzzy number:

$$\delta_H(\mu,\mu')(0) = t[\inf_{x\in\mathcal{S}} T[\mu(x), c(\mu'(x))], \inf_{x\in\mathcal{S}} T[\mu'(x), c(\mu(x))]], \tag{324}$$

$$\begin{aligned} \delta_H(\mu,\mu')(n) \quad = \quad & t[\inf_{x\in\mathcal{S}} T[D^n_\nu(\mu)(x), c(\mu'(x))], \inf_{x\in\mathcal{S}} T[D^n_\nu(\mu')(x), c(\mu(x))], \\ & T(\sup_{x\in\mathcal{S}} t[\mu(x), c(D^{n-1}_\nu(\mu')(x))], \sup_{x\in\mathcal{S}} t[\mu'(x), c(D^{n-1}_\nu(\mu)(x))])]. \end{aligned} \tag{325}$$

The above definitions of fuzzy nearest point and Hausdorff distances (defined as fuzzy numbers) between two fuzzy sets do not necessarily share the same properties as their crisp equivalent. This is due in particular to the fact that, depending on the choice of the involved t-norms and t-conorms, excluded-middle and non-contradiction laws may not be satisfied. All distances are positive, in the sense that the defined fuzzy numbers have always a support included in $\mathbb{R}^+$. By construction, all defined distances are symmetrical with respect to μ and μ'. The separability property is not always satisfied. However, if μ is normalized, we have for the nearest point distance $\delta_N(\mu,\mu)(0) = 1$ and $\delta_N(\mu,\mu)(n) = 0$ for $n > 1$. For the Hausdorff distance, $\delta_H(\mu,\mu')(0) = 1$ implies $\mu = \mu'$ for T being the bounded sum ($T(a,b) = \min(1, a+b)$), while it implies μ and μ' crisp and equal for $T = \max$. Also the triangular inequality is not satisfied in general.

Examples of distances between brain structures (see Figure 20) are shown in Table 9, as fuzzy numbers issued from the morphological definition of Hausdorff distance. The results are in agreement with what is expected: the model of v2 provided by an anatomical atlas is near from cn2 and v1, quite far from cn1

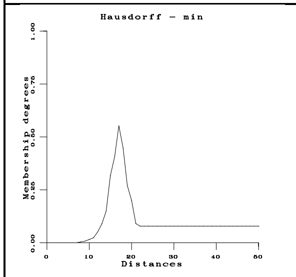
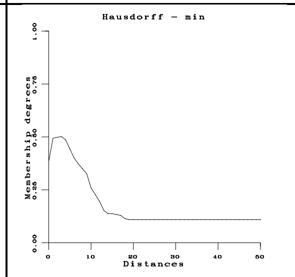
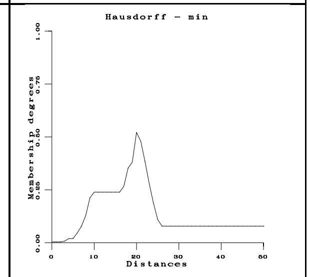
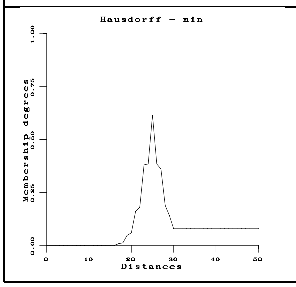
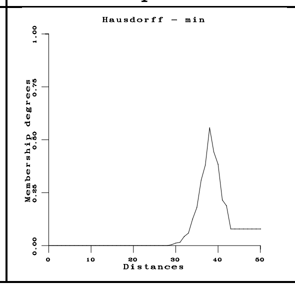
<table>
<tr><td colspan="3">Distance between a model of v2 and:</td></tr>
<tr><td>cn2</td><td>v2</td><td>v1</td></tr>
<tr><td>
</td><td>
</td><td>
</td></tr>
<tr><td>cn1</td><td>p1</td><td></td></tr>
<tr><td>
</td><td>
</td><td></td></tr>
</table>

Table 9: Distances between fuzzy sets using the morphological approach, for the Hausdorff distance, using minimum as t-norm. The distance is computed between each of the 5 structures and a model of v2 given by an anatomical atlas.

and very far from p1. We do not obtain a null value for v2, since it does not perfectly match the model of v2, but we obtain values that are still much lower than those obtained for the other structures. This shows that distances can be used both for identifying a structure among several possible ones, by using distance as a dissimilarity measure, and for describing the spatial arrangement of objects.

A **tolerance-based approach** has been developed in [121]. The basic idea is to combine spatial information and membership values by assuming a tolerance value τ, indicating the differences that can occur without saying that the objects are no more similar. The proposed definitions are semi-pseudometrics and are derived from the functional approach. The authors first define a local difference between μ and ν at a point x of $\mathcal{S}$ as:

$$d_x^\tau(\mu,\nu) = \inf_{y,z \in B(x,\tau)} |\mu(y) - \nu(z)|, \tag{326}$$

where $B(x,\tau)$ denotes the (spatial) closed ball centered at x of radius τ.

Then the functions d_p, d_∞ and d_{EssSup} are defined up to a tolerance τ as:

$$d_p^\tau(\mu,\nu) = [\int_{\mathcal{S}} [d_x^\tau(\mu,\nu)]^p dx]^{1/p}, \tag{327}$$

$$d_\infty^\tau(\mu,\nu) = \sup_{x \in \mathcal{S}} d_x^\tau(\mu,\nu), \tag{328}$$

$$d_{EssSup}^\tau(\mu,\nu) = \inf\{k \in \mathbb{R}, \lambda(\{x \in \mathcal{S}, d_x^\tau(\mu,\nu) > k\}) = 0\}. \tag{329}$$

Several results are proved in [121], in particular about convergence: $d_p^\tau(\mu,\nu)$ converges towards $d_{EssSup}^\tau(\mu,\nu)$ when p goes to infinity, all pseudo-metrics are decreasing with respect to τ, and converge towards d_p, d_∞ and d_{EssSup} when τ becomes infinitely small, for continuous fuzzy sets.

This approach has been extended in [120], by allowing the neighborhood around each point to depend on the point.

Note that this approach has strong links with morphological approaches, since the neighborhood considered around each point can be considered as a structuring element.

This approach has been illustrated on an example of noisy character recognition.

Finally, according to a **graph theoretic approach**, a similarity function between fuzzy graphs may also induce a distance between fuzzy sets. This approach contrasts with the previous ones, since the objects are no more represented directly as fuzzy sets on $\mathcal{S}$ or as vectors of attributes, but as higher level structures. Fuzzy graphs in cognitive vision can be used for representing objects, as in [124], or a scene, as in [107]. In the first case, nodes are parts of the objects and arcs are links between these parts. In the example presented in [124] for character recognition, nodes are fuzzy sets representing features of a character, extracted by some image processing. In the second case, nodes are objects of the scene and arcs are relationships between these objects. In the example of [107], the nodes represent clouds extracted from satellite images. These two examples use different ways to consider distances (or similarity) between fuzzy graphs.

Discussion In the first class of methods, the only way μ and ν are combined is by computation linking $\mu(x)$ and $\nu(x)$, i.e. only the memberships at the same point of $\mathcal{S}$. No spatial information is taken into account. A positive consequence is that the corresponding distances are easy to compute. The complexity is linear in the cardinality of $\mathcal{S}$. Considering image processing and vision applications, we suggest that the first class of methods (comparing membership functions only) be restricted when the two fuzzy sets to be compared represent the same structure or a structure and a model. Applications in model-based or case-based pattern recognition are foreseeable.

On the other hand, the definitions which combine spatial distance and fuzzy membership comparison allow for a more general analysis of structures in images, for applications where topological and spatial arrangement of the structures of interest is important (segmentation, classification, scene interpretation). This is permitted by the fact that these distances combine membership values at different points in the space, therefore taking into account their proximity or farness in $\mathcal{S}$. The price to pay is an increased complexity, generally quadratic in the cardinality of $\mathcal{S}$.

When facing the problem of choosing a distance, several criteria can be used. First, the type of application at hand plays an important role. While both classes of methods can be used for comparing an object and a model object, only the second class can be used for evaluating distances between objects in the same image. Among the distances of the first class, the results we obtained show that entropy and divergence based approaches are not satisfactory. Also normalized distances should be avoided in most cases. The choice among the remaining distances can be done by looking at the properties of the distances (for instance, do we need $d(\mu,\mu) = 0$ for the application at hand?), and at the computation time. Among the distances of the second class, similar choice criteria can be used.

5.5.4 Geodesic distance in a fuzzy set

Although the concept of geodesy is very important for crisp sets and should be promising as well for fuzzy sets, this topic has not been much addressed in the literature [17, 153].

Fuzzy geodesic distance defined as a number In [17] original definitions were proposed for the distance between two points in a fuzzy set, extending the notion of geodesic distance. Among these definitions, one proved to have desirable properties and was therefore considered as better than the others. We recall here this definition and the main results we obtained.

The geodesic distance between two points x and y represents the length of the shortest path between x and y that "goes out of μ as least as possible". A formal definition of this concept relies on the degree of connectivity, as defined by Rosenfeld [146]. In the case where $\mathcal{S}$ is a discrete bounded space (as is usually the case in image processing), the degree of connectivity in μ between any two points x and y of

$\mathcal{S}$ is defined as:

$$c_\mu(x,y) = \max_{L_i \in L}[\min_{t \in L_i} \mu(t)], \tag{330}$$

where L denotes the set of all paths from x to y. Each possible path L_i from x to y is constituted by a sequence of points of $\mathcal{S}$ according to the discrete connectivity defined on $\mathcal{S}$.

We denote by $L^*(x,y)$ a shortest path between x and y on which c_μ is reached (this path, not necessarily unique, can be interpreted as a geodesic path descending as least as possible in the membership degrees), and we denote by $l(L^*(x,y))$ its length (computed in the discrete case from the number of points belonging to the path). Then we define the geodesic distance in μ between x and y as:

$$d_\mu(x,y) = \frac{l(L^*(x,y))}{c_\mu(x,y)}. \tag{331}$$

If $c_\mu(x,y) = 0$, we have $d_\mu(x,y) = +\infty$, which corresponds to the result obtained with the classical geodesic distance in the case where x and y belong to different connected components (actually it corresponds to the generalized geodesic distance, where infinite values are allowed).

This definition corresponds to the weighted geodesic distance (in the classical sense) computed in the α-cut of μ at level $\alpha = c_\mu(x,y)$. In this α-cut, x and y belong to the same connected component (for the considered discrete crisp connectivity). This definition is illustrated in Figure 39.

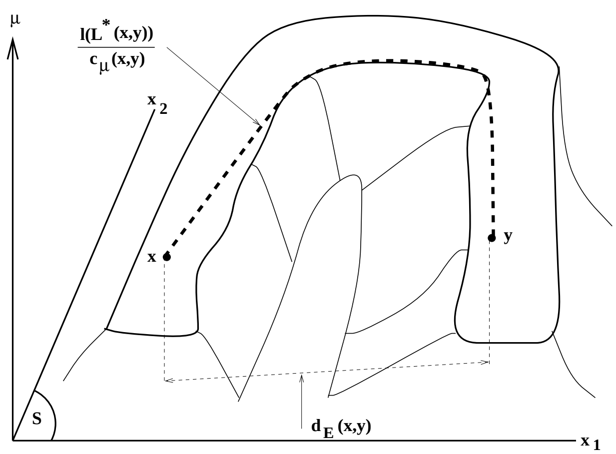


Figure 39: Illustration of the geodesic distance in a fuzzy set μ between two points x and y in a 2D space.

This definition satisfies the following set of properties (see [17] for the proof):

1. positivity: $\forall (x,y) \in \mathcal{S}^2, d_\mu(x,y) \geq 0$;
2. symmetry: $\forall (x,y) \in \mathcal{S}^2$, $d_\mu(x,y) = d_\mu(y,x)$;
3. separability: $\forall (x,y) \in \mathcal{S}^2$, $d_\mu(x,y) = 0 \Leftrightarrow x = y$;
4. d_μ depends on the shortest path between x and y that "goes out" of μ "as least as possible", and d_μ tends towards infinity if it is not possible to find a path between x and y without going through a point t such that $\mu(t) = 0$;
5. d_μ is decreasing with respect to $\mu(x)$ and $\mu(y)$;
6. d_μ is decreasing with respect to $c_\mu(x,y)$;
7. d_μ is equal to the classical geodesic distance if μ is crisp.

The triangular inequality is not satisfied, but from this definition, it is possible to build a true distance, satisfying triangular inequality, while keeping all other properties. This can be achieved in the following way (see [17] for proof and details):

$$d'_\mu(x,y) = \min_{t \in \mathcal{S}} \left[\frac{l(L^*(x,t))}{c_\mu(x,t)} + \frac{l(L^*(t,y))}{c_\mu(t,y)} \right].$$

Unfortunately this is computationally expensive.

These properties are in agreement with what can be required from a fuzzy geodesic distance, both mathematically and intuitively.

The definition proposed in [153] corresponds to one of the definitions proposed in [17] and is the length of the shortest path between the two considered points, the length being computed as the integral of the membership values along the path. Unfortunately, this definition does not meet all requirements we have here, since it does not satisfy the separability property and does not has the appropriate behavior with respect to the membership values (properties (4)-(6) above). Indeed the best path can go trough points with very low values (which tend to decrease the length), i.e. to go out of the set to some extent. However, one advantage of this distance is that it allows the authors in [153] to derive algorithms for computing the fuzzy distance transform.

Fuzzy geodesic distance defined as a fuzzy number In the previous approach, the geodesic distance between two points is defined as a crisp number (i.e. a standard number). It could be also defined as a fuzzy number, taking into account the fact that, if the set is imprecisely defined, geodesic distances in this set can be imprecise too. One solution to achieve this aim is to use the extension principle, based on a combination of the geodesic distances computed on each α-cut of μ. Let us denote by $d_{\mu_\alpha}(x,y)$ the geodesic distance between x and y in the crisp set μ_α. Using the extension principle, we define the degree to which the geodesic distance between x and y in μ is equal to d as:

$$\forall d \in \mathbb{R}^+,\ d_\mu(x,y)(d) = \sup\{\alpha \in [0,1],\ d_{\mu_\alpha}(x,y) = d\}. \quad (332)$$

This definition satisfies the following properties:

1. If $\alpha > c_\mu(x,y)$, then x and y belong to two distinct connected components of μ_α[11]. In this case, the (generalized) geodesic distance is infinite. If we restrict the evaluation of $d_\mu(x,y)(d)$ to finite distances d, then $d_\mu(x,y)(d) = 0$ for $d > d_{\mu_{c_\mu(x,y)}}$.

2. Let $d\ _{calS}(x,y)$ denote the Euclidean distance between x and y. It is the shortest of the geodesic distances that can be obtained in any crisp set that contains x and y. This set can be for instance the whole space $\mathcal{S}$, which can be assimilated to the α-cut of level 0 (μ_0). Therefore, for $d < d_\mathcal{S}(x,y)$, we have $d_\mu(x,y)(d) = 0$.

3. Since the α-cuts are nested ($\mu_\alpha \subset \mu_{\alpha'}$ for $\alpha > \alpha'$), it follows that $d_{\mu_\alpha}(x,y)$ is increasing in α, for $\alpha \leq c_\mu(x,y)$. Therefore, $d_\mu(x,y)$ is a fuzzy number, with a maximum value for $d_{\mu_{c_\mu(x,y)}}$, and with a discontinuity at this point.

This definition can be normalized by dividing all values by $c_\mu(x,y)$, in order to get a maximum membership value equal to 1.

One drawback of this definition is the discontinuity at $d_{\mu_{c_\mu(x,y)}}$. It corresponds to the discontinuity existing in the crisp case too when x and y belong to parts that become disconnected. Further work aims at exploiting features of fuzzy set theory in order to avoid this discontinuity, if this is found desirable.

[11]since $c_\mu(x,y)$ corresponds to "height" (in terms of membership values) of the point along the path that connects x and y, i.e. the maximum of the minimal height along paths from x to y.

5.6 Directional relative position between objects

Concepts related to directional relative position are rather ambiguous, they defy precise definitions. However, human beings have a rather intuitive and common way of understanding and interpreting them. From our every day experience, it is clear that any "all-or-nothing" definition leads to unsatisfactory results in several situations, even of moderate complexity such as those illustrated in Figure 40: on the left, the object A is to the right of R but it can also be considered to be to some extent above it; on the right, the object B is strongly to the right of R and above it. Fuzzy set theory appears then as an appropriate tool for such modeling. In the following, we denote by $\mathcal{S}$ the Euclidean space where the objects are defined. $\mathcal{S}$ is typically a 2D or 3D discrete space (as in image processing and vision).

5.6.1 Main fuzzy approaches

Fuzzy relations describing relative position In [129, 105], the angle between the segment joining two points a and b and the x-axis of the coordinate frame (in 2D) is computed. This angle, denoted by $\theta(a, b)$, takes values in $[-\pi, \pi]$, which constitutes the domain on which primitive directional relations are defined.

The four such relations "left","right", "above" and "below" are defined in [129] as $\cos^2\theta$ and $\sin^2\theta$ functions. Other functions are possible: in [105] trapezoidal shaped membership functions are used, for the same relations. Whatever the equations, the membership functions for these relations are denoted by μ_{left}, μ_{right}, μ_{above}, and μ_{below}, and are functions from $[-\pi, \pi]$ into [0,1]. The equations are chosen according to simplicity (e.g. cos or sin functions), to the fact that they define a fuzzy partition of $[-\pi, \pi]$, and to their invariance properties with respect to rotation (i.e. a rotation should correspond to a translation of the membership functions).

In the work relying on these definitions, only these four basic directions are used, other relations being expressed in terms of these. However, we can propose a straightforward extension to any direction. In 2D, a direction is defined by an angle α with the x-axis. Using this convention, the relationship "right" corresponds to $\alpha = 0$. From $\mu_{right} = \mu_0$, we derive μ_α, representing the relationship "in direction α", for any α as follows:

$$\forall\theta, \mu_\alpha(\theta) = \mu_0(\theta - \alpha) \tag{333}$$

with for instance:

$$\mu_0(\theta) = \begin{cases} \cos^2(\theta) & if\ \theta \in [-\frac{\pi}{2}, +\frac{\pi}{2}] \\ 0 & elsewhere \end{cases} \tag{334}$$

This makes the definitions based on angle computation more general. Moreover, it guarantees geometric invariance.

Another solution for defining relations intermediate between the four basic ones has been proposed in [130]. This solution is based on logical combinations of these four basic relations. For instance, "oblique right" is defined by "(above and right of) or (below and right of)". The membership function representing this relationship is computed from μ_{right}, μ_{above}, and μ_{below} using min (for "and") and max (for "or"). In a more general way, any t-norm and t-conorm could be used for that purpose. The advantage of this approach is that only four membership functions have to be defined, which is consistent with the usual way of speaking about relative position. The drawback is that, contrary to the definition proposed in Equation 333, we cannot achieve a great precision in direction using this approach. Also, the shape of the membership function will vary depending on the considered direction, leading to a high anisotropy and therefore a loss of rotation invariance, while it remains the same using Equation 333.

The extension to 3D images calls for a representation of a direction by two angles, denoted by α_1 and α_2 (with $\alpha_1 \in [0, 2\pi]$ and $\alpha_2 \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, the 2D case corresponding to $\alpha_2 = 0$).

In [130], the membership functions defining six basic directions in the 3D space are defined, again using squared cosinus and sinus functions. For instance, the relation "right" is defined as:

$$\mu_{right}(\alpha_1, \alpha_2) = \begin{cases} \cos^2(\alpha_1)\cos^2(\alpha_2) & if\ \alpha_1 \in [-\frac{\pi}{2}, +\frac{\pi}{2}]\ and\ \alpha_2 \in [-\frac{\pi}{2}, +\frac{\pi}{2}] \\ 0 & otherwise \end{cases} \tag{335}$$

The other relations are defined in a similar way.

Centroid method A first simple solution to evaluate a fuzzy relationship between two objects consists in representing each object by a characteristic point. This point is chosen as the object centroid in [114, 105]. Let c_R and c_A denote the centroids of objects R and A. The degree of satisfaction of the proposition "A is to the right of R" is then defined as:

$$\mu^R_{right}(A) = \mu_{right}(\theta(c_R, c_A)), \tag{336}$$

where the membership function μ_{right} is defined as in Section 5.6.1.

Note that, although not mentioned in the original paper, this method can be straightforward extended to 3D.

Extension to fuzzy objects can be done in two ways. One way consists in computing a weighted centroid, where the contribution of each object point is equal to its membership value. The second way consists in applying the definition for binary objects on each α-cut and then aggregating the results using a summation [65], or the extension principle [189]. However, this second method may be computationally expensive, depending on the quantization of the object membership values.

Histogram of angles: compatibility method The method proposed in [128, 129, 130] consists in computing the normalized histogram of angles and in defining a fuzzy set in [0,1] representing the compatibility between this histogram and the fuzzy relation. More precisely, the angle histogram is computed from the angle between any two points in both objects as defined before, and normalized by the maximum frequency. Let us denote $H^R(A)$ this normalized histogram, where R is the reference object and A the object the position of which with respect to R is evaluated. $H^R(A)$ represents the spatial directional relations of the object A with respect to the reference object R. Issues that arise here include expressing this in terms of the basic relations, and extracting a global representative evaluation of this spatial relation.

With respect to the first issue, operations of compatibility and matching of two fuzzy sets are considered. The compatibility set $\mu_{C(H,\mu_\alpha)}$ between $H^R(A)$ and μ_α is defined, for any $u \in [0,1]$, following the extension principle as:

$$\mu_{C(H,\mu_\alpha)}(u) = \begin{cases} 0 & if\ \mu_\alpha^{-1}(u) = \emptyset \\ \sup_{v|u=\mu_\alpha(v)} H^R(A)(v) & otherwise. \end{cases} \tag{337}$$

With respect to the second issue, a global evaluation of the relation can for instance be provided by the center of gravity of the compatibility fuzzy set:

$$\mu^R_\alpha(A) = \frac{\int_0^1 u\mu_{C(H,\mu_\alpha)}(u)du}{\int_0^1 \mu_{C(H,\mu_\alpha)}(u)du}. \tag{338}$$

Another solution for the first issue is to use a fuzzy pattern matching approach [49, 76] (between μ_α and $H^R(A)$), as suggested in [18]. Then the global evaluation is given in the form of a pair necessity/possibility.

Ultimately, this global evaluation, which can be done in many ways, has to be selected according to the type and goals of the application at hand.

The extension of this method to 3D objects amounts to computing a bi-dimensional histogram, i.e. as a function of two angles, and then applying the same principle using the relations defined in 3D. The computation of the histogram is heavy in 2D, and becomes even more so in 3D. Another problem when computing bi-histograms is that the domain of possible angle values may be under-represented, depending on the size and the sampling of the considered objects. This may result in a noisy and hole containing histogram. This effect already appears in 2D.

A possible direction for overcoming the computational burden would be to consider 2D restrictions of the 3D space, depending on the direction we are interested in. For instance, when assessing a left or right direction, it might be sufficient to look only at projections of the objects in the horizontal plane. This approach can be considered in particular for almost convex objects without holes. However, for more complex objects, too much information may be lost by this approach.

The fuzzy extension of this method is based on a weighted histogram [129]. Let us denote by μ_R and μ_A the membership functions of the fuzzy objects R and A. The weighted histogram is computed as:

$$H^R(A)(\theta) = \sum_{a,b,\theta(a,b)=\theta} \min[\mu_R(a), \mu_A(b)]. \tag{339}$$

This expression if equivalent to compute a histogram on each α-cut and to combine the obtained results by summation as in [65]. Indeed, let us consider, for instance, a discretization of the values of α, as $\alpha_1, ...\alpha_n$, with $\alpha_1 = 0$, $\alpha_n = 1$, and $\forall i, 1 \leq i \leq n-1, \alpha_i < \alpha_{i+1}$ (a similar reasoning holds in the continuous case). Let us denote by μ_{R_i} and μ_{A_i} the α-cuts of μ_R and μ_A at level α_i. Then we have:

$$\min[\mu_R(a), \mu_A(b)] = \sum_{i=1}^{n-1} (\alpha_{i+1} - \alpha_i) \min[\mu_{R_i}(a), \mu_{A_i}(b)]. \tag{340}$$

From this equality, we derive:

$$\begin{aligned} H^R(A)(\theta) &= \sum_{a,b,\theta(a,b)=\theta} \sum_{i=1}^{n-1} (\alpha_{i+1} - \alpha_i) \min[\mu_{R_i}(a), \mu_{A_i}(b)] \\ &= \sum_{i=1}^{n-1} (\alpha_{i+1} - \alpha_i) \sum_{a,b,\theta(a,b)=\theta} \min[\mu_{R_i}(a), \mu_{A_i}(b)] \\ &= \sum_{i=1}^{n-1} (\alpha_{i+1} - \alpha_i) H^{\mu_{R_i}}(\mu_{A_i})(\theta). \end{aligned} \tag{341}$$

This shows that the weighted histogram is equivalent to the summation of the histograms of the α-cuts. However the first form is computationally much less expensive, and does not involve any assumption on the quantification of the values of α.

Aggregation method An aggregation method has been proposed in [114, 105], which uses all points of both objects instead of only one characteristic point. For any pair of points i in R and j in A, the angle $\theta(i, j)$ is computed, and the corresponding membership value for a direction α (being one of the 4 considered relations) is computed as previously:

$$\mu_{ij} = \mu_\alpha(\theta(i, j)). \tag{342}$$

All these values are then aggregated. The aggregation operator suggested in [105] is a weighted mean:

$$\mu_\alpha^R(A) = \left[\sum_{i \in R} \sum_{j \in A} w_{ij} \mu_{ij}^p \right]^{1/p}, \tag{343}$$

where w_{ij} are weights the sum of which is equal to 1.

Learning approach In [106], a learning approach using neural networks is proposed in order to cope with the complexity and variety of spatial relationships. The idea is to learn membership functions of spatial relationships for a few types of shapes, for which it is possible to easily assign membership values. Four basic relations are considered. The angle histograms of the training data are the inputs of neural networks (one such network for each type of shape), the outputs of which are then combined using Choquet fuzzy integrals.

The main problem with this approach is the first assignment for the training data. It does not seem easy to define criteria that allow to distinguish between values such as 0.002 and 0.005, 0.97 and 0.93, etc.

Histogram of forces Instead of considering pairs of points as in angle histogram approaches, pairs of longitudinal sections are considered in [125], where the concept of F-histogram is introduced. The degree to which an object A is in the direction α with respect to a reference object R is computed using successively points, segments, and longitudinal sections. Information on points is translated by a function ϕ acting on the difference of coordinates on the α-axis between points of A and points of R. Therefore, distance information is explicitly taken into account. A second function f integrates ϕ on segments of A and R in the direction α. Finally, the contributions of segments constituting the longitudinal sections of A and R in the direction α are summed. A so called "histogram of forces" allows to compute the weight supporting a proposition like "object A is in direction α from object R".

The definitions of the functions involved in this construction are done in an axiomatic way, that guarantees that the obtained relationships have good properties. Note that a similar axiomatic treatment can be adopted for the other approaches without any change in the results. Basically, this approach amounts to considering a weighted angle histogram:

$$H^R(A)(\theta) = \sum_{a,b,\theta(a,b)=\theta} \varphi(||\vec{ab}||), \tag{344}$$

where φ is a decreasing function. Typically, $\varphi(x) = \frac{1}{x^r}$. For $r = 0$, the weighted histogram is equal to the angle histogram, and for $r \geq 1$, it gives more importance to points of A that are close to some points of R. This allows to deal with situations where A and R have very different partial extents, and to account only for the closest parts of them.

This approach has been extended to fuzzy objects using their α-cuts. Extension to 3D objects could be probably done, but with a high complexity.

Projection based approach The approach proposed in [109] is very different from the previous ones since it does not use any histogram. It is based on a projection of the considered object on the axis related to the direction to be assessed (e.g. the x-axis for evaluating the relations "left to" and "right to"). Let us detail the computation for the relation "A is left from R". The same construction applies for any direction. Let us denote by $R^f(x)$ the normalized projection of the set R on the x-axis. The degree for a point x to be left to R is defined as:

$$R^{\leftarrow}(x) = \frac{\int_x^{+\infty} R^f(y)dy}{\int_{-\infty}^{+\infty} R^f(y)dy}. \tag{345}$$

This definition provides a degree of 1 for points that are completely on the left of the support of R^f and a degree of 0 for points that are completely on the right of the support of R^f, and the degree decreases in-between.

Let us now introduce a second set A. The degree $(A \leftarrow R)^f(x)$ to which x is in the projection of A and to the left of R is expressed as a conjunction of $A^f(x)$ and $R^{\leftarrow}(x)$. The conjunction is taken as a product in [109]. The degree to which A is left from R is then deduced as the ratio of the areas below

$(A \leftarrow R)^f$ and A^f:

$$\mu_\alpha^R(A) = \frac{\int_{-\infty}^{+\infty} A^f(x) \int_x^{+\infty} R^f(y) dy dx}{\int_{-\infty}^{+\infty} A^f(y) dy \int_{-\infty}^{+\infty} R^f(y) dy}. \tag{346}$$

This approach can be generalized to fuzzy sets [109] by taking each point into account in the projection to the amount of its membership function, leading to similar properties than in the crisp case.

Morphological approach In [18, 23, 24] a morphological approach has been proposed in order to evaluate the degree to which an object A is in some direction with respect to a reference object R, consisting of two steps:

1. A *fuzzy landscape* is first defined around the reference object R as a fuzzy set such that the membership value of each point corresponds to the degree of satisfaction of the spatial relation under examination. This makes use here of a spatial representation of fuzzy sets, which already proved to be useful in image processing and vision [112, 19]. Therefore the fuzzy landscape is directly defined in the same space as the considered objects, contrary to the projection method [109], where the fuzzy area is defined on a one-dimensional axis.
2. Then the object A is compared to the fuzzy landscape attached to R, in order to evaluate how well the object matches with the areas having high membership values (i.e. areas that are in the desired direction). This is done using a fuzzy pattern matching approach, which provides an evaluation as an interval or a pair of numbers instead of one number only.

A 3D direction is defined by two angles α_1 and α_2, where $\alpha_1 \in [0, 2\pi]$ and $\alpha_2 \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. The direction in which the relative position of an object with respect to another one is evaluated is denoted by: $\vec{u}_{\alpha_1,\alpha_2} = (\cos\alpha_2 \cos\alpha_1, \cos\alpha_2 \sin\alpha_1, \sin\alpha_2)^t$, and we note $\alpha = (\alpha_1, \alpha_2)$. Let us denote by $\mu_\alpha(R)$ the fuzzy set defined in the image such that points of areas which satisfy to a high degree the relation "to be in the direction $\vec{u}_{\alpha_1,\alpha_2}$ with respect to reference object R" have high membership values. In other words, the membership function $\mu_\alpha(R)$ has to be an increasing function of the degree of satisfaction of the relation. It is a spatial fuzzy set (i.e. a function of the image $\mathcal{S}$ into [0,1]) and directly related to the shape of R. The precise definition of $\mu_\alpha(R)$ is given below.

Let us denote by μ_A the membership function of the object A, which is a function of $\mathcal{S}$ into [0,1]. The evaluation of relative position of A with respect to R is given by a function of $\mu_\alpha(R)(x)$ and $\mu_A(x)$ for all $x \in \mathcal{S}$. An appropriate tool for defining this function is the fuzzy pattern matching approach [76]. Following this approach, the evaluation of the matching between two possibility distributions consists of two numbers, a necessity degree N (a pessimistic evaluation) and a possibility degree Π (an optimistic evaluation), as often used in conjunction with fuzzy sets. In our application, they take the following forms:

$$\Pi_{\alpha_1,\alpha_2}^R(A) = \sup_{x \in \mathcal{S}} t[\mu_\alpha(R)(x), \mu_A(x)], \quad N_{\alpha_1,\alpha_2}^R(A) = \inf_{x \in \mathcal{S}} T[\mu_\alpha(R)(x), 1 - \mu_A(x)], \tag{347}$$

where t is a t-norm (fuzzy intersection) and T a t-conorm (fuzzy union) [70]. In the crisp case, these equations reduce to: $\Pi_{\alpha_1,\alpha_2}^R(A) = \sup_{x \in A} \mu_\alpha(R)(x)$, and $N_{\alpha_1,\alpha_2}^R(A) = \inf_{x \in A} \mu_\alpha(R)(x)$.

The possibility corresponds to a degree of intersection between the fuzzy sets A and $\mu_\alpha(R)$, while the necessity corresponds to a degree of inclusion of A in $\mu_\alpha(R)$. They can also be interpreted in terms of fuzzy mathematical morphology, since the possibility $\Pi_{\alpha_1,\alpha_2}^R(A)$ is equal to the dilation of μ_A by $\mu_\alpha(R)$ at origin, while the necessity $N_{\alpha_1,\alpha_2}^R(A)$ is equal to the erosion, as shown in [38]. These two interpretations, in terms of set theoretic operations and in terms of morphological ones, explain how the shape of the objects is taken into account.

Several other functions combining $\mu_\alpha(R)$ and $\mu_A(x)$ can be constructed. The extreme values provided by the fuzzy pattern matching are interesting because of their morphological interpretation, and because they provide a pair of extreme values and not only a single value and may better capture the ambiguity of the relation if any. One drawback of these measures is that they are sensitive to noise, since they rely on infimum and supremum computation. An average measure can also be useful from a practical point of view (it is much less sensitive to noise), and is defined as:

$$M^R_{\alpha_1,\alpha_2}(A) = \frac{1}{|A|} \sum_{x \in \mathcal{S}} t[\mu_A(x), \mu_\alpha(R)(x)], \qquad (348)$$

where $|A|$ denotes the fuzzy cardinality of A: $|A| = \sum_{x \in \mathcal{S}} \mu_A(x)$.

The key point in the previous definition relies in the definition of $\mu_\alpha(R)$. The requirements stated above for this fuzzy set are not strong and leave room for a large spectrum of possibilities. This flexibility allows the user to define any membership function according to the application at hand and the context requirements. The following definition looks precisely at the domains of space that are visible from a reference object point in the direction $\vec{u}_{\alpha_1,\alpha_2}$. This applies to any kind of objects, including those having strong concavities.

Let us denote by P any point of $\mathcal{S}$, and by Q any point of R. Let $\beta(P,Q)$ be the angle between the vector $\vec{QP}$ and the direction $\vec{u}_{\alpha_1,\alpha_2}$, computed in $[0,\pi]$:

$$\beta(P,Q) = \arccos\left[\frac{\vec{QP} \cdot \vec{u}_{\alpha_1,\alpha_2}}{\|\vec{QP}\|}\right], \quad \textit{and} \;\; \beta(P,P) = 0. \qquad (349)$$

We then determine for each point P the point Q of R leading to the smallest angle β, denoted by $\beta_{\min}$. In the crisp case, Q is the reference object point from which P is visible in the direction closest to $\vec{u}_{\alpha_1,\alpha_2}$: $\beta_{\min}(P) = \min_{Q \in R} \beta(P,Q)$. The fuzzy landscape $\mu_\alpha(R)$ at point P is then defined as: $\mu_\alpha(R)(P) = f(\beta_{\min}(P))$, where f is a decreasing function of $[0,\pi]$ into $[0,1]$. We can chose for instance a simple linear function: $\mu_\alpha(R)(P) = \max(0, 1 - \frac{2\beta_{\min}(P)}{\pi})$.

In the fuzzy case, this definition is extended as: $Q \in R$ and $f(\beta_{\min}) = \max_{Q \in R} f(\beta(P,Q))$ (since f is decreasing), which translates in fuzzy terms as:

$$\mu_\alpha(R)(P) = \max_{Q \in Supp(S)} t[\mu_R(Q), f(\beta(P,Q))], \qquad (350)$$

where t is a t-norm.

An advantage of this approach is its interpretation in terms of morphological operations. It can be shown that $\mu_\alpha(R)$ is exactly the fuzzy dilation of μ_R by ν, where ν is a fuzzy structuring element defined on $\mathcal{S}$ as:

$$\forall P \in \mathcal{S},\; \nu(P) = \max[0, 1 - \frac{2}{\pi} \arccos \frac{\vec{OP} \cdot \vec{u}_\alpha}{\|\vec{OP}\|}], \qquad (351)$$

where O is the center of the structuring element. The following definition is used for the fuzzy dilation (see [38] for more details about fuzzy morphological operations):

$$\forall P \in \mathcal{S},\; D_\nu(\mu)(P) = \max_{Q \in \mathcal{S}} t[\mu(Q), \nu(P-Q)], \qquad (352)$$

where t is a t-norm. This equivalence provides an additional morphological interpretation of this approach.

5.6.2 An example

A formal comparison of these approaches is given in [43], based on their properties, the type of basic elements on which they rely, their behavior in extreme situations, in case of concavities, of distant objects, and on their computational cost.

Another aspect that is very important is the type of questions each definition is able to answer, or dedicated to. These questions can take different forms, e.g.:

- what are the spatial relationships between two given objects?

- to which degree a given spatial relation holds between two objects?
- what are the regions of the space where a spatial relationship is satisfied (to some degree) with respect to a reference object?

An important feature of angle histogram and force histogram is that they provide a general description of the directional relationships. From this general information, several ones can be deduced, as the degree of satisfaction of one specific relationship (for a particular direction), or the dominant relationship. This is not easy to obtain with the morphological approach, that needs one computation for each direction of interest. This approach is more dedicated to cases where we are interested in specified relations.

For problems where we have to assess the relative position of several objects with one reference object, the morphological approach may be more appropriate if the computation time is a strong requirement.

Most works in spatial reasoning aim at defining and assessing spatial relationships between objects, given these objects. But we may take another point of view, and address the problem of the representation of knowledge about expected relationships, in order to guide the reasoning process in the space, for exploring the space and search for the object that satisfies some relationships with respect to already known objects [29, 85]. For the example of model-based pattern recognition, this leads to progressive recognition, where each object is detected and recognized by gathering constraints given by the model expressing relationships that this object has to satisfy with respect to previously recognized objects [87]. For this problem where the relationships are considered as constraints with respect to one object (rather than a relation between two objects), we can make use of a spatial representation, as fuzzy sets in the space. Each relationship is expressed as one fuzzy set, corresponding to a spatial constraint, restricting the space to the only regions where the relationship is satisfied [29]. This can be directly obtained using the morphological approach (it is the result of the first step), but is more difficult to obtain with other approaches, that do not work directly in the image space.

We conclude this section by an example illustrating the definitions. Two pairs of objects are shown in Figure 40.

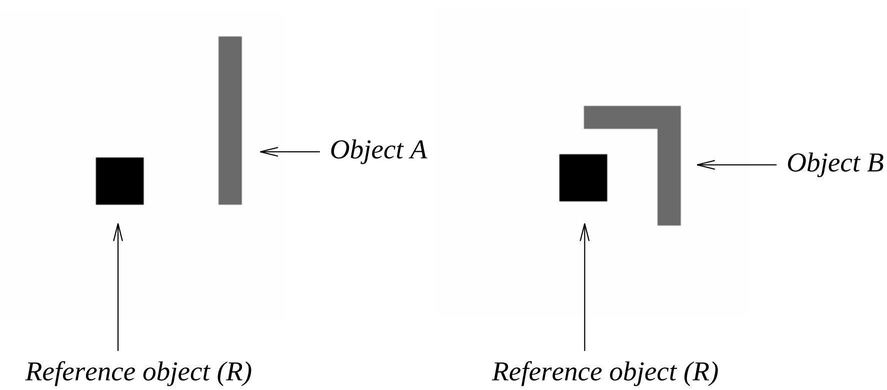


Figure 40: Two examples where the relative position of objects with respect to the reference object is difficult to define in a "all-or-nothing" manner.

Figure 41 presents the angle histograms for the two examples of Figure 40.

Illustrations of the definition of $\mu_\alpha(R)$ are given in Figure 42 for several reference objects. They show the consistency of the proposed approach: since the aim of the proposed definition is not to find only the dominant relationship, an object may satisfy several different relationships with high degrees. Therefore, "to be to the right of R" does not mean that the object should be completely to the right of the reference object, but only that at least part of the object is to the right of part of the region.

Tables 10 and 11 show the results for object A with respect to object R, according to various methods. They all agree to say that A is mainly to the right of R. The degree of being to the right increases with the value of r, since the part of A which is to the right of R is the closest one to R. On the contrary the

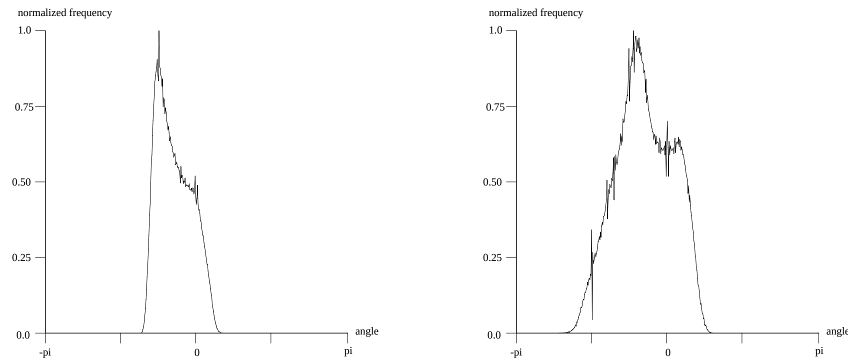


Figure 41: Angle histograms for the two examples of Figure 40. Left: object A w.r.t. reference object R; right: object B w.r.t reference object R.

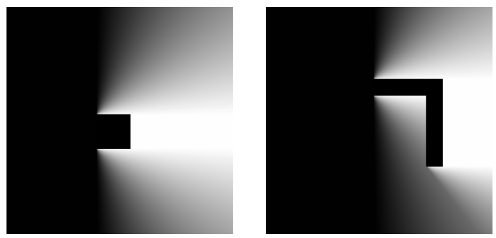


Figure 42: A few examples of $\mu_\alpha(R)$ for $\alpha_1 = \alpha_2 = 0$ corresponding to the relative position "right" (high grey values correspond to high membership values) using the morphological (angle of visibility) method, for different types of reference objects (reference objects are black).

degree of being above decreases with r. The values are somewhat different for all approaches, but since the ranking and the general behavior is the same, no conclusion concerning a more favorable approach can be derived from this example.

Object A with respect to object R							
Relation	Centroid	Aggregation			Compatibility		
		$r=0$	$r=2$	$r=5$	$r=0$	$r=2$	$r=5$
Left	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Right	0.76	0.73	0.79	0.86	0.62	0.67	0.75
Below	0.00	0.00	0.01	0.01	0.05	0.06	0.06
Above	0.24	0.27	0.20	0.13	0.38	0.33	0.25

Table 10: Relative position of object A (rectangle) with respect to object R (square) of Figure 40, using centroid, aggregation and compatibility methods. Angle or force histograms are computed using $r = 0$, $r = 2$ and $r = 5$ (the angle histogram method corresponds to $r = 0$).

5.7 Fuzzy classification

5.7.1 Pattern recognition

In the scientific literature, the words *pattern recognition* cover a broad set of concepts which lead from the observation of the universe to some decisions associated with the interpretation of some elements of this universe: under this light, recognition may cover the pre-processing stages, the feature detection, the

Object A with respect to object R				
Relation	FPM			Morphological approach
	$r=0$	$r=2$	$r=5$	
Left	[0.00, 0.00]	[0.00, 0.00]	[0.00, 0.00]	[0.00, 0.00] $M=0.00$
Right	[0.37, 0.68]	[0.39, 0.98]	[0.58, 1.00]	[0.50, 1.00] $M=0.81$
Below	[0.00, 0.10]	[0.00, 0.12]	[0.00, 0.13]	[0.00, 0.35] $M=0.05$
Above	[0.32, 0.63]	[0.02, 0.61]	[0.00, 0.42]	[0.00, 0.73] $M=0.44$

Table 11: Relative position of object A (rectangle) with respect to object R (square) of Figure 40, using fuzzy pattern matching approach (FPM) between relationships and angle or force histograms computed using $r=0$, $r=2$ and $r=5$, and using the morphological approach (the $[N, \Pi]$ intervals are given, as well as the average value).

classification and the derived diagnosis and appropriate actions. A popular example is to be found in the postal character recognition which starts from the handling of the enveloppe and stops when the letter is directed towards the appropriate railway station. Within this rendering, almost every picture processing is part of pattern recognition.

In the following lines, we will reduce the meaning of pattern recognition to the central part of the previous series of steps which consists of the design of appropriate algorithms to recognize prototypes or representative members of separate classes. Moreover, in this Section, we only concentrate to the recognition based on numerical arguments and not on the symbolic pieces of knowledge we have on the objects to be recognized. This part of pattern recognition is often identified as **statistical pattern recognition**, by opposition to **structural pattern recognition** which makes use of rules, graphs and grammars. Statistical pattern recognition referential text books are [84, 168, 61], while in structural pattern recognition, we may refer to [83]; [78] presents both aspects of pattern recognition in the framework of vision and image processing.

The general framework of statistical pattern recognition may be stated in simple words.

Starting from a set X of unknown objects $\{x_i, i = 1, ..., N\}$ called *variables* or *samples* each of them described by a vector of measurements $\mathbf{p}^i = [p_1^i, p_2^i, p_d^i]^t$ of dimension d, we want to affect each of them to one subset ω_j of X (called **a class**) with the constraint that the number of subsets ω_j be much smaller than N.

For some problems, the number of classes k may be known, as for instance it is the case for the recognition of zip-codes in postal codes.

For some other problems some kind of information may be available about the final classes in which the samples can be ordered, either because of some generic representation of each class (character recognition is again a good instance of this), or because prototypes of classified samples exist (for example when training sets are available or when an expert is supervising the classification).

But, from a general point of view, a classification problem, stated in the previous words, is solved by providing answers to the following mandatory steps:

1. choose an adequate space of representation to display the samples;
2. determine the number k of classes adapted to the objectives and to the data;
3. define the criteria on the basis of which to compare the samples to each individual class, and which governs the affectation of each sample to a class,
4. determine the criteria on which to appreciate the global quality of a classification,
5. determine the algorithm to achieve the best possible fullfilment of these criteria.

In the previous lines, we adopted the language of "conventional" pattern recognition. In brief, under this light, every sample belongs to one class, and may not belong to more. The decision is crisp and we will refer from now on to **crisp pattern recognition** when considering a problem from this point of view. Crisp pattern recognition is well adapted to treat some problems where a clear disjunction exists between classes. For instance dealing with the classification of pets, we have to attribute Hobbes[12] to the family of *cats*, *dogs* or *tigers*, and, unless unsufficiently informed, we cannot expect too much hesitation from any classifier, even if the *tigers* class is not likely to be well represented among the pet group. Crisp pattern recognition is a consequence of the *crisp* concept of membership of the conventional set theory.

Within a **fuzzy pattern recognition** point of view, the membership concept is changed (as we always did in this book up to now), into a degree of membership, and the injective association of samples to classes is no longer valid. Staying in the domestic zoology domain, we may take an illustration of the relevance of this idea by attempting to classify dogs. If we may certainly find some rare (and therefore expensive) examples of "crisp" terriers or mastiffs, we expect many crossbreedings which may belong, for instance, for one quarter to the terrier class and for three quarters to the mastiff class (if not even more oddly to 20 different families!).

Many solutions have been developed within the framework of crisp pattern recognition to deal with similar situations. One way is to introduce additional classes (which could be called *breeds*), another is to add a rejection class where samples which do not fulfill exactly some criteria are placed for further examination. But it is clear that the basic requirement of uniqueness at the root of crisp pattern recognition makes every tallied sample a stranger in the partition.

Therefore the question whether crisp or fuzzy pattern recognition should be employed is brought back to the problem at hand. Are you able to exhibit a clear cut partition from the world you observe? Can you guarantee that no interbreeding exists between classes? Is there a fundamental distinction between samples (even if it is hidden to the observer)? In these cases you should adopt a crisp point of view from which you may expect a more complete processing, since fuzzy pattern recognition must be followed by a stage of decision to treat the set of membership functions before any action.

If, on the contrary, you suppose your samples to be continuously distributed between some "pure" prototypes, if you do not expect some decisive fronteers to exist, or some precisive criterion to apply, then fuzzy pattern recognition will probably better fit your demand.

Crisp pattern recognition is ideally suited to treat problems where fundamentally there is no uncertainty in the objects to be recognized but where some imprecision comes from the observation system which doesn't allow a perfect access to the decisive parameters. If we consider the imprecision of the observation as a *noise* coming in between the object and us, then the dispersion of the samples due to noise makes the classes overlap. Pattern recognition is therefore the art of retrieving the classes from the distribution and, if possible, to attribute each sample to its original class. Under the light of imprecision, the words "statistical pattern recognition" take their full meaning. Noise, as a fundamental source of imprecision, spreads about well differentiated classes which could be later discovered from statistical studies on the distribution.

On the contrary, fuzzy pattern recognition is well suited to process under uncertainty. If pure classes may exist, they are not likely to cover every possible sample which may be a cross or a blend of several prototypes. The objective of fuzzy pattern recognition is more to qualify the relation of each sample to every prototype than to exhibit the "closest" one. One way to do it is to transform the uncertainty of the observed world on an imprecision of the classes. Under fuzzy pattern recognition, uncertainty and imprecision are mixed in a complex way which makes often difficult the task of identifying the role of each of them in the scattering of samples.

One of the main difficulties of the practitioner is to recognize from one experiment which point of view, crisp or fuzzy, is better representing the situation at hand. And the decision to choose one or the other representation is relevant from the **modelization** stage of the problem. We will come back later on the importance of this stage in the pattern recognition task. It may happen in some practical cases that the observed universe dictates the solution (our previous examples on pets are such schematic situations

[12] Thank you Mr. Watterson for this example.

where we do not hesitate), but most of the time our knowledge of the problem is not sufficient to impose one solution (and that is why pattern recognition is asked for).

Clearly the crisp point of view is simpler and more comfortable. It observes differences and decides on similarities. It rarely examines relations between classes and is not expected to quantify proximities. On the other hand, it incorporates the decision stage, this means that it makes one more step towards a future action than fuzzy pattern recognition which cautiously provides membership functions only.

5.7.2 Fuzzy C-means

The most familiar technique of fuzzy data classification is called Fuzzy C-means, it is issued from a family of crisp classification techniques based on the iterative modification of the clusters, the best representative of which is the k-means method. We briefly recall the main lines of the hard classification based on k-means.

k-means hard classification In the case of a hard classification, we want to obtain a k-partition P of the set X of samples, i.e. we are looking for k non-empty subsets of X, denoted $\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_k$, which verify:

1. $\forall i \neq j \qquad \mathcal{C}_i \bigcap \mathcal{C}_j = \emptyset$,
2. $\qquad \bigcup_i \mathcal{C}_i = X$.

We define the partition function of X in P as the binary function μ_{ij} equal to 1 if the sample x_i of X belongs to the class $\mathcal{C}_j$, and to 0 if not. N denotes the cardinality of X.

Let J be the objective functional of the classification problem. In the case of k-means, the functional is chosen as the sum of the intra-class variances, i.e., if m_j denotes the centroid of the class $\mathcal{C}_j$:

$$J = \sum_{i=1}^{N} \sum_{j=1}^{k} \mu_{ij} ||x_i - m_j||^2. \tag{353}$$

The objective of the classification is to minimize J by a proper choice of the partition P.

The k-means algorithm proceeds by the following steps:

- initiate the process by choosing the initial centroids (called seeds),
- iterate the following sequence:
  1. affect each sample to the class with the closest centroid (i.e. minimize J with respect to μ_i),
  2. compute a new set of centroids, according to the partition functions (i.e. minimize J with respect to m_j).
- stop the iteration when there is no more change in the partition or when the decay of the objective functional is small enough.

This technique has been widely described [80, 78] and adapted for many different problems (for instance when the samples are not all available at the same time, or when the number of classes is not known). It may be plunged into a more general family of iterative clustering techniques called "dynamic clusters" [64] where the distances are not taken to the centroids but to various possible manifolds of the underlying $\mathcal{S}$ space.

It is known that k-means are not converging towards the global minimum of functional J, but towards a local minimum which depends on the initial positions of the seeds. Therefore several different techniques have been proposed to initiate the seeds: if the random selection among the set X is the most popular initialization, other solutions start from a regular distribution of seeds in $\mathcal{S}$, from a choice of extreme points in $\mathcal{S}$, or benefit from some *a priori* knowledge on the final classification to choose an adapted initial configuration of seeds.

A great attention has been paid in the literature to the number k of clusters to be detected. In some rare cases, this number is fixed by the problem at hand (for instance in the case of character recognition we have prior information on k). Most of the time however, k has to be deduced from the data. The ISODATA method [79] has popularized a systematic approach where, in-between 2 limits (usually 2 and a known upper-bound of the class cardinality), a systematic classification is made with an increasing value of k and a criterion of classification quality is measured. The best number of clusters is the one which optimizes the criterion. Several different criteria may be proposed mostly heuristically deduced. We will discuss this important issue later.

The fuzzy case: method Within the fuzzy framework, a fuzzy k-partition[13] is defined by:

1. $\forall x_i \in X$ and $\forall \mathcal{C}_j \in P$ $\quad \mu_{ij} \in [0,1]$,
2. $\forall x_i \in X$ $\quad \sum_{j=1}^{k} \mu_{ij} = 1$,
3. $\forall j \leq k$ $\quad 0 < \sum_{i=1}^{N} \mu_{ij} < N$.

The fuzzy C-means algorithm [9] is based on a similar reasoning that k-means algorithm, i.e. on iteratively modifying the partition to minimize an objective functional. The objective functional is defined as:

$$J_m = \sum_{j=1}^{k} \sum_{i}^{N} \mu_{ij}^m ||x_i - m_j||^2, \tag{354}$$

where m is a parameter belonging to $]1, +\infty[$ called fuzzy factor.

The membership function is deduced from the cluster center position by the equation:

$$\mu_{ij} = \frac{1}{\sum_{j=1}^{k} [\frac{||x_i - m_i||}{||x_i - m_j||)}]^{\frac{2}{m-1}}}, \tag{355}$$

and the cluster center position is obtained by:

$$m_j = \frac{\sum_i \mu_{ij}^m x_i}{\sum_i \mu_{ij}^m}. \tag{356}$$

The scatter matrix for class $\mathcal{C}_j$ is also defined as:

$$M_j = \sum_{i=1}^{N} \mu_{ij} (x_i - m_j)(x_j - m_j)^t, \tag{357}$$

and the within-cluster scatter matrix as:

$$M = \sum_{j=1}^{k} M_j. \tag{358}$$

We verify easily that the objective functional J_m is the trace of the within-cluster scatter matrix M

The fuzzy C-means algorithm may now be written as:

- initiate the process by choosing the initial centers of the clusters,

[13] to respect the eponym notation we should have denoted by C the number of clusters, but in order to remain coherent with the notations of the rest of this monography we adopt the letter k as in the previous Section.

- iterate the following sequence:
  1. compute the membership function of any sample x_i with respect to any cluster $\mathcal{C}_j$ using Equation 355),
  2. compute a new set of cluster centers from Equation 356 .
- stop the iteration when there is no more change in the partition or when the decay of the objective fuctional is small enough.

One of the main advantages of fuzzy approach of classification is the possibility to differentiate in μ_{ij} the partition space covering X, this facility is not possible with the binary values μ_{ij} of 353. This benefit is exploited for demonstration of convergence for instance [9].

Discussion on the fuzzy case Let us discuss the different stages of this method.

At first, the functional 354 may be seen as an extension of Equation 353: it makes use of a non-binary weight μ_{ij}^m belonging to $[0,1]$ instead of the binary one of Equation 353. For $m=1$, each internal summation is equal to zero when m_j is the centroid of the set X taking the only membership values μ_{ij} into account, whatever the distribution of weights. For $m=2$, each internal summation provides the variance of the complete set X when the only memberships μ_{ij} are considered. For increasing values of m, the weights are becoming more binary, as depicted on Figure 43. Equation 355 constrains m to belong to the $]1,\infty[$ interval.

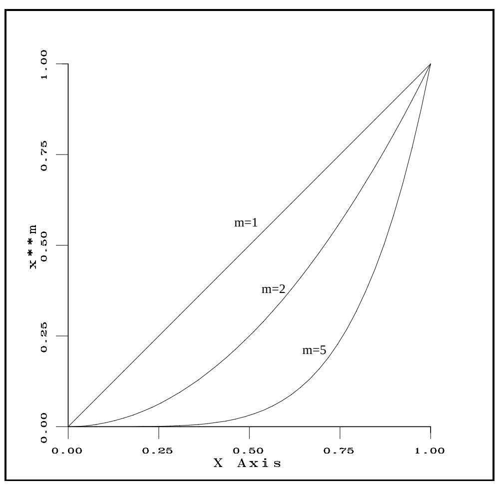


Figure 43: Variation of the term x^m for $x \in [0,1]$ and different values of m.

Equation 355, after one step of centroid updating, guarantees that the membership values remain in the $[0,1]$ interval, and Equation 356 defines the new centroid in agreement with functional J_m. We present on Figure 44, in the case of a 1-dimensionnal 2-class problem, the variations of the membership values of a sample as a function of its position along the x-axis. Both Equations 355 and 356 depend on the argument m. As for Equation 354, m is a way to modulate the membership of a sample, but it varies in a converse direction: from very fuzzy (high values of m) to almost crisp (low values of m). This behavior is illustrated in Figure 45, on the left, and the product μ_{ij}^m as it appears in the functional, on the right. We see from this last figure that the global behavior of the functional is towards the crisp one when m goes to 1.

We see from Figure 45, on the left and right parts of the curves, one of the main drawbacks of fuzzy C-means classification. Away from the centroid of its class, the membership function decays as given by Equation 355, and, as for a given x_i, the membership functions μ_{ij} sum up to 1 for every class $\mathcal{C}_j$, the membership to the other class increases, a behavior which is usually not wanted. Let us illustrate this point with a toy-example. Suppose we want to discriminate *long* snakes from *short* snakes described by

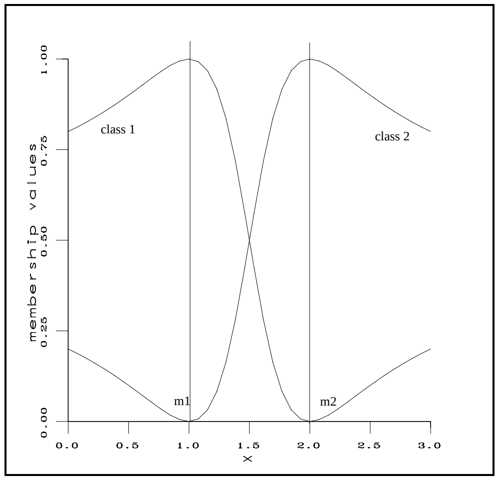
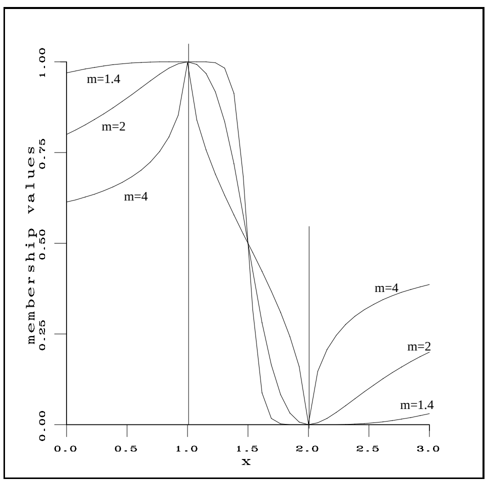


Figure 44: Membership values in a fuzzy C-means classification as a function of x in the case of 2 classes, with centroids in positions $x = 1$ and $x = 2$. In this example $m = 2$.



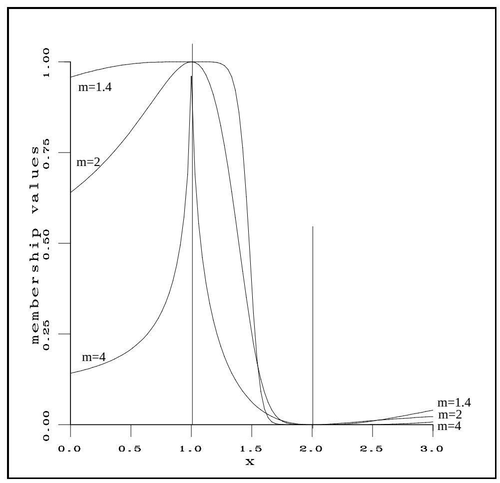


Figure 45: Fuzzy C-means classification in the case of 2 classes, with centroids in positions $x = 1$ and $x = 2$, for 3 different values of m: $m = 1.4$, $m = 2$ and $m = 4$. Left figure: membership value to class $\mathcal{C}_1$ as a function of x. Right figure: the term μ_{i1}^m of the functional J_m.

their length x expressed in meters. And suppose we find a positioning of the cluster centers respectively at 1m and at 2m. We may expect a hard classification to propose a limit somewhere around 1.5 meters. A fuzzy C-means classification may provide a classification as depicted in Figure 44. A problem arises with a 20cm long snake which will be less *short* and more *long* than a 1m snake!

A second consequence of normalization is that, in the case of a more than 2 classes, the effect of normalization in Equation 355 may provide unwanted ripples on the membership functions which are hard to interpret. This point is illustrated in Figure 46 where 2 classes ($\mathcal{C}1$ and $\mathcal{C}2$) have been chosen rather close to enhance the ripples.

If the normalization constraint is suppressed, the optimization of the functional equation converges towards the degenerated solution where all the membership values are zero.

Variations around fuzzy C-means: Ruspini's classifier In an approach anterior to the previous one, Ruspini proposed to optimize different objective functionals instead of the one of Equation 354 [150].

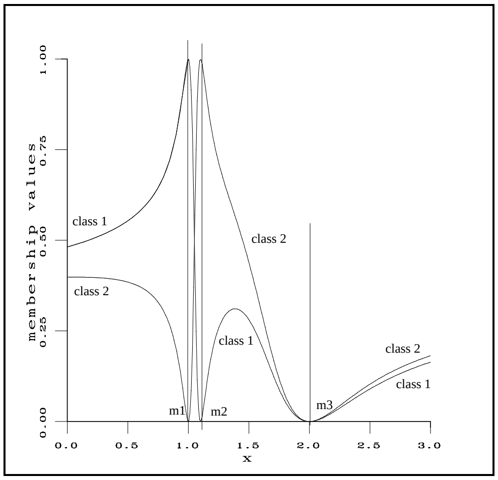


Figure 46: 3 class problem (with centroids at $x = 1$, $x = 1.1$ and $x = 2$). The membership functions to the only classes $\mathcal{C}_1$ and $\mathcal{C}_2$ are represented. They exhibit very irregular shapes with unwanted ripples close from the 'limits" of classes.

In the most popular work, he optimizes :

$$J = \sum_{i=1}^{N}\sum_{l=1}^{N}\left\{\left[\sum_{j=1}^{k}\sigma(\mu_{ij}-\mu_{lj})^2\right]-\delta_{il}^2\right\}^2 .$$

where σ is a constant and δ_{il} a dissimilarity measure between classes $\mathcal{C}_i$ and classes $\mathcal{C}_l$, by using a gradient technique taking benefit of the differentiability of space X with respect to the μ_{ij}. Despites the interesting properties of the method (merely in a 2 class problem where mathematical properties of the clusters may be mathematically stated), Ruspini's method has met little developments since its criterion is not intuitively very appealing and its computational efficiency is low.

5.7.3 Possibilistic C-means

An alternative solution to fuzzy C-means classification, which avoids the normalization drawbacks, is given by Possibilistic C-means [113].

The objective functional in the case of Possibilistic C-means is defined as:

$$J = \sum_{j=1}^{k}\sum_{i=1}^{N}\mu_{ij}^m.||x_i-m_j||^2 + \sum_{j=1}^{k}\eta_j\sum_{i=1}^{N}(1-\mu_{ij})^m.||x_i-m_j||^2 \tag{359}$$

and the membership function to the class $\mathcal{C}_j$ as:

$$\mu_{ij} = \frac{1}{1+\frac{||x_i-m_j||^2}{\eta_j}^{\frac{1}{m-1}}} \tag{360}$$

Here, η_j is a parameter associated to cluster $\mathcal{C}_j$ which controls the decay of the membership function and more precisely which determines the distance from the class center where the membership funtion equals $1/2$. The parameter m is similar to the one used in fuzzy C-means to control the amount of "fuzzines" in the classification and the behavior of the membership functions with respect to m is similar.

We see in Figure 47 that the Possibilistic C-means algorithm having not the normalization to 1 of the membership functions for a given x, behaves according to the intuition for large distances to the cluster center. Coming back to the snake example, if the 20 cm snake is still less a member of the *short snake*

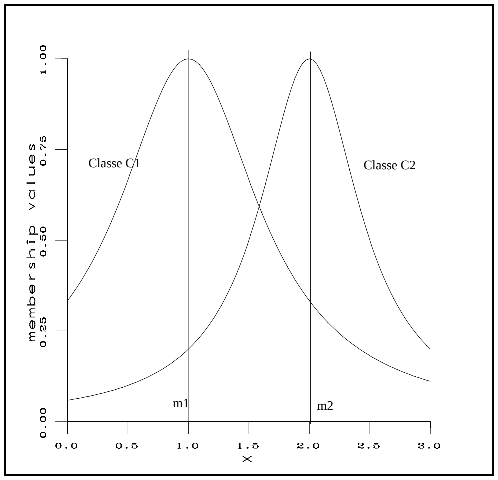


Figure 47: 2 class problem (with centroids at $x = 1$, $x = 2$). Membership functions to class $\mathcal{C}_1$ and $\mathcal{C}_2$ as obtained with Possibilistic C-means, with $m = 1$ and 2 different cluster dimensions: $\eta_1 = 0.5$ and $\eta_2 = 0.25$. These curves have to be compared with the 2 curves obtained with fuzzy C-means in the same conditions (Figure 44).

class than a 1 m snake (since it is farther from the centroid of the class), at least it is no longer more a member of *long snake* class.

The determination of the appropriate values of η_j is a difficult problem. If in some rare situations we have prior information about some relevant values of η_j, most of the time these values have to be derived from the data.

In [113], two solutions have been proposed:

1. to relate the width of the cluster to the within-cluster variance:

$$\eta_j = \frac{\sum_{i=1}^{N} \mu_{ij} ||x_i - m_j||^2}{\sum_{i=1}^{N} \mu_{ij}}$$

2. to relate the width of the cluster to those samples of the cluster the membership function of which is greater than a given value α :

$$\eta_j = \frac{\sum_{i \in \{\mu_{ij}\}_\alpha} ||x_i - m_j||^2}{|\{\mu_{ij}\}_\alpha|}$$

   where $\{\mu_{ij}\}_\alpha$ denotes the $\alpha-$cut of μ.

It is possible to change iteratively the values of η_j, but this doesn't guarantee that the method will still converge towards a stable classification.

5.7.4 Fuzzy k-nearest neighbors

The method of the k-nearest neighbors has been developed in the crisp case as an alternative to the Bayesian Maximum A Posteriori (MAP) classifier in the case where the probability densities of the different classes are unknown, but a set of yet classified samples is available. It may be traced back to the early 50's [81] but has been most popular since the publication of [56] which states many propreties of this classification technique.

Let $Y = \{y_i, i = 1, N'\}$ be the set of classified samples, and $X = \{x_i, i = 1, N\}$ the set of samples to be classified. In the k-nearest neighbors method, we first determine for each unknown sample x_i, the subset of the k samples of Y, closest from x_i than any other y_j. Let $\gamma_k(x_i)$ denote this subset. The k-nearest neighbors rule estimates the probability of x_i to belong to a class $\mathcal{C}_j$ as the ratio over k of the number n_j of samples of $\gamma_k(x_i)$ classified as members of $\mathcal{C}_j$:

$$\hat{p_j} = \frac{n_j}{k}$$

The decision to attribute x_i to a class $\mathcal{C}_j$ is taken according to the maximal estimated probability $\hat{p_j}$:

$$x_i \in \mathcal{C}_j \qquad iff \quad \hat{p_j} = max_{j'}\{\hat{p_{j'}}\}$$

The benefit of the k-nearest neighbors method over more conventional Bayesian methods like for instance Parzen's windows [78] is twofold:

1. no additional information is needed on the probability functions of the samples,
2. it is always possible to estimate the probability $\hat{p_j}$, even in case of very sparse data.

Good classification properties have been demonstrated in the case where the number N' tends to infinity. For instance, if $k = 1$, the method, called "nearest-neighbor", has a probability of error at most twice as large as the Bayesian error, and for increasing k, we have:

$$P_B \leq P_{k+q} \leq P_k \leq 2P_B$$

for any $k \geq 1$ and for any $q \geq 0$, where P_B is the Bayesian error, and P_k the k-nearest neighbor error.

5.8 Local operations for filtering or edge detection

In this section, we summarize the main techniques for local filtering in a broad sense, aiming at enhancing the contrast of an image, at suppressing noise, at extracting contours, etc. Note that these aims are different and often contradicting each other. However, the principles of the techniques are similar, and they can be grouped into two classes: techniques based on functional optimization on the one hand, and rule based techniques on the other hand. These aspects have been largely developed in the literature (see e.g. [10, 118, 167, 2]), and we provide here just the main lines.

5.8.1 Functional approaches

These techniques consist in minimizing or maximizing a functional, which can be interpreted as an analytical representation of some objective. For instance, enhancing the contrast of an image according to this technique amounts to reduce the fuzziness of the image. This can be performed by a simple modification of membership functions (for instance using intensification operators), by minimizing a fuzziness index such as entropy, or even by determining an optimal threshold value (for instance optimal in the sense of minimizing a fuzziness index) which provides an extreme enhancement (until binarization) [133, 134].

Other methods consist in modifying classical filters (median filter for instance) by incorporating fuzzy weighting functions [119].

5.8.2 Fuzzy rule-based techniques

Rule based techniques rely on ideal models (of filters, contours, etc.). These ideal cases being rare, variations and differences with respect these models are permitted through fuzzy representations of these models, as fuzzy rules.

For instance, a smoothing operator can be expressed by [151, 152]:

IF	a pixel is *darker* than its neighbors
THEN	*increase* its grey level
ELSE IF	the pixel is *lighter* than its neighbors
THEN	*decrease* its grey level
OTHERWISE	keep it unchanged

In this representation, the emphasized terms are defined by fuzzy sets or fuzzy operations. Typically, the grey level characteristics are defined by linguistic variables, the semantics of which are constituted by fuzzy sets on the grey level interval. Actions are fuzzy functions applied on grey levels and on pixels. The implementation of these fuzzy rules follows the general principles described above for fuzzy logic.

Rules are sometimes but a different representation of functional approaches. For instance, if a contour detector is defined as:

IF	a pixel *belongs to the contour*
THEN	*increase a lot* its grey level
OTHERWISE	*decrease a lot* its grey level

then it is clear that the final result is an image where the grey level of each pixel in an increasing function of the image gradient at that pixel, which is equivalent to a functional representation.

More complex rules can be found, for instance in [117, 137], where a contour detector is expressed by a set of rules involving the gradient, the symmetry and the stiffness of the contour. Fuzzy rule based systems have also been proposed for contour linking, based on proximity and alignment criteria.

One of the advantages of this approach is that the representation is close to intuition. This leads to an easy design of adaptive operators. For instance, rules can help in deciding which type of operator should be applied depending on the local context.

6 Possible extensions

This contribution could be extended in future specific actions in the following directions:

- knowledge representation and uncertainty in images, including:
  - numerical representations of imperfect knowledge,
  - symbolic representations of imperfect knowledge,
  - knowledge-based systems,
  - reasoning and inference modes;
- numerical methods for information fusion and decision making, in particular:
  - belief function theory,
  - fuzzy and possibilistic fusion;
- spatial information in fusion methods;
- logical approaches for fusion and decision making.

References

[1] *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, CA, 1991.

[2] K. Arakawa. Fuzzy Rule-Based Image Processing with Optimization. In E. E. Kerre and M. Nachtegael, editors, *Fuzzy Techniques in Image Processing*, Studies in Fuzziness and Soft Computing, chapter 8, pages 222–247. Physica-Verlag, Springer, 2000.

[3] B. De Baets. Idempotent Closing and Opening Operations in Fuzzy Mathematical Morphology. In *ISUMA-NAFIPS'95*, pages 228–233, College Park, MD, September 1995.

[4] B. De Baets and E. Kerre. An Introduction to Fuzzy Mathematical Morphology. In *NAFIPS'93*, pages 129–133, Allentown, Pennsylvania, 1993.

[5] H. Bandemer and W. Näther. *Fuzzy Data Analysis.* Theory and Decision Library, Serie B: Mathematical and Statistical Methods. Kluwer Academic Publisher, Dordrecht, 1992.

[6] W. Bandler and L. Kohout. Fuzzy Power Sets and Fuzzy Implication Operators. *Fuzzy Sets and Systems*, 4:13–30, 1980.

[7] G. Banon. Distinction entre plusieurs sous-ensembles de mesures floues. In *Colloque International sur la Théorie des Ensembles Flous*, Marseille, France, 1978.

[8] E. Benoit and L. Foulloy. Capteurs flous multicomposantes : applications à la reconnaissance des couleurs. In *Les Applications des Ensembles Flous*, pages 167–176, Nîmes, France, October 1993.

[9] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms.* Plenum, New-York, 1981.

[10] J. C. Bezdek, J. Keller, R. Krishnapuram, and N. R. Pal. *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing.* Handbooks of Fuzzy Sets series. Kluwer Academic Publisher, Boston, 1999.

[11] J. C. Bezdek and S. K. Pal. *Fuzzy Models for Pattern Recognition.* IEEE Press, New-York, 1992.

[12] D. Bhandari, N. R. Pal, and D. D. Majumder. Fuzzy Divergence, Probability Measure of Fuzzy Events and Image Thresholding. *Pattern Recognition Letters*, 13:857–867, 1992.

[13] E. Binaghi, A. Della Ventura, A. Rampini, and R. Schettini. Fuzzy Reasoning Approach to Similarity Evaluation in Image Analysis. *International Journal of Intelligent Systems*, 8:749–769, 1993.

[14] I. Bloch. About Properties of Fuzzy Mathematical Morphologies: Proofs of Main Results. Technical report, Télécom Paris 93D023, December 1993.

[15] I. Bloch. Triangular Norms as a Tool for Constructing Fuzzy Mathematical Morphologies. In *Int. Workshop on "Mathematical Morphology and its Applications to Signal Processing"*, pages 157–161, Barcelona, Spain, May 1993.

[16] I. Bloch. Distances in Fuzzy Sets for Image Processing derived from Fuzzy Mathematical Morphology (invited conference). In *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 1307–1312, Granada, Spain, July 1996.

[17] I. Bloch. Fuzzy Geodesic Distance in Images. In A. Ralescu and T. Martin, editors, *Lecture Notes in Artificial Intelligence: Fuzzy Logic in Artificial Intelligence, towards Intelligent Systems*, pages 153–166. Springer Verlag, 1996.

[18] I. Bloch. Fuzzy Relative Position between Objects in Images: a Morphological Approach. In *IEEE Int. Conf. on Image Processing ICIP'96*, volume II, pages 987–990, Lausanne, September 1996.

[19] I. Bloch. Image Information Processing using Fuzzy Sets. In *World Automation Congress, Soft Computing with Industrial Applications*, pages 79–84, Montpellier, France, May 1996.

[20] I. Bloch. Incertitude, imprécision et additivité en fusion de données : point de vue historique. *Traitement du Signal*, 13(4):267–288, 1996.

[21] I. Bloch. Information Combination Operators for Data Fusion: A Comparative Review with Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 26(1):52–67, 1996.

[22] I. Bloch. Using Fuzzy Mathematical Morphology in the Dempster-Shafer Framework for Image Fusion under Imprecision. In *IFSA'97*, pages 209–214, Prague, June 1997.

[23] I. Bloch. Fuzzy Relative Position between Objects in Image Processing: a Morphological Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7):657–664, 1999.

[24] I. Bloch. 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*, 7(2):99–133, 1999.

[25] I. Bloch. On Fuzzy Distances and their Use in Image Processing under Imprecision. *Pattern Recognition*, 32(11):1873–1895, 1999.

[26] I. Bloch. Fuzzy Mathematical Morphology and Derived Spatial Relationships. In E. Kerre and N. Nachtegael, editors, *Fuzzy Techniques in Image Processing*, chapter 4, pages 101–134. Springer Verlag, 2000.

[27] I. Bloch. Geodesic Balls in a Fuzzy Set and Fuzzy Geodesic Mathematical Morphology. *Pattern Recognition*, 33(6):897–905, 2000.

[28] I. Bloch. On Links between Mathematical Morphology and Rough Sets. *Pattern Recognition*, 33(9):1487–1496, 2000.

[29] I. Bloch. Spatial Representation of Spatial Relationships Knowledge. In A. G. Cohn, F. Giunchiglia, and B. Selman, editors, *7th International Conference on Principles of Knowledge Representation and Reasoning KR 2000*, pages 247–258, Breckenridge, CO, 2000. Morgan Kaufmann, San Francisco, CA.

[30] I. Bloch. Modal Logics based on Mathematical Morphology for Spatial Reasoning. *Journal of Applied Non Classical Logics*, 12(3-4):399–424, 2002.

[31] I. Bloch. On Fuzzy Spatial Distances. *Advances in Imaging and Electron Physics*, 2003.

[32] I. Bloch and A. Hunter (Eds). Fusion: General Concepts and Characteristics. *International Journal of Intelligent Systems*, 16(10):1107–1134, oct 2001.

[33] I. Bloch and J. Lang. Towards Mathematical Morpho-Logics. In B. Bouchon-Meunier, J. Gutierrez-Rios, L. Magdalena, and R. Yager, editors, *Technologies for Constructing Intelligent Systems*, pages 367–380. Springer, 2002.

[34] I. Bloch and H. Maître. Constructing a Fuzzy Mathematical Morphology: Alternative Ways. In *Second IEEE International Conference on Fuzzy Systems, FUZZ IEEE 93*, pages 1303–1308, San Fransisco, California, March 1993.

[35] I. Bloch and H. Maître. Mathematical Morphology on Fuzzy Sets. In *Int. Workshop on "Mathematical Morphology and its Applications to Signal Processing"*, pages 151–156, Barcelona, Spain, May 1993.

[36] I. Bloch and H. Maître. Fusion de données en traitement d'images : modèles d'information et décisions. *Traitement du Signal*, 11(6):435–446, 1994.

[37] I. Bloch and H. Maître. Fuzzy Mathematical Morphology. *Annals of Mathematics and Artificial Intelligence*, 10:55–84, 1994.

[38] I. Bloch and H. Maître. Fuzzy Mathematical Morphologies: A Comparative Study. *Pattern Recognition*, 28(9):1341–1387, 1995.

[39] I. Bloch and H. Maître. Fusion of Image Information under Imprecision. In B. Bouchon-Meunier, editor, *Aggregation and Fusion of Imperfect Information*, Series Studies in Fuzziness, pages 189–213. Physica Verlag, Springer, 1997.

[40] I. Bloch, H. Maître, and M. Anvari. Fuzzy Adjacency between Image Objects. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 5(6):615–653, 1997.

[41] I. Bloch, H. Maître, and M. Minoux. Optimal Matching of 3-D Convex Polyhedra with Applications to Pattern Recognition. *Pattern Recognition and Image Analysis*, 3(2):137–149, 1993.

[42] I. Bloch, C. Pellot, F. Sureda, and A. Herment. Fuzzy Modelling and Fuzzy Mathematical Morphology applied to 3D Reconstruction of Blood Vessels by Multi-Modality Data Fusion. In D. Dubois R. Yager and H. Prade, editors, *Fuzzy Set Methods in Information Engineering: A Guided Tour of Applications*, chapter 5, pages 93–110. John Wiley & Sons, New-York, 1996.

[43] I. Bloch and A. Ralescu. Directional Relative Position between Objects in Image Processing: A Comparison between Fuzzy Approaches. *Pattern Recognition*, 36:1563–1582, 2003.

[44] G. Borgefors. Distance Transforms in the Square Grid. In H. Maître, editor, *Progress in Picture Processing, Les Houches, Session LVIII, 1992*, chapter 1.4, pages 46–80. North-Holland, Amsterdam, 1996.

[45] B. Bouchon-Meunier, M. Rifqi, and S. Bothorel. Towards General Measures of Comparison of Objects. *Fuzzy Sets and Systems*, 84(2):143–153, September 1996.

[46] B. Bouchon-Meunier and R. R. Yager. Entropy of Similarity Relations in Questionnaires and Decision Trees. In *Second IEEE Int. Conf. on Fuzzy Systems*, pages 1225–1230, San Francisco, California, March 1993.

[47] L. Boxer. On Hausdorff-like Metrics for Fuzzy Sets. *Pattern Recognition Letters*, 18:115–118, 19997.

[48] J. J. Buckley and E. Eslami. Fuzzy Plane Geometry I: Points and Lines. *Fuzzy Sets and Systems*, 86:179–187, 1997.

[49] M. Cayrol, H. Farreny, and H. Prade. Fuzzy Pattern Matching. *Kybernetes*, 11:103–116, 1982.

[50] B. Charroux. *Image Analysis: Interpretation-Guided Cooperation between Segmentation Operators (in French)*. PhD thesis, University Paris XI, January 1996.

[51] B. B. Chauduri and A. Rosenfeld. On a Metric Distance between Fuzzy Sets. *Pattern Recognition Letters*, 17:1157–1160, 1996.

[52] S. Chauvin. *Evaluation of Decision Theories Applied to Data Fusion in Satellite Imaging (in French)*. PhD thesis, ENST and Nantes University, December 1995.

[53] S. M. Chen, M. S. Yeh, and P. Y. Hsio. A Comparison of Similarity Measures of Fuzzy Values. *Fuzzy Sets and Systems*, 72:79–89, 1995.

[54] Y. Cheng and R. L. Kashyap. A Study of Associative Evidential Reasoning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-11(6):623–631, 1989.

[55] C. C. Chu and J. K. Aggarwal. The Integration of Image Segmentation Maps using Region and Edge Information. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(12):1241–1252, 1993.

[56] T.M. Cover and P.E. Hart. Nearest Neighbor Pattern Classification. *IEEE Trans on Information Theory*, IT-13(1):21–27, 1967.

[57] V. Cross and C. Cabello. A Mathematical Relationship Between Set-Theoretic and Metric Compatibility Measures. In *ISUMA-NAFIPS'95*, pages 169–174, College Park, MD, September 1995.

[58] B. V. Dasarathy. Fusion Strategies for Enhancing Decision Reliability in Multi-Sensor Environments. *Optical Engineering*, 35(3):603–616, March 1996.

[59] L. C. de Barros, R. C. Bassanezi, and P. A. Tonelli. On the Continuity of the Zadeh's Extension. In *Seventh IFSA World Congress*, volume II, pages 3–8, Prague, June 1997.

[60] C. Demko and E. H. Zahzah. Image Understanding using Fuzzy Isomorphism of Fuzzy Structures. In *IEEE Int. Conf. on Fuzzy Systems*, pages 1665–1672, Yokohama, Japan, March 1995.

[61] P.A. Devijver and J. Kittler. *Pattern Recognition, a Statistical Approach*. Prentice-Hall, Englewood Cliffs, 1982.

[62] V. di Gesu. Mathematical Morphology and Image Analysis: A Fuzzy Approach. In *Workshop on Knowledge-Based Systems and Models of Logical Reasoning*, Reasoning, 1988.

[63] V. di Gesu, M. C. Maccarone, and M. Tripiciano. Mathematical Morphology based on Fuzzy Operators. In R. Lowen and M. Roubens, editors, *Fuzzy Logic*, pages 477–486. Kluwer Academic, 1993.

[64] E. Diday and J.C. Simon. Cluster Analysis. In Springer Verlag, editor, *Digital Pattern Recognition*, pages 47–94. K.S. Fu, Berlin, 1976.

[65] D. Dubois and M.-C. Jaulent. A General Approach to Parameter Evaluation in Fuzzy Digital Pictures. *Pattern Recognition Letters*, 6:251–259, 1987.

[66] D. Dubois, J. Lang, and H. Prade. Fuzzy Sets in Approximate Reasoning, Part II: Logical Approaches. *Fuzzy Sets and Systems*, 40:203–244, 1991.

[67] D. Dubois and H. Prade. *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, New-York, 1980.

[68] D. Dubois and H. Prade. Inverse Operations for Fuzzy Numbers. In E. Sanchez and M. Gupta, editors, *Fuzzy Information, Knowledge Representation and Decision Analysis, IFAC Symposium*, pages 391–396, Marseille, France, July 1983.

[69] D. Dubois and H. Prade. On Distance between Fuzzy Points and their Use for Plausible Reasoning. In *Int. Conf. Systems, Man, and Cybernetics*, pages 300–303, 1983.

[70] D. Dubois and H. Prade. A Review of Fuzzy Set Aggregation Connectives. *Information Sciences*, 36:85–121, 1985.

[71] D. Dubois and H. Prade. *Possibility Theory*. Plenum Press, New-York, 1988.

[72] D. Dubois and H. Prade. Fuzzy Sets in Approximate Reasoning, Part I: Inference with Possibility Distributions. *Fuzzy Sets and Systems*, 40:143–202, 1991.

[73] D. Dubois and H. Prade. A Glance at non-Standard Models and Logics of Uncertainty and Vagueness. Technical Report IRIT/91-97/R, IRIT, Toulouse, France, 1991.

[74] D. Dubois and H. Prade. Combination of Information in the Framework of Possibility Theory. In M. Al Abidi et al., editor, *Data Fusion in Robotics and Machine Intelligence*. Academic Press, 1992.

[75] D. Dubois and H. Prade. A Unifying View of Comparison Indices in a Fuzzy Set-Theoretic Framework. In R. R. Yager, editor, *Fuzzy Sets and Possibility Theory*, pages 3–13. Pergamon Press, 1992.

[76] D. Dubois, H. Prade, and C. Testemale. Weighted Fuzzy Pattern Matching. *Fuzzy Sets and Systems*, 28:313–331, 1988.

[77] D. Dubois, H. Prade, and R. Yager. Merging Fuzzy Information. In J.C. Bezdek, D. Dubois, and H. Prade, editors, *Handbook of Fuzzy Sets Series, Approximate Reasoning and Information Systems*, chapter 6. Kluwer, 1999.

[78] R.O. Duda and O.E. Hart. *Pattern Classification and Scene Analysis.* John Wiley and Sons, New York (USA), 1973.

[79] J.C. Dunn. A Fuzzy Relative of the ISODATA Process and its Use in Detecting Compact Well Separated Clusters. *J. Cybern.*, 3:32–57, 1974.

[80] A. Edwards and L. Cavalli-Sforza. A Method for Cluster Analysis. *Biometrics*, 21:362–376, 1965.

[81] E. Fix and J.L. Hodges. Discriminatory Analysis, Non-parametric Discrimination: Consistency Properties. Technical Report 4, USAF School of Aviation Medicine, Randolf Fields, TX, (USA), 1951.

[82] M. J. Frank. On the Simultaneous Associativity of $f(x, y)$ and $x + y - f(x, y)$. *Aequationes Mathematicae*, 19:194–226, 1979.

[83] K.S. Fu. *Syntactic Methods in Pattern Recognition.* Academic Press, New York (USA), 1974.

[84] K. Fukunaga. *Introduction to Statistical Pattern Recognition.* Academic Press, New York (USA), 1972.

[85] J. Gasos and A. Ralescu. Towards a Linguistic Instructions-Based Navigation Support System Using Environment Information for Guiding Scene Interpretation. In *FUZZ-IEEE/IFES'95*, pages 1261–1266, Yokohama, Japan, 1995.

[86] S. J. Gee and A. M. Newman. RADIUS: Automating Image Analysis Through Model-Supported Exploitation. In *Image Understanding Workshop*, pages 185–196, Washington D.C., 1993.

[87] T. Géraud, I. Bloch, and H. Maître. Atlas-guided Recognition of Cerebral Structures in MRI using Fusion of Fuzzy Structural Information. In *CIMAF'99 Symposium on Artificial Intelligence*, pages 99–106, La Havana, Cuba, 1999.

[88] G. Gerla and L. Scarpati. Extension Principles for Fuzzy Set Theory. *Journal of Information Sciences*, 106:49–69, 1998.

[89] C. R. Giardina and D. Sinha. Image Processing using Pointed Fuzzy Sets. In *SPIE Intelligent Robots and Computer Vision VIII: Algorithms and Techniques*, volume 1192, pages 659–668, 1989.

[90] V. Goetcherian. From Binary to Grey Tone Image Processing using Fuzzy Logic Concepts. *Pattern Recognition*, 12:7–15, 1980.

[91] R. Goetschel and W. Voxman. Topological Properties of Fuzzy Numbers. *Fuzzy Sets and Systems*, 10:87–99, 1983.

[92] R. Goetschel and W. Woxman. Elementary Fuzzy Calculus. *Fuzzy Sets and Systems*, 18:31–43, 1986.

[93] J. A. Goguen. The Logic of Inexact Concepts. *Synthese*, 19:325–373, 1969.

[94] M. Grabisch. Fuzzy Integral in Multicriteria Decision Making. *Fuzzy Sets and Systems*, 69:279–298, 1995.

[95] M. Grabisch, T. Murofushi, and M. Sugeno. Fuzzy Measures of Fuzzy Events Defined by Fuzzy Integrals. *Fuzzy Sets and Systems*, 50:293–313, 1992.

[96] H. Hamacher. Ueber logische Verknupfungen Unscharfer Aussagen und deren Zugehoerige Bewertungsfunktionen. *Progress in Cybernetics and System Research*, 3:276–287, 1978.

[97] D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge. Comparing Images using the Hausdorff Distance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(9):850–863, 1993.

[98] M. Ishizuka. Inference Methods based on Extended Dempster Shafer's Theory for Problems with Uncertainty / Fuzziness. *New Generation Computing*, 1:159–168, 1983.

[99] R. Jain, S. N. J. Murthy, and P. L. J. Chen. Similarity Measures for Image Databases. In *IEEE Int. Conf. on Fuzzy Systems*, pages 1247–1254, Yokohama, Japan, March 1995.

[100] M. C. Jaulent and A. Yang. Application of Fuzzy Pattern Matching to the Flexible Interrogation of a Digital Angiographies Database. In *IPMU*, pages 904–909, Paris, France, 1994.

[101] A. Kandel and W. J. Byatt. Fuzzy Sets, Fuzzy Algebra, and Fuzzy Statistics. *Proceedings of the IEEE*, 66(12):1619–1639, 1978.

[102] A. Kaufman. *Introduction to the Theory of Fuzzy Subsets: Fundamental Theoretical Elements.* Academic Press, New York, 1973.

[103] A. Kaufmann. *Introduction to the Theory of Fuzzy Subsets.* Academic Press, New-York, 1975.

[104] A. Kaufmann and M. M. Gupta. *Fuzzy Mathematical Models in Engineering and Management Science.* North-Holland, Amsterdam, 1988.

[105] J. M. Keller and X. Wang. Comparison of Spatial Relation Definitions in Computer Vision. In *ISUMA-NAFIPS'95*, pages 679–684, College Park, MD, September 1995.

[106] J. M. Keller and X. Wang. Learning Spatial Relationships in Computer Vision. In *FUZZ-IEEE'96*, pages 118–124, New Orleans, 1996.

[107] A. Kitamoto and M. Takagi. Retrieval of Satellite Cloud Imagery Based on Subjective Similarity. In *9th Scandinavian Conference on Image Analysis*, pages 449–456, Uppsala, Sweden, 1995.

[108] G. J. Klir and T. A. Folger. *Fuzzy Sets, Uncertainty, and Information.* Englewood Cliffs, 1988.

[109] L. T. Koczy. On the Description of Relative Position of Fuzzy Patterns. *Pattern Recognition Letters*, 8:21–28, 1988.

[110] M. Koshelev. Fuzzy Logic Explains the Golden Proportion. *International Journal of Intelligent Systems*, 12(5):415–417, 1997.

[111] B. Kosko. Fuzziness vs. Probability. *International Jounal of General Systems*, 17:211–240, 1990.

[112] R. Krishnapuram and J. M. Keller. Fuzzy Set Theoretic Approach to Computer Vision: an Overview. In *IEEE Int. Conf. on Fuzzy Systems*, pages 135–142, San Diego, CA, 1992.

[113] R. Krishnapuram and J. M. Keller. A Possibilistic Approach to Clustering. *IEEE Transactions on Fuzzy Systems*, 1(2):98–110, 1993.

[114] R. Krishnapuram, J. M. Keller, and Y. Ma. Quantitative Analysis of Properties and Spatial Relations of Fuzzy Image Regions. *IEEE Transactions on Fuzzy Systems*, 1(3):222–233, 1993.

[115] S. Kullback. *Information Theory and Statistics.* Wiley, New York, 1959.

[116] P. A. Laplante and C. R. Giardina. Fast Dilation and Erosion of Time Varying Grey Valued Images with Uncertainty. In *SPIE Image Algebra and Morphological Image Processing II*, volume 1568, pages 295–302, 1991.

[117] T. Law, H. Itoh, and H. Seki. Image Filtering, Edge Detection and Edge Tracing using Fuzzy Reasoning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:481–491, 1996.

[118] C. S. Lee and Y. H. Kuo. Adaptive Fuzzy Filter and its Applications to Image Enhancement. In E. E. Kerre and M. Nachtegael, editors, *Fuzzy Techniques in Image Processing*, Studies in Fuzziness and Soft Computing, chapter 6, pages 172–193. Physica-Verlag, Springer, 2000.

[119] C. S. Lee, Y. H. Kuo, and P. T. Yu. Weighted Fuzzy Mean Filters for Image Processing. *Fuzzy Sets and Systems*, 89:157–180, 1997.

[120] R. Lowen and W. Peeters. Anisotropic Semi-Pseudometrics. *Journal of Intelligent and Fuzzy Systems (Submitted)*, 1997.

[121] R. Lowen and W. Peeters. Distances between Fuzzy Sets Representing Grey Level Images. *Fuzzy Sets and Systems*, 99(2):135–150, 1998.

[122] A. De Luca and S. Termini. A Definition of Non-Probabilistic Entropy in the Setting of Fuzzy Set Theory. *Information and Control*, 20:301–312, 1972.

[123] H. Maître, I. Bloch, H. Moissinac, and C. Gouinaud. Cooperative Use of Aerial Images and Maps for the Interpretation of Urban Scenes (invited conference). In *Ascona Workshop "Automatic Extraction of Man-Made Objects from Aerial and Space Images"*, pages 297–306, Ascona, Suisse, 1995.

[124] G. T. Man and J. C. Poon. A Fuzzy-Attributed Graph Approach to Handwritten Character Recognition. In *Second IEEE Int. Conf. on Fuzzy Systems*, pages 570–575, San Francisco, March 1993.

[125] P. Matsakis and L. Wendling. A New Way to Represent the Relative Position between Areal Objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(7):634–642, 1999.

[126] D. McKeown, W. A. Harvey, and J. McDermott. Rule-Based Interpretation of Aerial Imagery. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 7(5):570–585, 1985.

[127] K. Menger. Statistical Metrics. *Proc. National Academy of Siences USA*, 28:535–537, 1942.

[128] K. Miyajima and A. Ralescu. Spatial Organization in 2D Images. In *Third IEEE Int. Conf. on Fuzzy Systems, FUZZ-IEEE'94*, pages 100–105, Orlando, FL, June 1994.

[129] K. Miyajima and A. Ralescu. Spatial Organization in 2D Segmented Images: Representation and Recognition of Primitive Spatial Relations. *Fuzzy Sets and Systems*, 65:225–236, 1994.

[130] K. Miyajima and A. Ralescu. A Fuzzy Logic based Method for Describing the Spatial Relations in an Image (unpublished manuscript). 1997.

[131] T. Murofushi and M. Sugeno. An Interpretation of Fuzzy Measure and the Choquet Integral as an Integral with respect to a Fuzzy Measure. *Fuzzy Sets and Systems*, 29:201–227, 1989.

[132] M. Nakatsuyama. Fuzzy Mathematical Morphology for Image Processing. In *ANZIIS-93*, pages 75–79, Perth, Western Australia, 1993.

[133] S. K. Pal, R. A. King, and A. A. Hashim. Automatic Grey-Level Thresholding through Index of Fuzziness and Entropy. *Pattern Recognition Letters*, 1:141–146, 1983.

[134] S. K. Pal and A. Rosenfeld. Image Enhancement and Thresholding by Optimization of Fuzzy Compactness. *Pattern Recognition Letters*, 7:77–86, 1988.

[135] C. P. Pappis and N. I. Karacapilidis. A Comparative Assessment of Measures of Similarity of Fuzzy Values. *Fuzzy Sets and Systems*, 56:171–174, 1993.

[136] C. Pellot, I. Bloch, A. Herment, and F. Sureda. An Attempt to 3D Reconstruct Vessel Morphology from X-Ray Projections and Intravascular Ultrasounds: Modelling and Fusion. *Computerized Medical Imaging and Graphics*, 20(3):141–151, 1996.

[137] O. Perez-Oramas. *Contribution à une méthodologie d'intégration de connaissances pour le traitement d'images. Application à la détection de contours par règles linguistiques floues.* PhD thesis, Université de Nancy, 2000.

[138] A. Popov. Convexity Indicators based on Fuzzy Morphology. *Pattern Recognition Letters*, 18:259–267, 1997.

[139] A. T. Popov. Morphological Operations on Fuzzy Sets. In *IEE Image Processing and its Applications*, pages 837–840, Edinburgh, UK, July 1995.

[140] W.K. Pratt. *Digital Image Processing.* Wiley Interscience, New-York (USA), 1978.

[141] M. L. Puri and D. A. Ralescu. Différentielle d'une fonction floue. *C. R. Acad. Sc. Paris, Série I*, 293:237–239, 1981.

[142] M. L. Puri and D. A. Ralescu. Differentials of Fuzzy Functions. *Journal of Mathematical Analysis and Applications*, 91:552–558, 1983.

[143] N. S. V. Rao and S. S. Iyengar. Distributed Decision Fusion under Unknown Distributions. *Optical Engineering*, 35(3):617–624, March 1996.

[144] A. Rosenfeld. Fuzzy Graphs. In Zadeh et al., editor, *Fuzzy Sets and their Applications to Cognitive and Decision Processes*, pages 77–95. Academic Press, New York, 1975.

[145] A. Rosenfeld. Fuzzy Digital Topology. *Information and Control*, 40:76–87, 1979.

[146] A. Rosenfeld. The Fuzzy Geometry of Image Subsets. *Pattern Recognition Letters*, 2:311–317, 1984.

[147] A. Rosenfeld. Distances between Fuzzy Sets. *Pattern Recognition Letters*, 3:229–233, 1985.

[148] A. Rosenfeld and S. Haber. The Perimeter of a Fuzzy Set. Technical Report CAR-TR-8, Center for Automation Research, University of Maryland, College Park, Maryland, 1983.

[149] A. Rosenfeld and R. Klette. Degree of Adjacency or Surroundness. *Pattern Recognition*, 18(2):169–177, 1985.

[150] E. Ruspini. Numerical Methods for Fuzzy Clustering. *Inf. Sci.*, 2:319–350, 1970.

[151] F. Russo and G. Ramponi. Introducing the Fuzzy Median Filter. In *Signal Processing VII: Theories and Applications*, pages 963–966, 1994.

[152] F. Russo and G. Ramponi. An Image Enhancement Technique based on the FIRE Operator. In *IEEE Int. Conf. on Image Processing*, volume I, pages 155–158, Washington DC, 1995.

[153] P. K. Saha, F. W. Wehrli, and B. R. Gomberg. Fuzzy Distance Transform: Theory, Algorithms, and Applications. *Fuzzy Sets and Systems*, 86:171–190, 2003.

[154] E. Sanchez. Solutions of Fuzzy Equations with Extended Operations. *Fuzzy Sets and Systems*, 12:237–248, 1984.

[155] B. Schweizer and A. Sklar. Associative Functions and Abstract Semigroups. *Publ. Math. Debrecen*, 10:69–81, 1963.

[156] B. Schweizer and A. Sklar. *Probabilistic Metric Spaces.* North Holland, Amsterdam, 1983.

[157] J. Serra. *Image Analysis and Mathematical Morphology.* Academic Press, London, 1982.

[158] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, 1976.

[159] J. C. Simon. *From Pixels to Features.* V-X, North Holland, Amsterdam, 1989.

[160] D. Sinha and E. Dougherty. Fuzzy Mathematical Morphology. *Journal of Visual Communication and Image Representation*, 3(3):286–302, 1992.

[161] D. Sinha and E. R. Dougherty. Fuzzification of Set Inclusion: Theory and Applications. *Fuzzy Sets and Systems*, 55:15–42, 1993.

[162] D. Sinha, P. Sinha, E. R. Dougherty, and S. Batman. Design and Analysis of Fuzzy Morphological Algorithms for Image Processing. *IEEE Trans. on Fuzzy Systems*, 5(4):570–584, 1997.

[163] C. Sokic and G. Pavlovic-Lazetic. Homogeneous Images in Fuzzy Databases. In *Seventh IFSA World Congress*, volume IV, pages 297–303, Prague, June 1997.

[164] Léa Sombé. *Raisonnements sur des informations incomplètes en intelligence artificielle.* Teknea, Marseille, 1989.

[165] M. Sugeno. *Theory of Fuzzy Integral and its Applications.* Ph.d. thesis, Tokyo Institute of Technology, Tokyo, 1974.

[166] S. C. A. Thomopoulos. Sensor Integration and Data Fusion. *Journal of Robotics Systems*, 7(3):337–372, 1990.

[167] H. R. Tizhoosh. Fuzzy Image Enhancement: An Overview. In E. E. Kerre and M. Nachtegael, editors, *Fuzzy Techniques in Image Processing*, Studies in Fuzziness and Soft Computing, chapter 5, pages 137–171. Physica-Verlag, Springer, 2000.

[168] J.T. Tou and R. C. Gonzalez. *Pattern Recognition Principles.* Addison Wesley Pub. Comp., Reading (England), 1974.

[169] S. Turpin-Dhilly. *Adaptation des outils de la morphologie floue l'analyse de donnes multidimensionnelles.* PhD thesis, Université de Lille, 2000.

[170] A. Tversky. Features of Similarity. *Psychological Review*, 84(4):327–352, 1977.

[171] J. K. Udupa and S. Samarasekera. Fuzzy Connectedness and Object Definition. In *SPIE Medical Imaging*, volume 2431, pages 2–11, San Diego, CA, February 1995.

[172] J. K. Udupa and S. Samarasekera. Fuzzy Connectedness and Object Definition: Theory, Algorithms, and Applications in Image Segmentation. *Graphical Models and Image Processing*, 58(3):246–261, 1996.

[173] L. Wald. Some Terms of Reference in Data Fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3):1190–1193, 1999.

[174] E. L. Walker. Fuzzy Relations for Feature-Model Correspondence in 3D Object Recognition. In *NAFIPS*, pages 28–32, Berkeley, CA, June 1996. IEEE.

[175] W.-J. Wang. New Similarity Measures on Fuzzy Sets and on Elements. *Fuzzy Sets and Systems*, 85:305–309, 1997.

[176] X. Wang, B. De Baets, and E. Kerre. A Comparative Study of Similarity Measures. *Fuzzy Sets and Systems*, 73:259–268, 1995.

[177] R. Willmott. Two Fuzzier Implication Operators in the Theory of Fuzzy Power Sets. *Fuzzy Sets and Systems*, 4:31–36, 1980.

[178] R. R. Yager. On the Measure of Fuzziness and Negation. *International Journal of General Systems*, 5:221–229, 1979.

[179] R. R. Yager. On a General Class of Fuzzy Connectives. *Fuzzy Sets and Systems*, 4:235–242, 1980.

[180] R. R. Yager. On Ordered Weighted Averaging Aggregation Operators in Multi-Criteria Decision Making. *IEEE Transactions on Systems, Man, and Cybernetics*, 18(1):183–190, 1988.

[181] R. R. Yager. Connectives and Quantifiers in Fuzzy Sets. *Fuzzy Sets and Systems*, 40:39–75, 1991.

[182] R. Y. Yager. Entropy Measures under Similarity Relations. *Int. J. General Systems*, 20:341–358, 1992.

[183] R. T. Yeh and S. Y. Bang. Fuzzy Relations, Fuzzy Graphs and their Applications to Cluster Analysis. In K.S. Fu L. A. Zadeh and M. Shimura, editors, *Fuzzy Sets and Their Applications*, pages 125–149. Academic Press, NY, 1975.

[184] V. R. Young. Fuzzy Subsethood. *Fuzzy Sets and Systems*, 77:371–384, 1996.

[185] L. Zadeh. Fuzzy Sets and Information Granularity. In M. Gupta, R. Ragade, and R. Yager, editors, *Advances in Fuzzy Set Theory and Applications*, pages 3–18. North-Holland, Amsterdam, 1979.

[186] L. A. Zadeh. Fuzzy Sets. *Information and Control*, 8:338–353, 1965.

[187] L. A. Zadeh. Similarity Relations and Fuzzy Orderings. *Information Sciences*, 3:177–200, 1971.

[188] L. A. Zadeh. The Concept of a Linguistic Variable and its Application to Approximate Reasoning. *Information Sciences*, 8:199–249, 1975.

[189] L. A. Zadeh. Fuzzy Sets as a Basis for a Theory of Possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.

[190] L. A. Zadeh. Fuzzy Logic = Computing with Words. *IEEE Transactions on Fuzzy Systems*, 4(2):103–111, 1996.

[191] H. J. Zimmermann. *Fuzzy Set Theory and its Applications*. Kluwer Academic Publisher, Boston, 1991.

[192] H. J. Zimmermann and P. Zysno. Latent Connectives in Human Decision Making. *Fuzzy Sets and Systems*, 4:37–51, 1980.

[193] R. Zwick, E. Carlstein, and D. V. Budescu. Measures of Similarity Among Fuzzy Concepts: A Comparative Analysis. *International Journal of Approximate Reasoning*, 1:221–242, 1987.