



DÉPARTEMENT TSI

Les méthodes de raisonnement dans les images

Brique VOIR - Module RASIM

Isabelle BLOCH, Henri MAÎTRE

ENST, département TSI, CNRS UMR 5141 LTCI, 46 rue Barrault, 75634 Paris Cedex 13
Tél : 01 45 81 75 85, Fax : 01 45 81 37 94, E-mail : Isabelle.Bloch@enst.fr

Avril 2004

Table des matières

1	Introduction	9
2	Représentation des connaissances et incertitude	11
2.1	Introduction	11
2.2	Traitement de l'information	11
2.3	Représentations numériques de connaissances imparfaites	12
2.4	Représentations symboliques de connaissances imparfaites	13
2.5	Systèmes à base de connaissances	14
2.6	Modes de raisonnement et inférence	16
3	Probabilités : point de vue historique	19
3.1	Les probabilités dans l'histoire	20
3.1.1	Avant 1660	20
3.1.2	Vers la formulation mathématique bayésienne	21
3.1.3	La prédominance de l'approche fréquentiste : les « objectivistes » . .	23
3.1.4	20ème siècle : retour au subjectivisme	24
3.2	Classes de probabilités, objectivistes et subjectivistes	26
3.3	Postulats fondamentaux pour une logique inductive	27
3.3.1	Postulats fondamentaux	28
3.3.2	Première équation fonctionnelle	29
3.3.3	Deuxième équation fonctionnelle	30
3.3.4	Probabilités déduites des équations fonctionnelles	30
3.3.5	Mesure d'incertitude et théorie de l'information	30
3.3.6	De Finetti et la théorie du pari	31
4	Bases de la théorie des ensembles flous et des possibilités	35
4.1	Définition des concepts fondamentaux des ensembles flous	35

4.1.1	Ensembles flous	35
4.1.2	Opérations ensemblistes : définitions originales de L. Zadeh	37
4.1.3	Structure et types d'ensembles flous	38
4.1.4	α -coupes	38
4.1.5	Cardinal	39
4.1.6	Convexité	39
4.1.7	Nombre flou	40
4.2	Mesures floues	41
4.2.1	Mesure floue d'un ensemble net	41
4.2.2	Exemples de mesures floues	41
4.2.3	Intégrales floues	42
4.2.4	Mesures d'ensembles flous	43
4.2.5	Mesures de flou	44
4.3	Éléments de théorie des possibilités	45
4.3.1	Nécessité et possibilité	45
4.3.2	Distribution de possibilité	46
4.3.3	Sémantique	47
4.4	Opérateurs	47
4.4.1	Complémentation floue	48
4.4.2	Normes et conormes triangulaires	49
4.4.3	Opérateurs de moyenne	56
4.4.4	Sommes symétriques	59
4.4.5	Opérateurs adaptatifs	60
4.5	Variables linguistiques	61
4.5.1	Définition	62
4.5.2	Exemple de variable linguistique	62
4.5.3	Modificateurs	63
4.6	Relations floues	64
4.6.1	Définitions	65
4.6.2	Propriétés d'une relation floue	65
4.6.3	Composition de relations	66
4.6.4	Relations de similarité	67
4.6.5	Relations d'ordre	68
4.7	Logique floue et possibiliste	69

4.7.1	Logique floue	70
4.7.2	Logique possibiliste	73
4.8	Principes généraux de construction d'opérations floues à partir d'opérations binaires	74
4.8.1	Principe d'extension	75
4.8.2	Combinaison d'opérations sur les α -coupes	77
4.8.3	Traduction d'expressions binaires en expressions floues	79
4.8.4	Comparaison	80
5	Théorie des ensembles flous et des possibilités en traitement d'images	81
5.1	Introduction	81
5.2	Représentation de l'information spatiale	82
5.2.1	Objets flous spatiaux	82
5.2.2	Opérations ensemblistes	84
5.2.3	Ensembles flous géométriques	85
5.2.4	Mesures géométriques d'objets flous	86
5.2.5	Mesures géométriques floues d'objets flous	87
5.2.6	Transformations géométriques	88
5.3	Morphologie mathématique floue	89
5.3.1	Définitions	90
5.3.2	Propriétés	92
5.3.3	Exemples	93
5.3.4	Conclusion et extensions	95
5.4	Topologie floue	96
5.4.1	Connexité floue et voisinage	96
5.4.2	Frontière d'un objet flou	97
5.4.3	Adjacence entre deux objets flous	98
5.5	Distances	100
5.5.1	Représentations	100
5.5.2	Distance d'un point à un objet flou	101
5.5.3	Distance entre deux objets flous	102
5.5.4	Distance géodésique dans un ensemble flou	108
5.6	Position relative directionnelle entre deux objets	108
5.6.1	Les principales approches floues	108
5.6.2	Un exemple simple	110

5.7	Classification floue	113
5.7.1	C-moyennes floues	113
5.7.2	C-moyennes possibilistes	115
5.7.3	k-plus proches voisins flous	116
5.8	Opérations locales d'amélioration ou de détection	117
5.8.1	Approches fonctionnelles	117
5.8.2	Techniques à base de règles	117
6	Définitions générales de la fusion d'informations	119
6.1	Choix d'une définition	119
6.2	Caractéristiques générales des données	121
6.3	Numérique / symbolique	124
6.3.1	Données et informations	124
6.3.2	Traitements	124
6.3.3	Représentations	125
6.4	Systèmes de fusion et types d'architecture	125
6.5	Fusion en traitement d'images et fusion dans d'autres domaines	127
7	Fusion en traitement des images	129
7.1	Objectifs de la fusion en traitement des images	129
7.2	Les situations de fusion	132
7.3	Caractéristiques des données en fusion d'images	133
7.4	Contraintes	135
7.5	Aspects numériques et symboliques en fusion d'images	136
8	Méthodes numériques de fusion d'informations	137
8.1	Fusion probabiliste et bayésienne	137
8.1.1	Mesures d'information	137
8.1.2	Modélisation et estimation	139
8.1.3	Combinaison dans un cadre bayésien	139
8.1.4	Combinaison vue comme un problème d'estimation	140
8.1.5	Décision	140
8.1.6	Autres approches en détection	141
8.1.7	Un exemple de fusion bayésienne en imagerie satellitaire	141
8.1.8	Discussion	143

8.2	Fusion dans la théorie des fonctions de croyance	147
8.2.1	Modélisation	147
8.2.2	Estimation des fonctions de masse	150
8.2.3	Combinaison conjonctive	154
8.2.4	Autres modes de combinaison	159
8.2.5	Décision	159
8.2.6	Déduction axiomatique de la règle de combinaison de Dempster-Shafer	161
8.2.7	Exemple d'application en imagerie médicale	165
8.3	Fusion floue et possibiliste	171
8.3.1	Modélisation	171
8.3.2	Définition des fonctions d'appartenance ou des distributions de possibilité	172
8.3.3	Combinaison	174
8.3.4	Décision	180
8.3.5	Exemples d'application	180
8.4	Introduction de l'information spatiale	184
8.4.1	Au niveau de la modélisation	185
8.4.2	Au niveau de la décision	186
8.4.3	Au niveau de la combinaison	186
8.4.4	Exemples d'applications	187
8.5	Conclusion	193
8.5.1	Quelques acquis	193
8.5.2	Quelques perspectives	195
9	Les systèmes à base de connaissance en traitement des images	199
9.1	Un système à base de connaissance en bref	200
9.1.1	Schéma d'ensemble	200
9.1.2	Les ingrédients des systèmes à base de connaissances	201
9.2	La structure d'un SBC pour la vision	203
9.2.1	Réseaux sémantiques	203
9.2.2	Réseaux bayésiens	205
9.3	Bibliographie thématique commentée	205
9.3.1	Les grandes familles de systèmes experts pour le traitement d'images .	205
9.3.2	Articles généraux sur la gestion des connaissances et leur représentation	206
9.3.3	Applications thématiques	206

Chapitre 1

Introduction

Ce cours présente les méthodes avancées de traitement d'images, des méthodes de raisonnement, de fusion et de gestion de l'incertain. Ces approches sont indispensables dans la plupart des applications où la complexité des informations à traiter et des objectifs poursuivis nécessitent de compléter les méthodes classiques d'analyse et de traitement d'images par des méthodes de haut niveau. Cette complexité peut résulter de la géométrie des images, des informations qu'elles contiennent, de l'agencement spatial des structures dans les images, de la multiplication des modes d'acquisition des images, de leurs imperfections et hétérogénéités. Ce cours s'attaque à ces difficultés et présente des solutions ainsi que leurs évolutions les plus récentes. Il montre également comment aborder des problèmes complexes de modélisation d'informations dans les images, d'interprétation de scènes, de fusion d'images, de raisonnement et de prise de décision dans les images.

Le chapitre 2 présente la problématique de la représentation de connaissances imparfaites en traitement d'images et introduit brièvement les principales classes de méthodes qui seront reprises dans les chapitres suivants. Le chapitre 3 donne un point de vue historique sur les probabilités, permettant de mettre en évidence l'évolution du formalisme et du type de connaissances manipulées dans ce cadre. Le chapitre 4 présente les bases de la théorie des ensembles flous et des possibilités. Leur utilisation en traitement d'images pour modéliser et manipuler l'information spatiale imprécise est décrite dans le chapitre 5. La multiplication des systèmes d'acquisition d'images, la diversité des sources d'informations et leur complémentarité ont fait émerger le thème de la fusion d'informations en traitement d'images. Le domaine de la fusion d'informations connaît depuis plusieurs années une forte évolution, rapide et foisonnante. Après une phase de questions, de discussions, voire d'errements, où le domaine de la fusion en traitement du signal et des images n'était pas bien défini, nous sommes maintenant arrivés à une bonne maîtrise des outils de base (souvent importés d'autres domaines) et il est désormais possible d'une part de mener des applications de bout en bout, et d'autre part de développer des outils plus complexes et plus sophistiqués [322]. Il reste néanmoins beaucoup de travail théorique à conduire pour élargir les bases de ces méthodes et de travaux expérimentaux pour en valider l'utilisation. Les principes généraux de la fusion sont décrits dans le chapitre 6 et les spécificités en traitement d'images sont résumées dans le chapitre 7. Les deux chapitres suivants présentent les principales classes de méthodes de fusion d'informations, numériques dans le chapitre 8, et symboliques dans le chapitre 9. Le premier décrit les méthodes s'appuyant sur les théories des probabilités, des ensembles flous et des possibilités, et des fonctions de croyance. Le second s'intéresse aux

systèmes à base de connaissances, aux formalismes logiques, et présente quelques éléments de raisonnement spatial.

Chapitre 2

Représentation des connaissances et incertitude

2.1 Introduction

Dans ce chapitre, nous présentons succinctement les différents modes de représentation de l'information et des connaissances utilisés en traitement et interprétation des images. Les représentations numériques reposant sur les théories des probabilités, des fonctions de croyances, des ensembles flous et des possibilités seront reprises de manière plus détaillée dans les chapitres 3, 8, 4 et 5. Les systèmes à base de connaissances, qui permettent de structurer les informations, les connaissances et les modes d'inférence pour les combiner seront présentés dans leurs grandes lignes au chapitre 9. Les approches symboliques, ainsi que les modes de raisonnement dans différentes logiques, seront juste mentionnées. Elles sortent du cadre numérique, cependant leurs propriétés mériteraient que l'on s'y intéresse plus en image, en particulier dans le cadre de la fusion d'informations.

2.2 Traitement de l'information

Nous considérons l'information au sens le plus large. Ainsi, le terme information s'applique à tout élément susceptible d'être codé pour être conservé, traité ou communiqué [131]. En traitement et interprétation des images, il s'agit le plus souvent d'informations attachées à des mondes réels (observations, mesures, connaissances génériques sur des phénomènes réels, etc.), mais il peut s'agir également de mondes virtuels, dans l'expression des buts de l'utilisateur, et de ses préférences.

Cette notion générique se divise classiquement en deux catégories, celle des connaissances, faisant référence à des classes d'objets, et celle des données, correspondant à des cas, faits, ou objets particuliers.

Selon la distinction proposée dans [131], l'information peut être constituée de données, de faits, et concerne alors une situation particulière bien déterminée (*il y a une forêt à cet endroit*). Elle peut également porter sur l'existence (ou sur une autre propriété) d'une situation indéterminée (*il existe des régions recouvertes de forêts*). Constituée de données statistiques, de prototypes ou d'exemples typiques, elle porte sur un ensemble de situations

particulières. Enfin, s'il s'agit de classes de situations, elle prendra la forme de contraintes, règles génériques, avec ou sans exceptions, de connaissances générales.

Les informations manipulées sont le plus souvent imparfaites. C'est d'ailleurs dans ces imperfections que se trouve une des raisons d'être de la fusion comme nous le verrons au chapitre 6. Ces imperfections se manifestent sous de multiples formes : ambiguïté, biais, bruit, incomplétude, imprécision, incertitude, incohérence et conflit... Il faut y ajouter le caractère variable et évolutif des informations relatives au monde dynamique.

Nous reviendrons largement sur ces imperfections au chapitre 6, et les particulariserons ensuite dans le chapitre 7 pour les images. En résumé, elles sont dues :

- aux phénomènes observés ;
- aux limites des capteurs ;
- aux algorithmes de reconstruction et de traitement ;
- au bruit ;
- au manque de fiabilité (souvent dû aux limites précédentes) ;
- au mode de représentation ;
- aux connaissances et concepts manipulés.

Il est particulièrement important d'inclure ces imperfections dans les représentations et dans les modes de raisonnement.

Les problématiques de l'interprétation d'images et de la fusion rejoignent celles du traitement de l'information en général [131]. Il s'agira ainsi de :

- représenter l'information (afin de la mettre sous forme utile) ;
- stocker, retrouver, expliciter l'information ;
- exploiter l'information pour décider et agir ;
- communiquer l'information.

La difficulté de la résolution de ces problèmes est bien sûr accrue en raison des imperfections. Pour cela, trois positions peuvent être adoptées :

1. une première attitude consiste à éliminer les imperfections autant que possible. Cela passe par exemple par l'amélioration des capteurs et la multiplication des acquisitions ;
2. une deuxième action possible est de tolérer l'imprécision en produisant des algorithmes et des programmes robustes, et en les associant à des procédures pour réparer les échecs ;
3. la troisième possibilité est de raisonner avec l'imperfection. Celle-ci est alors considérée comme un type de connaissance ou d'information, et sa prise en compte nécessite de la modéliser, de développer des modes de raisonnement approximatif, et d'utiliser des métaconnaissances, c'est-à-dire des connaissances sur ces imperfections.

Nous privilégions ici la troisième approche, qui met en œuvre explicitement des techniques de fusion d'informations et de décision.

2.3 Représentations numériques de connaissances imparfaites

Les principales théories numériques permettant de représenter les connaissances imparfaites et de raisonner à partir de celles-ci sont :

- les probabilités (chapitre 3) ;
- les fonctions de croyance (chapitre 8) ;
- les ensemble flous et les possibilités (chapitre 5).

Dans les représentations probabilistes, le langage est constitué de distributions de probabilités sur un référentiel. Elles permettent de prendre en compte de manière rigoureuse des incertitudes aléatoires ou stochastiques. Il est plus difficile de rendre compte des autres formes d'imperfections, à la fois formellement et sémantiquement. L'inférence bayésienne, souvent utilisée en fusion dans les domaines qui nous intéressent, permet un raisonnement abductif (les différents types d'inférence sont présentés dans la section 2.6).

La théorie des fonctions de croyances (ou théorie de Dempster-Shafer [302]) repose sur un langage défini par des fonctions (appelées dans ce cadre fonctions de masse, de croyance et de plausibilité) sur l'ensemble des parties du référentiel. Les représentations permettent de tenir compte à la fois de l'imprécision et de l'incertitude (y compris sous sa forme subjective), de l'ignorance, de l'incomplétude, et donnent accès au conflit. L'inférence par la règle de Dempster réalise une agrégation de type conjonctif des informations combinées.

Dans la théorie des ensembles flous et des possibilités [120, 126, 358, 361], le langage est formé de sous-ensembles flous du référentiel ou de distributions de possibilité sur celui-ci. Il permet de représenter des informations qualitatives, imprécises, vagues. L'inférence se fait par des règles logiques (ou leur équivalent sous forme numérique), réalisant essentiellement un raisonnement de type déductif, pouvant être qualitatif.

Nous reviendrons en détails sur ces trois théories dans les chapitres suivants, mais dès maintenant il faut retenir que :

- elles ne modélisent pas exactement les mêmes concepts ni les mêmes aspects de l'information ;
- elles n'ont pas la même sémantique ;
- elles n'ont pas le même pouvoir de représentation ;
- elles n'ont pas le même pouvoir de raisonnement.

En particulier les deux premiers points font qu'il est illusoire et trompeur de vouloir comparer leur performances sur les mêmes applications¹.

Ces constatations sont également une motivation pour des techniques hybrides de représentations, permettant de représenter simultanément des informations dont les types d'imperfections sont différents. On peut ainsi définir des probabilités d'événements flous, des fonctions de croyances de sous-ensembles flous, etc. Cependant ces approches sont encore peu exploitées en fusion d'informations.

2.4 Représentations symboliques de connaissances imparfaites

L'intelligence artificielle est classiquement définie (dans les travaux de Minsky et de McCarthy par exemple) de deux points de vue :

- du point de vue cognitif : il s'agit de construire des modèles calculables de processus cognitifs, donc des programmes simulant les performances humaines ;
- du point de vue informatique et d'ingénieur : il s'agit d'affecter aux ordinateurs des tâches qui seraient considérées comme intelligentes si elles étaient effectuées par des humains, donc d'étendre les capacités des ordinateurs.

De l'intelligence artificielle sont nées des représentations symboliques des connaissances. Le domaine de la représentation de connaissances (*knowledge representation*) se caractérise par :

¹Cela explique en partie les conclusions contradictoires que l'on trouve à ce sujet dans la littérature.

- la définition d'une représentation comme un ensemble de conventions syntaxiques et sémantiques pour décrire un élément de connaissance ;
- des représentations logiques (dont l'expressivité dépend de la logique utilisée, voir section 2.6) ;
- des représentations compactes (seules les propriétés et caractéristiques pertinentes sont explicitées) ;
- une facilité de manipulation ;
- ce qui est important est effectivement explicite.

La plupart des données manipulées dans les domaines de l'image sont analogiques ou numériques. Les représentations analogiques nécessitent une description complète du monde. Le passage à des représentations logiques, moins coûteuses, plus compactes, nécessite une conversion de l'analogique vers le symbolique.

Les exigences des représentations symboliques se situent à plusieurs niveaux :

- au niveau ontologique : tous les concepts importants doivent être pris en compte ;
- au niveau épistémique : on ne doit pas être obligé d'exprimer ce qui n'est pas connu ;
- au niveau computationnel : la représentation doit permettre un calcul efficace des propriétés exprimées.

Les deux premiers niveaux induisent des contraintes sur le langage et le troisième sur les mécanismes d'inférence.

La communauté de la représentation de connaissances (symbolique) s'intéresse au raisonnement non monotone, au raisonnement automatique, aux logiques de description, aux représentations subjectives (préférences, souhaits...), aux ontologies, etc. [278]. Plus près de notre propos, elle s'intéresse également à l'apprentissage, à l'intégration et la fusion de bases de connaissances, à la décision et au diagnostic, au raisonnement temporel et spatial, à la représentation des actions et à la planification. Il y a sans aucun doute des pistes à explorer dans ce sens.

2.5 Systèmes à base de connaissances

Les systèmes à base de connaissances seront décrits dans le chapitre 9. Précisons ici ce qui relève plus particulièrement de la représentation des connaissances.

De tels systèmes peuvent être vus comme des extensions des systèmes experts classiques, en permettant des modes différents de représentation des connaissances et de raisonnement.

La construction d'un système à base de connaissance implique des acteurs ayant trois rôles distincts :

- l'utilisateur remplit la base de faits avec les données à traiter ;
- l'expert construit la base de connaissances ;
- le développeur construit le moteur d'inférence et la stratégie de raisonnement.

Parmi les différents types de connaissances représentées dans ces systèmes, on distingue les connaissances :

- déclaratives (comment sont les choses) ;
- procédurales (comment on fait) ;
- épisodiques (relatives à l'expérience précédente) ;
- et les métaconnaissances (connaissances sur la connaissance).

Il est en général extrêmement difficile de passer d'un type de connaissance à l'autre (et

même souvent impossible) et donc ces connaissances sont par essence bien différentes et peuvent toutes être nécessaires.

Le contrôle consiste à rechercher des chemins entre les connaissances initiales et les buts, par des techniques de chaînage avant (application de règles d'inférence lorsque de nouvelles données sont déclarées, les conséquences pouvant alors déclencher de nouvelles règles d'inférence), ou chaînage arrière (application de règles d'inférence lorsque de nouvelles requêtes sont formulées, les prémisses non encore vérifiées de ces règles conduisant alors au déclenchement de nouvelles règles).

Les exemples les plus classiques de systèmes à base de connaissances comportent :

- les règles de production ;
- les *frames* ;
- les réseaux sémantiques ;
- les systèmes avec incertitude : Mycin [307], etc.

Les systèmes de règles de production (de la forme *si... et/ou... alors...*) sont des systèmes faciles à adapter ou à étendre et dont le fonctionnement et les résultats peuvent être facilement expliqués. Ils ont l'inconvénient d'une représentation fragmentée de la connaissance induisant un manque d'efficacité. Leur pouvoir d'expression dépend essentiellement du type de logique utilisée. Par exemple la logique du premier ordre ou des prédictats permet de manipuler des variables et des quantificateurs, par opposition à la logique propositionnelle où tout est constant.

Les *frames* constituent une forme déclarative de systèmes à base de connaissances, dans lesquels une liste d'attributs ou de propriétés est assortie de caractéristiques et de valeurs de ces caractéristiques. Ils trouvent leur utilité dans la description de concepts généraux, de classes d'objets. Des liens hiérarchiques, d'héritage, de spécialisation et d'instanciation permettent de manipuler des classes de granularités différentes. Ces systèmes sont le plus souvent statiques, mais une certaine dynamique peut y être introduite en affectant des procédures aux attributs.

Les réseaux sémantiques s'appuient sur une représentation graphique de la base de connaissances, dans laquelle les noeuds représentent les concepts et les objets, et les arcs représentent des relations. Les règles d'inférence s'appuient sur des propriétés d'héritage lorsque l'on passe par des arcs d'une classe à une classe plus spécifique. Ces réseaux sont beaucoup utilisés pour le traitement du langage naturel par exemple.

En vision et en traitement d'images, où l'on opère dans des environnements incomplètement spécifiés et seulement partiellement connus, on trouve des exemples de systèmes à base de connaissances essentiellement pour la supervision de programmes [85, 255, 331] et pour l'interprétation d'images [108, 152, 174, 236, 238].

En vision, des tâches spécifiques de focalisation et adaptation (avec leurs mécanismes attentionnel, de révision ou réparation et de maintien de la cohérence), de coopération et fusion (confrontative, augmentative, intégrative), de coordination (délibérative, réactive, optimale) sont ajoutées aux systèmes à base de connaissances. Ces approches sont détaillées dans [149] et ne sont pas reprises ici.

2.6 Modes de raisonnement et inférence

Les modes d'inférence utilisés dans les systèmes à base de connaissances sont plus variés que dans les systèmes experts classiques. On distingue les modes suivants :

- la déduction, qui fournit des conséquences à partir de faits (par exemple si la base de faits contient A et la proposition $A \rightarrow B$, alors on peut déduire B) ;
- la contraposition permet de raisonner sur les non-observations (par exemple si l'on a $A \rightarrow B$ et non B , on peut inférer non A) ;
- l'abduction cherche à remonter aux causes expliquant les observations (par exemple de $A \rightarrow B$ et de l'observation de B , on infère que A est une cause possible de B) ;
- l'induction permet d'inférer des règles à partir d'observations régulières ou habituelles (par exemple, si l'on a B à chaque fois que l'on a A , on peut inférer $A \rightarrow B$) ;
- la projection fournit des conséquences à partir d'actions (si la base de faits contient la proposition $A \rightarrow B$ et que l'on fait A , on s'attend à ce que B soit réalisé) ;
- la planification établit les actions à effectuer pour attendre des buts (si l'on veut B et que la base contient $A \rightarrow B$, alors on infère l'action A).

Ces deux derniers modes d'inférence sont particulièrement développés dans les systèmes à base de connaissances embarqués, tels que ceux qui sont utilisés par exemple en robotique mobile [289].

L'interprétation d'images et la fusion d'informations imposent souvent de faire appel à différents modes de raisonnement, afin de mieux saisir et représenter les finesse et les subtilités du raisonnement humain.

En raisonnement monotone, l'obtention de plus d'informations conduit naturellement à plus de conclusions : si d'une base KB on déduit A , on déduira également A de $KB \cup B$. Les logiques classiques, propositionnelle et du premier ordre, relèvent de ce mode de raisonnement.

En raisonnement non monotone, une nouvelle information peut invalider des conclusions précédentes. En présence d'informations et de connaissances imparfaites comme c'est le cas en fusion d'informations, les sources de non-monotonie dans le raisonnement viennent essentiellement des hypothèses et des restrictions qui sont effectuées. Celles-ci sont nécessaires pour pouvoir raisonner, mais peuvent être remises en cause si de nouvelles informations ou éléments de connaissance sont disponibles. Ces hypothèses à la source de la non-monotonie comportent :

- l'utilisation de propriétés typiques ;
- les exceptions possibles ;
- l'hypothèse de monde fermé, sur laquelle nous reviendrons en particulier dans le chapitre 8.

Les logiques non monotones reposent sur des notions de préférence (quel monde, quelle situation sont plus « normaux » que d'autres, quels sont les buts préférés si tous ne peuvent être atteints...), de changement ou révision de croyances (les fameux postulats AGM [4]), et bien sûr sur un certain nombre de postulats gérant la non-monotonie, appelés postulats de rationalité. Un exemple de tels postulats, de monotonie « prudente » exprime que si une base KB permet de déduire $A \rightarrow B$ et permet de déduire C , alors elle permet de déduire $A \wedge C \rightarrow B$.

Les notions de contingence, de vérité nécessaire ou possible, ne sont pas bien représentées en logique classique. Manipuler de tels concepts implique l'introduction de modalités dans

la logique. Les logiques modales [78, 183] permettent de raisonner sur des propositions A (A est vrai), $\Box A$ (A est nécessaire), $\Diamond A$ (A est possible). Des formes numériques de ces notions se retrouvent dans la théorie des fonctions de croyance en termes de croyance et plausibilité (voir chapitre 9) et dans la théorie des possibilités en termes de nécessité et possibilité (voir chapitre 5). Nous ne détaillerons pas les acceptations logiques de ces notions dans cet ouvrage.

Enfin, les notions d'imprécision et d'incertitude, dont nous avons déjà beaucoup parlé, peuvent être représentées dans les logiques floues et possibilistes, dont nous donnerons un aperçu au chapitre 5.

Chapitre 3

Probabilités : point de vue historique

Parmi les méthodes de représentations des connaissances, les méthodes numériques cherchant à modéliser l'imprécision et l'incertitude des données et de la connaissance sont largement employées pour des problèmes aussi variés que l'agrégation multi-critères, la combinaison de témoignages, ou encore la fusion d'images hétérogènes. Les méthodes probabilistes sont certainement les plus populaires, mais suscitent pourtant encore de nombreuses polémiques, en particulier entre approches fréquentistes¹, objectivistes, et approches subjectivistes. Si les subjectivistes semblent prendre le pas dans de nombreux domaines, les principes fréquentistes sont souvent d'une grande utilité pratique, en particulier dès qu'il s'agit d'apprendre une loi sur de grands échantillons, par exemple pour reconnaître des cultures dans une image aérienne.

Un parcours historique des différentes acceptations des probabilités permet d'expliquer les origines de ces polémiques et montre que le choix d'une approche peut être raisonnable et justifié par le problème posé et l'interprétation que l'on souhaite donner aux probabilités. La partie 3.1 sera consacrée à ce point de vue historique et la partie 3.2 à la caractérisation des différentes classes de probabilités. Nous nous sommes pour cela largement inspirés des articles de synthèse cités en référence.

Il est remarquable que l'hypothèse d'additivité des probabilités², aujourd'hui communément admise, n'est apparue que très tard. Cette hypothèse est posée de manière axiomatique dans la théorie de Kolmogorov. Cependant les travaux de Cox montrent que ces « axiomes » peuvent être déduits d'un certain nombre de postulats de base dictés par l'intuition (partie 3.3).

Dans la partie 8.1.8, nous donnerons quelques exemples montrant les limites des probabilités additives, dues aux contraintes souvent trop fortes qu'elles imposent. La modification des postulats de base pour dépasser ces limites conduit à des théories numériques différentes, ne satisfaisant plus les mêmes propriétés, et on retrouve ainsi des approches telles que les

¹On trouve aussi le terme de « fréquentiste » dans la littérature, et il semble qu'il n'y ait pas encore de consensus sur cette terminologie.

²La relation d'additivité exprime que pour deux événements A et B exclusifs, la probabilité de la réunion notée $A + B$ vaut $p(A + B) = p(A) + p(B)$. En particulier, on en déduit que $p(A) + p(\bar{A}) = 1$, où \bar{A} désigne le contraire de A (ou son complémentaire en termes ensemblistes).

ensembles flous ou la théorie des croyances de Dempster-Shafer (voir chapitre 8). Il a souvent été reproché à cette dernière que la règle de combinaison orthogonale de Dempster n'avait pas de justification théorique. Plusieurs auteurs ont répondu à ces critiques, et nous présenterons dans le chapitre 8 (section 8.2.6) les arguments de Smets, permettant de déduire cette règle d'axiomes plus facilement justifiables. Nous établissons ensuite les liens entre ces axiomes et ceux de Cox, expliquant ainsi les origines des différences entre les deux théories.

3.1 Les probabilités dans l'histoire

Ce sont les travaux de Shafer, et en particulier ses remarquables articles synthétiques d'histoire des sciences [303, 304], qui ont motivé cette étude. Nous nous en sommes largement inspirés dans la présentation historique qui suit. Ils en constituent la base, et des compléments ont été apportés à partir des articles ou ouvrages [93, 105, 126, 164, 180, 185, 186, 192, 256, 320, 333].

3.1.1 Avant 1660

Dès l'antiquité apparaît l'opposition entre connaissance et opinion, que l'on trouve en particulier chez Platon, autour des années 400 avant notre ère, et des termes comme nécessaire, possible, probable commencent à être définis. On trouve par exemple chez Aristote (au tour des années -350) des assertions du type « si un événement est nécessaire, c'est que son contraire est impossible » (il faudra attendre la théorie des possibilités pour que soit développée une théorie cohérente capable de modéliser cette phrase) ou encore « le probable est ce qui se passe habituellement » (faisant référence à la répétition de phénomènes, base de la théorie fréquentiste). Les anciens distinguaient trois catégories épistémologiques. Dans la première, une connaissance certaine est possible. Cela correspond à la notion de connaissance ou de science chez Platon. La deuxième catégorie comprend les événements pour lesquels une connaissance probable est possible. On retrouve ici la notion d'opinion de Platon, et ainsi la probabilité apparaît comme un attribut de l'opinion. La troisième catégorie, absente de la philosophie de Platon, correspond aux événements pour lesquels aucune connaissance n'est possible, donc au domaine de l'aléatoire. Ce terme s'entend dans le sens de l'exclusion de régularité statistique. Transposées en termes plus modernes, ces notions correspondent, nous semble-t-il, à celles de déduction pour la première catégorie, et d'induction pour la deuxième. La troisième correspondrait à des phénomènes ne suivant (apparemment) aucune loi prédictible ou apprise. Cette troisième catégorie semble exclure toute possibilité d'une théorie mathématique des chances.

Ces notions restent toutefois très primitives et aucune théorie n'est encore développée. Cependant leur importance est incontestable puisqu'elles sont liées à la théorie de la connaissance, considérée comme essentielle, et dont se préoccupaient beaucoup les philosophes. Le fait que les probabilités soient considérées comme « guide de la vie » par Cicéron (-60) l'atteste. Il est difficile de résister ici au plaisir de reprendre ces quelques phrases de Sénèque, citées dans [234] :

Voici en quoi nous ne sommes pas d'accord avec les Étrusques, spécialistes de l'interprétation des foudres. Selon nous, c'est parce qu'il y a collision de nuages que la foudre fait explosion. Selon eux, il n'y a collision que pour que l'explosion se fasse. Comme ils rap-

portent tout à la divinité, ils sont persuadés, non pas que les foudres annoncent l'avenir parce qu'elles ont été formées, mais qu'elles se forment parce qu'elles doivent annoncer l'avenir. (Sénèque, *Questions Naturelles*, II, 32).

Ces phrases illustrent bien le caractère subjectif de l'opinion : il est impossible d'imaginer une expérience qui permettrait de prouver ou de réfuter l'une ou l'autre de ces opinions. Elles illustrent aussi la différence entre causalité et liens logiques, souvent confondus. Les probabilités traduisent des liens logiques mais pas de relation de causalité [105].

Puis ces catégories épistémologiques disparaissent, sans que l'on sache vraiment l'expliquer. À la renaissance, deux notions complètement indépendantes sont manipulées : celle de chance ou d'aléatoire, et celle de probabilité, vue comme attribut de l'opinion et à laquelle n'est attachée aucune valeur numérique. La notion de chance est fortement liée à la théorie des jeux. On en trouve les prémisses dans le *Purgatoire* de Dante (1310), où sont décrites les différentes sommes que l'on peut obtenir en lançant trois dés. La théorie des jeux est ensuite largement développée à la fin du 16ème siècle et dans la première moitié du 17ème siècle. Cardano (1560) et Galilée (1620) recensent les différents résultats qui peuvent se produire dans un jeu et comptent les cas où chacun des résultats se produit. Apparaît ici pour la première fois la notion de « cas équiprobables ». C'est à Pascal et Fermat que l'on attribue les origines de la théorie mathématique des probabilités (bien qu'ils n'emploient pas ce terme), puisque, dans leur correspondance (autour de 1654), ils commencent à résoudre les premiers problèmes non triviaux. En 1657, est publié le premier livre sur le sujet de la théorie des jeux, écrit par Huygens. Ces trois mathématiciens et philosophes essayent de résoudre le « problème des points » : un jeu entre deux joueurs nécessitant qu'un joueur ait trois points pour gagner reste inachevé ; comment alors partager équitablement les enjeux si un joueur a un point et l'autre en a deux ? Ils expliquent avec des cas équiprobables pourquoi des fréquences proportionnelles apparaissent dans une grande série d'essais. Cependant, ils sont générés dans la résolution de leur problème à cause du déterminisme que leur impose le christianisme. Les anciens au contraire acceptaient très bien l'indéterminisme. En résumé, tous ces travaux traitent du problème des jeux et en déduisent une théorie de la chance, mais ne parlent jamais de probabilité, même s'ils introduisent un vocabulaire un peu plus épistémologique. Il demeure une grande confusion entre statistiques et connaissance a priori, conduisant au mélange de deux classes de probabilités, à partir de fréquences (liées aux statistiques) et à partir des cas équiprobables (liés à une connaissance a priori).

3.1.2 Vers la formulation mathématique bayésienne

Le premier lien entre la théorie des jeux et les probabilités apparaît en 1662 où il est introduit par Arnault dans *l'Art de Penser*. Arnault établit une analogie entre les jeux et la vie de tous les jours et suggère qu'un point de vue épistémologique des chances permettrait d'appliquer la théorie aux probabilités (toujours considérées comme attribut de l'opinion). Il s'arrête juste avant le concept de probabilité numérique, mais ses travaux marquent clairement un tournant dans l'évolution de la notion de probabilité. L'analogie entre jeux et vie est exploitée à la fin du 17ème siècle par les démographes qui calculent des tables d'espérance de vie à l'aide de la théorie des jeux, mais sans introduire la notion de probabilité.

Une contribution d'un tout autre domaine est due à Leibniz qui propose dans le *De Conditionibus* (1665) de représenter les droits légaux des personnes par des nombres. L'absence de droit est représentée par 0, un droit pur par 1 et un droit conditionnel par une fraction entre

0 et 1. Cette classification des droits repose sur la condition sur laquelle est fondée le droit : une condition impossible conduit à l'absence de droit, si elle est nécessaire le droit est pur, si elle est contingente³, le droit est conditionnel. Ces notions de contingence ou de nécessité se retrouveront chez J. Bernoulli pour le problème de la combinaison de témoignages. Leibniz propose de relier la probabilité d'existence de la condition à la « grandeur » du droit, et semblait donc se diriger vers une conception numérique des probabilités, sans s'inspirer de la théorie des jeux. Il ne devient familier avec cette théorie que plus tard. Si ses essais sur la chance n'apportent rien de neuf du point de vue mathématique, ils reconnaissent le lien entre probabilités et théorie des jeux.

Dans le domaine de la combinaison de témoignages, les travaux de Hooper en 1699 (*A Calculation of the Credibility of Human Testimony*) conduisent à la définition de fonctions de confiance non bayésiennes, représentant la crédibilité d'un témoin, ainsi qu'à deux règles de combinaison, l'une pour des témoignages successifs, et l'autre pour des témoignages simultanés. Ces deux règles, très populaires au 18ème siècle, seront complètement abandonnées au 19ème siècle.

Une des plus importantes contributions à la relation entre théorie des jeux et probabilités à la fin du 17ème siècle est certainement celle de J. Bernoulli, en particulier dans son ouvrage *Ars Conjectandi* (publié en 1713). Bernoulli propose dès 1680 une théorie mathématique des probabilités et de leur combinaison. Il utilise la théorie des jeux pour calculer les probabilités mais conserve aux probabilités leur aspect épistémologique et leur rôle dans le jugement des individus. Ainsi, les premiers formalismes permettant de manipuler des probabilités numériques étaient consacrés à l'étude de probabilités subjectives ! La plus grande partie de la théorie de Bernoulli est consacrée à des probabilités qui sont subjectives, et qui sont des mesures de la connaissance. Elles sont calculées à partir du concept d'«argument» et leurs propriétés dépendent de la nature de ces arguments. En particulier, elles ne sont pas toujours additives. Les règles de combinaison proposées par Bernoulli, plus complètes que celles de Hooper, prennent, selon les arguments, différentes formes, dont seules certaines correspondent aux règles probabilistes usuelles. Dans la dernière partie de son travail, Bernoulli établit la fameuse loi des grands nombres. Cette loi permet d'estimer a posteriori une probabilité inconnue a priori à partir de l'observation des fréquences d'occurrence. Ce théorème porte donc plus sur les probabilités aléatoires que sur les probabilités épistémologiques (selon la distinction qu'introduira Lambert) puisqu'une chance qui ne peut être connue qu'a posteriori n'est pas initialement une caractéristique de notre connaissance.

Les successeurs de Bernoulli simplifient sa théorie et en réduisent inconsciemment la portée. Tout d'abord, ils ne sont pas convaincus par Bernoulli dont la théorie, qui leur semble souvent compliquée, n'est pas aussi bien établie que celle des jeux. De plus, ils retiennent surtout la loi des grands nombres et identifient probabilité et chance d'apparaître, conduisant à une approche essentiellement fréquentiste. Citons, parmi les successeurs de Bernoulli, Montmort (*Essai d'analyse sur les jeux de hasard*, 1708) qui essaye d'appliquer la théorie des jeux à d'autres domaines et Moivre (*De Mensura Sortis*, 1711, *Doctrine of Chances*, 1718) où l'on trouve la première règle explicite d'additivité et une représentation des probabilités entre 0 et 1. Sa définition devient la définition classique. Ces deux auteurs parlent de probabilités mais leur théorie traite surtout des chances. Les probabilités y sont définies comme le rapport du nombre de cas favorables sur le nombre de cas possibles, mais les auteurs sont confrontés au problème du comptage des cas qui n'est pas possible dans tous

³contingent est employé ici dans le sens : qui peut se produire ou non

les domaines. Notons que Moivre avait déjà découvert la loi gaussienne.

Au 18ème siècle, seul Lambert poursuit les travaux de Bernoulli et distingue probabilités aléatoires et probabilités épistémologiques. Les premières sont celles qui peuvent être connues soit *a priori*, comme dans la théorie des jeux, soit *a posteriori*, données par l'expérience. Les secondes sont affectées à des événements par inférence à partir d'effets ou de circonstances, et ont un caractère subjectif. Dans *Photometrica* (1760), il traite de la théorie des erreurs et propose une méthode connue aujourd'hui sous le nom de maximum de vraisemblance. Dans le *Neues Organon* (1764), il généralise la théorie des arguments de Bernoulli, corrige et généralise ses lois de combinaison et traite le cas des jeux, des syllogismes et des témoignages de divers types. Ses lois sont un cas particulier de la règle de combinaison de Dempster, et là encore les probabilités sont non additives.

Puis les travaux de Bayes (*An Essay Towards Solving a Problem in the Doctrine of Chances*, 1763) proposent une synthèse qui fait disparaître la distinction entre probabilités aléatoires et probabilités épistémologiques. Bayes inverse le théorème de Bernoulli : alors que Bernoulli estime le nombre de succès à partir de la connaissance de la probabilité, Bayes cherche à calculer la probabilité connaissant le nombre de succès dans un échantillon et l'exprime en termes de probabilités initiales, finales et de vraisemblance. Il ne manipule que des probabilités additives. Bayes n'a jamais cherché à faire connaître ses travaux. Son essai n'a été retrouvé qu'après sa mort, avec d'autres travaux (en particulier sur les corps chargés électriquement) écrits avec des abréviations qui n'ont pu être toutes déchiffrées [178]. Cela a conduit à s'interroger sur l'origine exacte du théorème de Bayes (on pourra consulter par exemple les travaux de Stigler [324, 325], qui propose une résolution ... bayésienne de cette interrogation).

Ces travaux sont repris par Laplace (*Théorie analytique des probabilités*, 1812). Cette époque voit l'essor des probabilités inverses, vues du point de vue subjectif. Cette théorie fait apparaître la distinction entre probabilité « initiale » d'une hypothèse (ou *a priori*), probabilité « finale » (après l'expérience), et probabilité de « vraisemblance » (probabilité de l'expérience sachant l'hypothèse). Laplace expose également le principe de raison insuffisante : des réalisations sont considérées équiprobables si l'on n'a aucune raison de penser autrement. Ce principe sera généralement adopté jusqu'au milieu du 20ème siècle.

3.1.3 La prédominance de l'approche fréquentiste : les « objectivistes »

Au 19ème siècle et au début du 20ème siècle, l'essor des sciences physiques conduit à négliger la modélisation du raisonnement humain. À cette époque apparaît une nouvelle discipline : la statistique. La notion de probabilité est alors reliée très souvent à l'observation de phénomènes physiques, à leur répétition dans des séquences longues. On reproche alors à la théorie de Bayes et de Laplace son côté subjectif, on l'accuse de manque de rigueur et on rejette la notion de probabilité *a priori* qui semble trop vague.

Les travaux de Cournot (1843), Ellis (1843), Venn (1866) définissent alors des probabilités physiques, en termes de fréquences. Comme le souligne Good [164], ces travaux sont confrontés à des problèmes insolubles. Par exemple, si en jetant une pièce, on observe la séquence de Pile (P) et Face (F) suivante : PFPFPFPFPF..., on en déduit que la probabilité d'avoir Pile est 1/2, mais cela ne permet pas de conclure sur l'honnêteté du jeu.

Un des problèmes soulevés par ces approches est celui de la longueur des séquences sur

lesquelles sont calculées les fréquences. Elles doivent être longues, mais de combien ? Les théories sont élaborées avec des séquences infinies, mais ne disent pas comment procéder en pratique. C'est le cas de la limite de Venn, des populations hypothétiques infinies de Fisher (1912), ou encore des séquences aléatoires infinies de von Mises (1919). Von Mises pose clairement la distinction entre théorie mathématique abstraite et l'application de cette théorie : la propriété essentielle de ses séquences aléatoires infinies doit être que la probabilité de succès doit être la même quelle que soit la sous-séquence (infinie), ce qui est une notion abstraite, et il se restreint en pratique aux domaines où cette définition est raisonnable. Son argument est qu'il n'est pas nécessaire de répéter effectivement indéfiniment l'expérience pour que la probabilité existe, et il se limite donc à des probabilités physiques et des processus aléatoires, en excluant les problèmes où l'on se demande par exemple quelle est la probabilité pour que X meure à 60 ans.

Le 19ème siècle voit également se développer la loi gaussienne. Déjà connue de Moivre, elle est obtenue par Gauss (1823) en utilisant le principe du maximum de vraisemblance dans des problèmes d'estimation de l'erreur d'observation. Au milieu du 19ème siècle, elle est retrouvée d'une part par Herschel à partir de considérations géométriques pour estimer des erreurs de mesure dans la position d'une étoile, et d'autre part par Maxwell lors de ses études sur les distributions de vitesses des molécules d'un gaz [105].

Cependant, malgré le fort contexte fréquentiste de cette époque, des distinctions proches de celles faites aujourd'hui sont formulées. Par exemple, Poisson, dans ses recherches sur la probabilité des jugements (1837), distingue chance et probabilité. La chance est caractéristique de l'événement lui-même, indépendamment de notre connaissance, alors que la probabilité est relative à notre connaissance. Cette distinction est voisine de celle faite par Lambert. La distinction entre probabilité objective et probabilité subjective est également explicite dans *l'Exposition de la Théorie des Chances et des Probabilités* de Cournot (1843). Mais, même si la distinction est affichée, les théories développées au 19ème siècle permettent seulement de résoudre les problèmes liés aux probabilités physiques ou objectives.

Les travaux de Boole constituent une sorte de lien entre les approches fréquentistes et épistémologiques. Dans son ouvrage *Laws of Thought* (1854), il essaye de combiner à un niveau épistémologique des évaluations faites localement sur les divers attributs de l'information. Il maintient que les probabilités sont déduites de fréquences, mais reconnaît l'impossibilité d'estimer, dans de nombreuses situations, les fréquences jointes, qui doivent alors être générées à un niveau subjectif.

3.1.4 20ème siècle : retour au subjectivisme

Au 20ème siècle, les approches classiques continuent à être développées, avec des bases mathématiques de plus en plus solides, en particulier sous l'impulsion de Kolmogorov, et l'approche fréquentiste demeure très présente et forte (en particulier en traitement du signal et des images), bénéficiant des travaux de Neyman, Pearson, Feller [141]. Parallèlement, avec la naissance de l'intelligence artificielle et son importance grandissante, le raisonnement humain et sa modélisation regagnent de l'intérêt et conduisent les chercheurs à revenir à une conception plus subjectiviste des probabilités. Deux écoles voient le jour, celle des probabilités additives, et celle des probabilités non additives.

La première école s'appuie sur des postulats de base pour arriver avec plus de rigueur et moins d'arbitraire aux probabilités, à leurs propriétés, au théorème de Bayes, etc. On

y trouve les travaux de Keynes [193], Kemble [192], Cox [93], Jaynes [185], Jeffreys [186], Tribus [333]. Nous détaillerons l'approche de Cox dans la partie 3.3. Celle de Tribus en est directement inspirée. Jaynes et Kemble se placent dans le domaine de la mécanique statistique et montrent que l'approche subjectiviste y est indispensable. Jeffreys, par un raisonnement du même type que celui de Cox, mais où les nombres sont introduits par convention, déduit les propriétés des probabilités d'un certain nombre de principes. Ces principes en rejettent certains autres considérés comme fondamentaux dans d'autres théories (par exemple les définitions des probabilités en termes d'ensemble infini d'observations possibles, en termes de propriétés du monde, le principe de causalité, etc.). L'essence de sa théorie est qu'aucune des probabilités directes, *a priori* ou *a posteriori*, n'est une fréquence. Même si la probabilité est calculée à partir d'une fréquence, elle n'est pas identique à la fréquence et un degré de confiance raisonnable est nécessaire avant son utilisation. Le but de la théorie de Jeffreys n'est pas de justifier l'induction mais d'en assurer la cohérence mathématique. De manière très proche, un grand nombre de chercheurs s'intéressent au point de vue philosophique des probabilités subjectives par rapport aux probabilités objectives, souvent sans remettre en cause l'additivité (Keynes, Jeffreys, Ramsey, de Finetti, Koopman, Russel, Carnap, Good, Savage, etc.). En particulier, les travaux de Savage et de Finetti montrent que l'approche subjective de la théorie bayésienne est la plus légitime et la plus cohérente [98]. Dans ses travaux, de Finetti adopte une approche résolument subjective (plus que celle de Cox) et raisonne en termes de cohérence de l'opinion des individus, et même en termes de psychologie collective afin d'expliquer la coïncidence des opinions d'individus différents. Cette approche particulièrement intéressante sera décrite succinctement à la fin de la partie 3.3.

La deuxième école remet complètement en cause l'additivité, en s'appuyant en particulier sur les travaux des pionniers tels que Bernoulli et Lambert [164, 304]. Koopman, dans les années 40, introduit la notion de probabilités inférieure et supérieure, définissant ainsi une probabilité subjective par une inégalité et non plus comme une valeur réelle précise, en s'inspirant des travaux de Boole (*Laws of Thought*, 1854), qui annonçaient déjà cette évolution. Il est suivi dans ses recherches par plusieurs autres chercheurs (Good, Dempster, etc.). En particulier, Dempster généralise les règles de Lambert, qui ne peuvent traiter que des arguments portant sur une seule conclusion, au cas où plusieurs hypothèses sont à envisager. Des applications de ces nouvelles théories se trouvent dans le domaine de l'économie, où Shackle par exemple propose des modèles économiques faisant appel à des notions proches de la théorie des possibilités, ou encore dans le domaine de la jurisprudence, avec en particulier les travaux d'Ekelöf (*Rättegang*, 1963) qui propose trois lois de combinaison de témoignages : la combinaison de témoignages successifs et celle de témoignages simultanés concordants suivent les règles de Hooper, alors que la combinaison de témoignages conflictuels est reliée à un cas particulier de la règle de Lambert.

À partir des années 60 apparaissent des théories qui ne sont plus directement reliées aux probabilités. Zadeh invente les ensembles flous en 1965 [358], Shortliffe et Buchanan construisent le système MYCIN à partir de la notion de facteurs de certitude en 1975 [306], Shafer développe la théorie des croyances (*A Mathematical Theory of Evidence*, 1976) [302], et Zadeh introduit la théorie des possibilités en 1978 [361]. Ces nouvelles théories connaissent aujourd'hui un grand développement en intelligence artificielle et en fusion de données.

3.2 Classes de probabilités, objectivistes et subjectivistes

La partie 3.1 fait apparaître plusieurs classes de probabilités, que nous synthétisons ici, en nous inspirant de la classification donnée par Good [164].

1. La définition classique est issue de la théorie des jeux et repose essentiellement sur la notion de cas équiprobables. Le calcul des probabilités se fait alors par dénombrement, en comptant les cas.
2. Une version plus subjective de cette définition introduit une information supplémentaire liée à la connaissance que l'on a, par exemple une connaissance sur l'honnêteté ou non du jeu. Cette deuxième classe ne considère donc que des probabilités conditionnelles. On peut inclure également dans cette classe les probabilités subjectives de Savage ou de Finetti, estimées proportionnellement à la somme d'argent qu'une personne serait disposée à donner si ce qu'elle affirmait se révélait faux [98, 126].
3. Une troisième classe est celle des probabilités inverses, suivant Bayes et Laplace. Il s'agit de la probabilité finale d'une hypothèse (après que des expériences ont été menées) estimée à partir de la probabilité *a priori* (en l'absence d'expériences) et de la probabilité conditionnelle (de vraisemblance), ou probabilité des expériences étant donné l'hypothèse.
4. Les probabilités physiques utilisées au 19ème siècle n'ont plus de caractère subjectif et recherchent au contraire l'objectivité en calculant des probabilités conditionnelles à des expériences réalisées.
5. L'approche purement fréquentiste calcule des fréquences d'occurrence dans des grandes séries (limite de Venn, population infinie de Fisher, etc.).
6. Enfin, la dernière classe, que Good appelle probabilité subjective « néo-classique », est la plus large. Les probabilités représentent des degrés de confiance, relatifs à un état de connaissance, prenant en compte de l'information à la fois objective et subjective. Cette définition englobe toutes les autres et peut être beaucoup plus générale. Elle s'appuie sur une théorie mathématique fondée sur quelques axiomes, permettant d'assurer la cohérence de l'ensemble de degrés de confiance. Elle peut être étendue à une théorie du comportement rationnel en introduisant des utilités. Enfin, Good suggère même de représenter ces probabilités subjectives par des inégalités, ce qui s'approche de la théorie des croyances de Dempster-Shafer par exemple.

L'opposition entre objectivistes et subjectivistes provient en fait de différences fondamentales sur le type de problèmes qu'ils cherchent à résoudre et sur leur modélisation. En effet, les fréquentistes objectivistes recherchent des fréquences dans un ensemble, ce qui suppose la possibilité de répétitions infinies dans des conditions semblables, mais donne également un moyen opérationnel de calcul. La probabilité est caractéristique de l'ensemble et n'existe pas sans lui, mais les données peuvent être hypothétiques (il n'est pas forcément nécessaire de réaliser toutes les répétitions). Cela conduit les objectivistes à refuser des problèmes car dépourvus de sens, comme des événements qui ne se produisent qu'une fois. Des énoncés sont considérés objectifs s'ils peuvent être réfutables (par des contre-exemples), même s'ils ne peuvent pas être prouvés rigoureusement [234]. Au contraire, les subjectivistes considèrent les probabilités comme des mesures de confiance, d'espérance raisonnable [93], de codage numérique d'un état de connaissance [105], d'artifice mental approprié, et peuvent donc traiter des problèmes pour lesquels il n'existe pas d'ensemble, en particulier des phénomènes

uniques. Pour de tels phénomènes, il n'existe pas de probabilité en soi, mais seulement des modèles probabilistes [234]. Les hypothèses sont évaluées en fonction de données observées et de probabilités a priori, même si la connaissance est incomplète. Les subjectivistes ne cherchent pas le meilleur comportement asymptotique comme le font les statisticiens, mais cherchent à faire la meilleure inférence possible avec les données dont ils disposent [192, 105]. Les fréquentistes traitent donc de probabilités aléatoires et les subjectivistes de probabilités épistémologiques [303, 304]. Les premières sont caractéristiques de l'événement lui-même et ne sont pas modifiées quand la connaissance change [192]. Les secondes au contraire sont toujours conditionnelles et changent avec la connaissance. Elles permettent de formuler des conclusions possibles, entre la certitude et l'impossibilité, et constituent donc une logique étendue [193]. Les subjectivistes rejettent le principe selon lequel les mêmes causes ont les mêmes effets, non parce qu'ils le jugent faux, mais parce qu'il est dénué de sens, les causes n'étant jamais identiques. Curieusement, l'objectivité a été introduite pour éliminer l'arbitraire et la subjectivité de Bayes et Laplace mais a nécessité d'utiliser des critères statistiques qui ne sont pas universels et dont le choix ajoute à nouveau de l'arbitraire [105].

Enfin, la dernière différence entre les deux approches, à la fois mathématique et de signification, est fondamentale et concerne l'additivité. Les probabilités aléatoires sont nécessairement additives, puiqu'elles sont liées à l'aspect fréquentiste. Les probabilités épistémologiques ne le sont pas nécessairement, bien que cela soit toujours sujet à controverse. Nous reviendrons sur ce point dans les parties suivantes.

En résumé, trois types de personnes se préoccupent de probabilités : les mathématiciens proposent des modèles sans s'occuper de leur adéquation à une réalité ni de l'utilisation qui va en être faite, les physiciens déduisent des lois d'observations et d'expériences, et les philosophes s'interrogent sur le sens de tout cela.

3.3 Postulats fondamentaux pour une logique inductive

Plutôt que d'accepter les « axiomes » des probabilités tels qu'ils sont présentés par exemple dans l'approche classique de Kolmogorov, les approches plus subjectivistes partent de postulats intuitifs, directement liés à ce que l'on attend d'une logique inductive⁴, dont ils déduisent les règles des probabilités. Cette approche est due essentiellement à Cox [93], et est reprise de manière détaillée par exemple dans [333, 105, 263], où l'on pourra également trouver les démonstrations des principaux résultats. Nous présentons ici ces postulats fondamentaux et les grandes lignes du raisonnement. Nous présentons succinctement à la fin de cette partie les travaux menés par de Finetti [98]. Moins connus des traiteurs des signaux et des images, ils ont cependant un intérêt double, à la fois par leur côté fondamentalement subjectiviste, et par la simplicité de la démonstration.

⁴La logique inductive cherche déterminer la solution la plus vraisemblable étant donnée l'information disponible, le vrai et le faux étant les cas extrêmes, par opposition à la logique déductive pour laquelle les seuls cas possibles sont le vrai, le faux et l'ignorance totale.

3.3.1 Postulats fondamentaux

Les postulats fondamentaux que pose Cox sont les suivants [333] (ceux proposés par Jeffreys [186] en sont très proches⁵) :

1. **Cohérence ou non-contradiction** : si on peut établir une conclusion de plusieurs manières, elles doivent toutes conduire au même résultat ; on ne doit pas avoir de conclusions contradictoires à partir des mêmes données ; de plus, à des propositions qui ont toutes la même valeur de vérité, on doit attribuer des confiances égales.
2. **Continuité de la méthode** : les opérations effectuées doivent être continues, et si un changement faible des données intervient, il ne doit pas entraîner de changement brutal dans le résultat.
3. **Universalité ou complétude** : on doit pouvoir attribuer un degré de confiance à toute proposition bien définie, et les degrés de confiance doivent pouvoir être comparés.
4. **Énoncés sans équivoque** : les propositions doivent être bien définies, c'est-à-dire qu'il doit être théoriquement possible de déterminer si une proposition est vraie ou fausse. Cela correspond à ce que Horvitz appelle la clarté [180].
5. **Pas de refus d'information** : il ne faut pas tirer de conclusion à partir d'informations partielles, c'est-à-dire que toutes les informations, expériences ou connaissances disponibles relatives à la proposition à évaluer doivent être prises en compte, et, en particulier, il est important de tenir compte la dépendance du contexte. Ce postulat répond aux théories classiques des probabilités, où, pour atteindre l'objectivité, certains types d'informations sont écartés.

Les postulats 2 et 3 conduisent à utiliser des nombres réels pour représenter et comparer des degrés de confiance : un seul nombre réel est nécessaire et suffisant pour représenter un degré de confiance, et on passe continûment du vrai au faux.

Le postulat 1 entraîne l'existence de relations fonctionnelles entre degrés de confiance.

Le postulat 4 impose que la logique symbolique classique, déductive, se retrouve comme cas particulier.

Le postulat 5 conduit au conditionnement hypothétique : le degré de confiance dans une proposition A n'est connu que conditionnellement à un état de connaissance e qui représente des informations reliées à la confiance dans A et qui sont supposées (ou crues) vraies. Un tel degré de confiance est noté $[A|e]$.

Le postulat de cohérence et le conditionnement hypothétique imposent alors qu'il existe une équation fonctionnelle T reliant $[AB|e]$ (degré de confiance dans « A et B » pour l'état de connaissance e) et au moins deux des quantités $[A|e]$, $[A|Be]$, $[B|e]$, $[B|Ae]$, et qu'il existe une relation fonctionnelle S entre les degrés de confiance dans une proposition $[A|e]$ et dans sa négation $[\bar{A}|e]$.

⁵ Jeffreys, cherchant définir une méthode générale d'induction, pose les postulats suivants : toutes les hypothèses doivent être exprimées, et les conclusions déduites des hypothèses ; la théorie doit être cohérente et non contradictoire ; toute règle doit pouvoir être applicable en pratique ; la théorie doit fournir des indicateurs d'inférence éventuellement fausse, ainsi que des possibilités de révision si de nouvelles informations sont disponibles ; la théorie ne doit pas refuser a priori d'information empirique. De plus, Jeffreys suggère de s'appuyer sur les guides suivants : le nombre de postulats doit être réduit à un minimum ; la théorie doit être en accord avec le raisonnement humain ; l'induction étant plus complexe que la déduction, on ne peut pas espérer la développer plus loin que la déduction. La démarche de Jeffreys consiste alors à traduire ces postulats en axiomes plus formels, à introduire les nombres pour représenter des probabilités, et enfin à démontrer les résultats classiques [186].

Paris [263], dans une démonstration rigoureuse du résultat de Cox, insiste sur une hypothèse souvent omise mais indispensable pour la démonstration :

$$\forall(\alpha, \beta, \gamma) \in [0, 1]^3, \forall \varepsilon > 0, \exists A, B, C, D, \text{ vérifiant le postulat de cohérence, tq}$$

$$|[D|ABC] - \alpha| < \varepsilon, |[C|AB] - \beta| < \varepsilon, |[B|A] - \gamma| < \varepsilon.$$

En particulier, cette hypothèse ne peut pas être vérifiée sur un référentiel fini.

3.3.2 Première équation fonctionnelle

Pour la relation T , 11 fonctions sont possibles (6 de 2 arguments, 4 de 3 arguments et 1 de 4 arguments). Pour des raisons de symétrie des rôles de A et B , ces fonctions peuvent être réduites à 7 seulement. La suite du raisonnement permet d'éliminer certaines formes de fonctions en examinant des cas particuliers conduisant à des absurdités.

Si l'état de connaissance e stipule que « A et B sont indépendants », alors $[A|e] = [A|Be]$. Parmi les 7 formes de T , celle qui est fonction de $[A|e]$ et $[A|Be]$ n'est alors plus fonction que de $[A|e]$ et ne dépend plus de B . Cette forme doit donc être éliminée.

Si $e = \langle A = \bar{B} \rangle$, alors $[AB|e] = [B|Ae] = i$ (où i représente le degré de confiance affecté aux propositions impossibles). Cette valeur constante élimine la forme de T qui est fonction de $[A|e]$ et $[B|e]$.

Si l'on examine maintenant le cas « A est impossible », alors $[AB|e] = [A|e] = [A|Be] = i$ et $[B|Ae]$ est indéfini. Ce cas permet d'éliminer les 4 formes de T fonctions respectivement de $[A|e]$ et $[B|Ae]$, de $[A|e]$, $[A|Be]$ et $[B|Ae]$, de $[B|e]$, $[B|Ae]$ et $[A|e]$, et de $[A|e]$, $[A|Be]$, $[B|e]$ et $[B|Ae]$.

La seule forme possible est donc :

$$[AB|e] = T([A|Be], [B|e]) = T([B|Ae], [A|e]), \quad (3.1)$$

dans laquelle A et B jouent des rôles interchangeables, et où le postulat de continuité impose que T soit une fonction continue.

La logique déductive classique impose que pour trois propositions A , B et C on ait $(AB)C = A(BC)$. En appliquant cette règle, on en déduit que T doit être associative. La solution générale de cette équation fonctionnelle est un produit :

$$Kf[T([A|Be], [B|e])] = f([A|Be])f([B|e]), \quad (3.2)$$

où K est une constante que l'on peut prendre égale à 1 par commodité, et f est une fonction monotone. La démonstration originale de ce résultat [93] suppose que T est deux fois différentiable. Cependant les résultats d'Aczél sur les équations fonctionnelles permettent de réduire ces hypothèses [2, 3] : il suffit que T soit associative, continue, et strictement croissante par rapport à chacun des arguments ; la différentiabilité n'est pas nécessaire⁶.

Si l'on suppose maintenant « $A = B$ » alors $[A|Be] = c$ (où c est le degré de confiance attribué à une proposition certaine). On en déduit que $f(c) = 1$. De manière analogue, en supposant « $A = \bar{B}$ », on trouve que $f(i)$ doit être égal à 0 ou à $+\infty$. Par convention, on choisit $f(i) = 0$. La fonction f est donc une fonction positive et croissante de 0 à 1.

⁶D. Dubois et H. Prade ont montré que si l'on accepte que T soit seulement croissante au sens large, on peut choisir $T = \min$.

3.3.3 Deuxième équation fonctionnelle

Examinons maintenant la relation fonctionnelle S entre $[A|e]$ et $[\bar{A}|e]$. En appliquant deux fois S , on obtient $S^2 = Id$. La cohérence avec la première équation fonctionnelle entraîne que S doit vérifier l'équation :

$$yS\left[\frac{S(x)}{y}\right] = xS\left[\frac{S(y)}{x}\right] \quad (3.3)$$

dont la solution générale est :

$$f([A|e])^k + f([\bar{A}|e])^k = 1 \quad (3.4)$$

en supposant S deux fois différentiable.

3.3.4 Probabilités déduites des équations fonctionnelles

On pose alors par convention $p(A|e) = f([A|e])^k$, appelée probabilité de A conditionnellement à e . Les deux équations fonctionnelles deviennent alors :

$$p(AB|e) = p(A|Be)p(B|e), \quad (3.5)$$

$$p(A|e) + p(\bar{A}|e) = 1. \quad (3.6)$$

On a ainsi démontré les relations imposées axiomatiquement dans l'approche classique (Kolmogorov). De plus, on manipule d'emblée des probabilités conditionnelles (relatives à un état de connaissance), alors que celles-ci ne sont introduites qu'après dans la théorie classique. Enfin, on en déduit la relation donnant la probabilité de la réunion :

$$p(A + B|e) = p(A|e) + p(B|e) - p(AB|e), \quad (3.7)$$

et donc l'additivité des probabilités d'événements exclusifs (également imposée axiomatiquement dans la théorie classique). On en déduit aussi la règle de Bayes :

$$p(A|Be) = \frac{p(B|Ae)p(A|e)}{p(B|e)}. \quad (3.8)$$

Il faut noter que cette approche conduit à des probabilités subjectives qui sont additives, contrairement à la pensée du 17ème siècle et à celle d'une des écoles du 20ème siècle.

3.3.5 Mesure d'incertitude et théorie de l'information

Dans une démarche similaire à celle de Cox, Jaynes définit une série de critères pour en déduire une mesure d'incertitude [185]. Dans ses travaux, il cherche à rapprocher mécanique statistique et théorie de l'information. Il exprime ses recherches comme un problème de spécification de probabilités dans le cas où l'on a peu d'informations. Reprenant les deux démarches objectivistes et subjectivistes, il retient la deuxième, qui permet, en représentant un état de connaissances, de formuler des conclusions possibles si l'on n'a pas assez d'information pour avoir des conclusions certaines. Elle est donc plus générale et Jaynes l'adopte pour la mécanique statistique.

Les critères intuitifs que Jaynes demande à une mesure d'incertitude sont les suivants :

1. la mesure doit être positive et continue,
2. elle doit augmenter quand l'incertitude augmente,
3. elle doit être additive si les sources sont indépendantes.

Il en déduit une mesure unique de l'incertitude représentée par une distribution de probabilité discrète qui correspond à ces critères intuitifs :

$$H(p_1, \dots, p_n) = -K \sum_{i=1}^n p_i \log p_i. \quad (3.9)$$

De même que Cox retrouve les relations probabilistes à partir de ses postulats, Jaynes retrouve ainsi à partir de ses critères l'entropie de la mécanique statistique et simultanément celle de Shannon [305].

Le principe du maximum d'entropie peut alors être considéré comme l'analogue du principe de raison insuffisante de Laplace. La différence essentielle est que le principe de Laplace a un caractère arbitraire et peut engendrer des paradoxes (la notion de cas équiprobables change si on change de variable). Au contraire, le principe du maximum d'entropie permet d'effectuer des inférences sur la base d'une information partielle de manière non biaisée. Il peut être choisi pour la bonne raison que l'entropie est déterminée de manière unique comme la valeur qui « s'engage le moins » vis-à-vis de l'information manquante, et non pour la raison négative que l'on n'a pas de raison de penser autrement. Cependant, il peut être criticable car les résultats qu'il permet d'obtenir dépendent de la manière de poser le problème. Cette critique s'applique aussi au principe de raison insuffisante et nous y reviendrons dans la partie 8.1.8.

3.3.6 De Finetti et la théorie du pari

L'approche proposée par de Finetti, antérieure à celle de Cox, repose également sur des axiomes simples et intuitifs dont sont déduites les propriétés des probabilités [98]. On retrouve bien sûr les axiomes de croissance et de comparaison universelle, et surtout un axiome de cohérence sur lequel s'appuie l'essentiel de la démonstration.

De Finetti développe une théorie des paris pour expliquer son raisonnement : la probabilité p attribuée par un individu à un événement E est donnée par les conditions dans lesquelles cet individu serait prêt à parier sur cet événement, c'est-à-dire dans lesquelles il misserait la somme pS pour gagner S si l'événement E se produit. À partir de cette définition, de Finetti montre d'abord que la somme des probabilités d'événements incompatibles doit être égale à 1. Soit $\{E_1, \dots, E_n\}$ une classe complète d'événements incompatibles, p_i leurs probabilités (toujours évaluées par un individu), et S_i les enjeux correspondant à chacun d'eux. Si l'événement E_k est réalisé, le gain G_k est défini par la différence entre l'enjeu correspondant S_k et la somme des mises, soit :

$$G_k = S_k - \sum_{i=1}^n p_i S_i. \quad (3.10)$$

On obtient n équations de ce type, correspondant aux n réalisations possibles. En considérant ces équations comme un système de n équations à n inconnues qui sont les S_i , le déterminant

de ce système vaut :

$$D = \begin{bmatrix} 1 - p_1 & -p_2 & \dots & -p_n \\ -p_1 & 1 - p_2 & \dots & -p_n \\ \dots & \dots & \dots & \dots \\ -p_1 & -p_2 & \dots & 1 - p_n \end{bmatrix} = 1 - (p_1 + p_2 + \dots + p_n). \quad (3.11)$$

Si le déterminant est non nul, le système admet une solution quels que soient les G_k , y compris s'ils sont tous positifs. Cela ne serait pas cohérent avec la notion de pari. On imagine mal un jeu auquel on puisse toujours gagner ou donner la possibilité à un adversaire de gagner sûrement ! La seule solution cohérente est donc celle obtenue lorsque le déterminant est nul, c'est-à-dire lorsque :

$$\sum_{i=1}^n p_i = 1. \quad (3.12)$$

Cette condition est de plus suffisante, puisque l'on a alors :

$$\sum_{i=1}^n p_i G_i = 0, \quad (3.13)$$

et donc les gains ne sont pas tous positifs.

De Finetti interprète le résultat de la manière suivante : chaque évaluation (subjective) des p_i telle que $\sum_{i=1}^n p_i = 1$ est une évaluation admissible, c'est-à-dire correspondant à une opinion cohérente. Le choix d'une évaluation parmi celles qui sont admissibles n'a alors plus rien d'objectif.

On déduit de l'équation 3.12 l'additivité des probabilités d'événements disjoints.

La deuxième étape du raisonnement est dédiée à la détermination des probabilités conditionnelles. Pour cela, on considère trois paris :

1. un pari sur $E_1 E_2$ (E_1 et E_2), d'enjeu S_1 , et de mise $p_1 S_1$,
2. un pari sur E_2 , d'enjeu S_2 , et de mise $p_2 S_2$,
3. un pari sur $E_1|E_2$, d'enjeu S , et de mise pS , le gain pour ce pari étant :
 - $(1 - p)S$ si $E_1|E_2$ est vrai,
 - $(-p)S$ si $E_1|E_2$ est faux,
 - 0 si E_2 n'a pas lieu (on considère que le jeu est nul dans ce cas, et que la mise est remboursée).

Trois réalisations sont alors possibles :

1. si E_1 et E_2 se réalisent, le gain est alors :

$$G_1 = (1 - p_1)S_1 + (1 - p_2)S_2 + (1 - p)S; \quad (3.14)$$

2. si E_2 se réalise mais pas E_1 , le gain est alors :

$$G_2 = -p_1 S_1 + (1 - p_2)S_2 - pS; \quad (3.15)$$

3. si E_2 ne se réalise pas, le gain est alors :

$$G_3 = -p_1 S_1 - p_2 S_2. \quad (3.16)$$

Considérons les équations 3.14, 3.15 et 3.16 comme un système de 3 équations à 3 inconnues, comme précédemment. Le déterminant vaut :

$$p_1 - pp_2 \quad (3.17)$$

et, toujours pour des raisons de cohérence, doit être égal à 0. On obtient alors la relation :

$$p(E_1|E_2) = \frac{p(E_1E_2)}{p(E_2)} \quad (3.18)$$

dont on déduit le théorème de Bayes.

On remarque que dans ce cas, l'espérance de gain vaut :

$$p_1G_1 + (p_2 - p_1)G_2 + (1 - p_3)G_3 = (p_1 - p_2p)S = 0. \quad (3.19)$$

Donc l'espérance de gain est nulle pour tous S, S_1, S_2 .

De Finetti adopte une philosophie subjectiviste dans le sens où il considère que les éléments subjectivistes, loin de devoir être éliminés comme le suggèrent les objectivistes afin de rendre la notion de probabilité plus « scientifique », sont essentiels, et inhérents à la notion de probabilité. Cela rejoint le point de vue selon lequel la probabilité exprime l'opinion d'un individu, et n'a de signification que par rapport à cet individu, qui s'oppose au point de vue objectiviste qui considère que la probabilité existe en dehors des individus et est une propriété du monde physique.

Chapitre 4

Bases de la théorie des ensembles flous et des possibilités

Dans ce chapitre, nous donnons les bases de la théorie des ensembles flous et des possibilités. Leur utilisation plus spécifique en traitement et interprétation d'images sera vue dans le chapitre suivant. Cette théorie a été introduite par L. Zadeh et le premier article sur le sujet date de 1965 [358]. Les références [189, 120, 365] contiennent l'essentiel de la théorie.

Nous introduisons d'abord la définition des ensembles flous et les opérations ensemblistes initialement proposées par Zadeh (section 4.1). La section 4.2 présente la notion de mesure floue. Une introduction à la théorie des possibilités est proposée dans la section 4.3. Les principaux opérateurs flous sont présentés dans la section 4.4. Ils seront exploités en particulier dans le chapitre 8 en temps qu'opérateurs de fusion d'informations. Les concepts de variable linguistique et de modificateurs sont introduits dans la section 4.5. Nous nous intéresserons ensuite aux relations floues (section 4.6), en particulier aux relations de similarité et d'ordre, qui sont utiles en particulier en reconnaissance des formes. Les principes de la logique floue et du raisonnement approximatif sont décrits dans la section 4.7. Enfin, dans la section 4.8 nous présentons les principes généraux permettant de passer d'opérations classiques sur des ensembles binaires à des opérations sur des ensembles flous. Ces principes seront utilisés de manière intensive dans le chapitre suivant.

4.1 Définition des concepts fondamentaux des ensembles flous

4.1.1 Ensembles flous

Soit \mathcal{S} l'univers, ou espace de référence. C'est un ensemble classique (ou net). On notera par x, y, \dots ses éléments (ou points). En traitement d'images, \mathcal{S} sera typiquement l'espace dans lequel est définie l'image (\mathbb{Z}^n ou \mathbb{R}^n , avec $n = 2, 3, \dots$). Les éléments de \mathcal{S} sont alors les points de l'image (pixels, voxels). L'univers peut également être un ensemble de valeurs prises par des caractéristiques de l'image telles que l'échelle des niveaux de gris. Les éléments sont alors des valeurs (niveaux de gris). L'ensemble \mathcal{S} peut aussi être un ensemble de primitives ou objets extraits des images (segments, régions, objets...) dans une représentation de plus haut niveau du contenu de l'image.

Un sous-ensemble X de \mathcal{S} est défini par sa fonction caractéristique μ_X , telle que :

$$\mu_X(x) = \begin{cases} 1 & \text{si } x \in X \\ 0 & \text{si } x \notin X \end{cases} \quad (4.1)$$

La fonction caractéristique μ_X est une fonction binaire, spécifiant l'appartenance (binaire) de chaque point de \mathcal{S} à X .

La théorie des ensembles flous traite de l'appartenance graduelle. Un sous-ensemble flou de \mathcal{S} est défini par sa fonction d'appartenance μ de \mathcal{S} dans $[0, 1]^1$. Pour tout x de \mathcal{S} , $\mu(x)$ est la valeur dans $[0, 1]$ représentant le degré d'appartenance de x au sous-ensemble flou (on dit souvent plus simplement « ensemble flou »).

Différentes notations sont souvent utilisées pour désigner un ensemble flou. L'ensemble

$$\{(x, \mu(x)), x \in X\}, \quad (4.2)$$

définit complètement l'ensemble flou, et est parfois noté :

$$\int_{\mathcal{S}} \mu(x)/x \quad (4.3)$$

ou dans le cas discret :

$$\sum_{i=1}^N \mu(x_i)/x_i \quad (4.4)$$

où N désigne le cardinal de \mathcal{S} .

Comme la donnée de l'ensemble de tous les couples $(x, \mu(x))$ est complètement équivalente à la définition de la fonction d'appartenance μ , dans la suite on simplifiera les notations et on utilisera la notation fonctionnelle μ (fonction de \mathcal{S} dans $[0, 1]$) pour désigner à la fois l'ensemble flou et sa fonction d'appartenance.

Le support d'un ensemble flou μ est l'ensemble des points qui ont une appartenance strictement positive à μ (c'est un ensemble binaire) :

$$Supp(\mu) = \{x \in \mathcal{S}, \mu(x) > 0\}. \quad (4.5)$$

Le noyau d'un ensemble flou μ est l'ensemble des points qui appartiennent complètement à μ (c'est également un ensemble binaire) :

$$Core(\mu) = \{x \in \mathcal{S}, \mu(x) = 1\}. \quad (4.6)$$

Un ensemble flou μ est dit normalisé si au moins un point appartient complètement à l'ensemble μ (ce qui est équivalent à $Core(\mu) \neq \emptyset$) :

$$\exists x \in \mathcal{S}, \mu(x) = 1. \quad (4.7)$$

Un ensemble flou μ est dit unimodal s'il existe un unique point x tel que $\mu(x) = 1$. Une définition moins contraignante permet au noyau d'être un ensemble compact, pas forcément réduit à un point.

¹L'intervalle $[0, 1]$ est le plus utilisé, mais n'importe quel intervalle ou autre ensemble (un treillis typiquement) pourrait être utilisé.

4.1.2 Opérations ensemblistes : définitions originales de L. Zadeh

Puisque les ensembles flous ont été introduits par L. Zadeh dans [358] pour généraliser la notion d'ensembles, les premières opérations qui ont été définies ont été des opérations ensemblistes. Dans cette partie nous présentons les définitions originales de L. Zadeh. Des classes plus générales d'opérations seront introduites dans la section 4.4.

L'égalité de deux ensembles flous est définie par l'égalité de leurs fonctions d'appartenance :

$$\mu = \nu \Leftrightarrow \forall x \in \mathcal{S}, \mu(x) = \nu(x). \quad (4.8)$$

L'inclusion d'un ensemble flou dans un autre est définie par une inégalité entre les fonctions d'appartenance :

$$\mu \subseteq \nu \Leftrightarrow \forall x \in \mathcal{S}, \mu(x) \leq \nu(x). \quad (4.9)$$

L'égalité de μ et ν est bien sûr équivalente à l'inclusion dans les deux sens.

Notons que ces notions donnent un résultat binaire. Nous verrons dans la suite comment définir un degré d'inclusion entre deux ensembles flous.

L'intersection (respectivement la réunion) de deux ensembles flous est définie par le minimum (respectivement le maximum) point à point entre les fonctions d'appartenance :

$$\forall x \in \mathcal{S}, (\mu \cap \nu)(x) = \min[\mu(x), \nu(x)], \quad (4.10)$$

$$\forall x \in \mathcal{S}, (\mu \cup \nu)(x) = \max[\mu(x), \nu(x)]. \quad (4.11)$$

Le complémentaire d'un ensemble flou est défini par :

$$\forall x \in \mathcal{S}, \mu^C(x) = 1 - \mu(x). \quad (4.12)$$

Les propriétés principales de ces opérations sont les suivantes :

- elles sont toutes cohérentes avec les opérations ensemblistes : dans le cas particulier où les fonctions d'appartenance ne prennent que les valeurs 0 et 1 (les ensembles sont alors binaires), ces définitions se réduisent aux définitions binaires classiques (cette propriété est importante et c'est le moins qu'on puisse demander à l'extension floue d'une opération binaire) ;
- $\mu = \nu \Leftrightarrow \mu \subseteq \nu$ et $\nu \subseteq \mu$;
- la complémentation floue est involutive : $(\mu^C)^C = \mu$;
- l'intersection et la réunion sont commutatives et associatives ;
- l'intersection et la réunion sont idempotentes et mutuellement distributives ;
- l'intersection et la réunion sont duales par rapport à la complémentation : $(\mu \cap \nu)^C = \mu^C \cup \nu^C$;
- si l'on considère que l'ensemble vide \emptyset est un ensemble flou de fonction d'appartenance identiquement nulle, alors on a $\mu \cap \emptyset = \emptyset$ et $\mu \cup \emptyset = \mu$, pour tout ensemble flou μ défini sur \mathcal{S} ;
- s'il on considère l'univers comme un ensemble flou de fonction d'appartenance identiquement égale à 1, alors on a $\mu \cap \mathcal{S} = \mu$ et $\mu \cup \mathcal{S} = \mathcal{S}$, pour tout ensemble flou μ défini sur \mathcal{S} .

Ces propriétés sont les mêmes que les opérations binaires correspondantes. Cependant, certaines propriétés vérifiées dans le cas binaire sont perdues dans le cas flou, comme le tiers-exclu ($X \cup X^C = \mathcal{S}$) et la non-contradiction ($X \cap X^C = \emptyset$). En effet, on a en général :

$$\mu \cup \mu^C \neq \mathcal{S}, \quad (4.13)$$

$$\mu \cap \mu^C \neq \emptyset. \quad (4.14)$$

4.1.3 Structure et types d'ensembles flous

Soit \mathcal{C} l'ensemble des ensembles nets de \mathcal{S} , et \mathcal{F} l'ensemble des ensembles flous de \mathcal{S} . L'ensemble \mathcal{C} est un treillis booléen (complémenté et distributif) pour l'intersection et la réunion. Il peut être considéré comme le treillis induit par la structure de $\{0, 1\}$. L'intervalle $[0, 1]$ est un treillis pseudo-complémenté et distributif (dans la terminologie des treillis, la complémentation à 1 est une pseudo-complémentation), qui induit une structure de treillis pseudo-complémenté et distributif sur \mathcal{F} .

Plusieurs types d'ensembles flous peuvent être considérés. Jusqu'à présent, nous nous sommes limités au cas où les valeurs prises par les fonctions d'appartenance sont des nombres. De tels ensembles flous sont appelés de type 1. Mais les valeurs d'appartenance peuvent aussi être des ensembles flous, en faisant appel à la structure de treillis de l'ensemble des ensembles flous. Un ensemble flou de type 2 est un ensemble flou dont les valeurs d'appartenance sont des ensembles flous de type 1. Plus généralement, un ensemble flou de type m est un ensemble flou dont les valeurs d'appartenance sont des ensembles flous de type $(m - 1)$ (pour $m > 1$). De telles extensions des ensembles flous sont particulièrement utiles lorsque la valeur d'appartenance que l'on peut affecter à un élément ne peut pas être définie de manière précise. Les opérations qui sont définies dans la suite de ce chapitre peuvent être généralisées aux ensembles flous de type m en utilisant le principe d'extension qui sera présenté dans la section 4.8.

4.1.4 α -coupes

L' α -coupe d'un ensemble flou μ est l'ensemble binaire défini par :

$$\mu_\alpha = \{x \in \mathcal{S}, \mu(x) \geq \alpha\}. \quad (4.15)$$

Les α -coupes strictes (ou fortes) sont définies par :

$$\mu_{\alpha_S} = \{x \in \mathcal{S}, \mu(x) > \alpha\}. \quad (4.16)$$

Un ensemble flou peut être interprété comme un empilement de ses α -coupes. Il peut être reconstruit à partir de celles-ci selon plusieurs formules, dont les plus usitées sont :

$$\mu(x) = \int_0^1 \mu_\alpha(x) d\alpha, \quad (4.17)$$

$$\mu(x) = \sup_{\alpha \in]0,1]} \min(\alpha, \mu_\alpha(x)), \quad (4.18)$$

$$\mu(x) = \sup_{\alpha \in]0,1]} (\alpha \mu_\alpha(x)). \quad (4.19)$$

La plupart des opérations définies jusqu'à maintenant commutent avec les coupes. Plus précisément, on a les relations suivantes :

$$\forall(\mu, \nu) \in \mathcal{F}^2, \mu = \nu \Leftrightarrow \forall \alpha \in [0, 1], \mu_\alpha = \nu_\alpha, \quad (4.20)$$

$$\forall(\mu, \nu) \in \mathcal{F}^2, \mu \subseteq \nu \Leftrightarrow \forall \alpha \in [0, 1], \mu_\alpha \subseteq \nu_\alpha, \quad (4.21)$$

$$\forall(\mu, \nu) \in \mathcal{F}^2, \forall \alpha \in [0, 1], (\mu \cap \nu)_\alpha = \mu_\alpha \cap \nu_\alpha, \quad (4.22)$$

$$\forall(\mu, \nu) \in \mathcal{F}^2, \forall \alpha \in [0, 1], (\mu \cup \nu)_\alpha = \mu_\alpha \cup \nu_\alpha, \quad (4.23)$$

$$\forall \mu \in \mathcal{F}, \forall \alpha \in [0, 1], (\mu^C)_\alpha = (\mu_{1-\alpha})^C. \quad (4.24)$$

Choisir une α -coupe d'un ensemble flou revient à seuiller la fonction d'appartenance afin de sélectionner les points ayant un degré d'appartenance d'au moins α . Cette opération peut être vue comme une « défuzzification ».

4.1.5 Cardinal

Dans cette section nous nous restreignons aux ensembles flous définis sur un domaine fini, ou qui ont un support fini (cela sera toujours le cas dans les applications en image).

Le cardinal d'un ensemble flou μ est défini par :

$$|\mu| = \sum_{x \in \mathcal{S}} \mu(x), \quad (4.25)$$

ou, si seul le support de μ est fini :

$$|\mu| = \sum_{x \in \text{Supp}(\mu)} \mu(x). \quad (4.26)$$

Cette définition est cohérente avec la notion classique de cardinal d'un ensemble binaire. Dans le cas d'un ensemble flou, chaque point compte pour une quantité égale à son degré d'appartenance. Le cardinal est également appelé puissance de l'ensemble flou (par exemple [219]).

Cette définition peut être étendue au cas où \mathcal{S} n'est pas fini mais mesurable. Soit M une mesure sur \mathcal{S} (telle que $\int_{\mathcal{S}} dM(x) = 1$). Le cardinal de μ est alors défini par :

$$|\mu| = \int_{\mathcal{S}} \mu(x) dM(x). \quad (4.27)$$

4.1.6 Convexité

Dans cette section, \mathcal{S} est l'espace euclidien (de dimension quelconque).

La convexité d'un ensemble flou est définie à partir de ses α -coupes : μ est convexe si et seulement si toutes ses α -coupes sont convexes. Attention, cette définition n'est pas équivalente à la convexité de la fonction d'appartenance au sens de la convexité classique des fonctions². L'expression analytique équivalente de la convexité floue est : μ est convexe si et seulement si :

$$\forall(x, y) \in \mathcal{S}^2, \forall \lambda \in [0, 1], \min(\mu(x), \mu(y)) \leq \mu(\lambda x + (1 - \lambda)y). \quad (4.28)$$

²La fonction d'une fonction f est définie par $\forall(x, y), f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y)$.

4.1.7 Nombre flou

Dans cette section, nous supposons que $\mathcal{S} = \mathbb{R}$.

Une quantité floue est un ensemble flou μ sur \mathbb{R} . Un intervalle flou est une quantité floue convexe (toutes ses α -coupes sont des intervalles). La semi-continuité supérieure de μ est équivalente au fait que les α -coupes soient des intervalles fermés.

Un nombre flou est un intervalle flou semi-continu supérieurement (s.c.s.) de support compact et unimodal. Un exemple de nombre flou représentant « environ 10 » est illustré sur la figure 4.1.

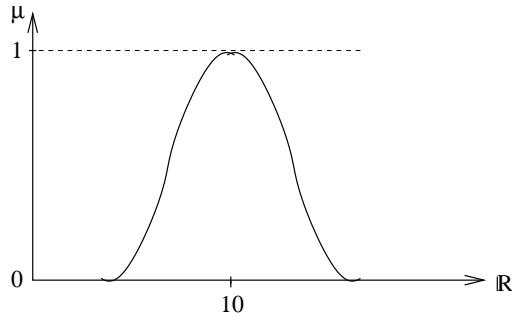


FIG. 4.1 – Nombre flou représentant « environ 10 ».

On peut trouver également des définitions moins strictes, en particulier en acceptant un intervalle de valeurs modales, c'est-à-dire qu'il existe quatre réels a, b, c, d , avec $a \leq b \leq c \leq d$ tels que $\mu(x) = 0$ en dehors de l'intervalle $[a, d]$, μ est croissante sur $[a, b]$, décroissante sur $[c, d]$ et vaut 1 sur $[b, c]$ [161, 162].

Un nombre flou peut être interprété comme une représentation souple d'une quantité imprécise, qui est plus générale qu'un intervalle classique.

Revenons maintenant à la notion de cardinal d'un ensemble flou, défini plus haut comme un nombre. Si l'ensemble est mal défini, on peut s'attendre à ce que toute mesure de cet ensemble soit imprécise également, en particulier son cardinal, qu'on préférera alors définir comme un nombre flou [120] :

$$|\mu|_f(n) = \sup\{\alpha \in [0, 1], |\mu_\alpha| = n\}. \quad (4.29)$$

La quantité $|\mu|_f(n)$ représente le degré avec lequel le cardinal de μ vaut n .

Une classe très courante de nombres flous est constituée des nombres flous *L-R*. Ils sont définis par une représentation paramétrique de leur fonction d'appartenance :

$$\forall x \in \mathbb{R}, \mu(x) = \begin{cases} L\left(\frac{m-x}{\alpha}\right) & \text{si } x \leq m \\ R\left(\frac{x-m}{\beta}\right) & \text{si } x \geq m \end{cases} \quad (4.30)$$

où α et β sont des nombres strictement positifs appelés étendues gauche et droite, m est un nombre appelé valeur moyenne et L et R sont des fonctions ayant les propriétés suivantes :

- $\forall x \in \mathbb{R}, L(x) = L(-x)$,
 - $L(0) = 1$,
 - L est non-croissante sur $[0, +\infty[$,
- et des propriétés similaires pour R .

Un des principaux avantages de tels nombres flous est qu'ils bénéficient d'une représentation compacte permettant des calculs simples.

4.2 Mesures floues

Les définitions et quelques exemples de mesures floues sont présentés ici. On trouvera des présentations plus détaillées dans [327, 120].

4.2.1 Mesure floue d'un ensemble net

Une mesure floue est une fonction f de \mathcal{C} (donc définie sur des ensembles nets) dans $[0, 1]$ satisfaisant les conditions suivantes :

1. $f(\emptyset) = 0$;
2. $f(\mathcal{S}) = 1$;
3. monotonie : $\forall (A, B) \in \mathcal{C}^2, A \subseteq B \Rightarrow f(A) \leq f(B)$;
4. continuité :

$$\begin{aligned} \forall i \in \mathbb{N}, \forall A_i \in \mathcal{C}, A_1 \subseteq A_2 \dots \subseteq A_n \dots \text{ ou } A_1 \supseteq A_2 \dots \supseteq A_n \dots \\ \Rightarrow \lim_{i \rightarrow \infty} f(A_i) = f\left(\lim_{i \rightarrow \infty} A_i\right). \end{aligned}$$

Des propriétés remarquables des mesures floues sont :

$$\forall (A, B) \in \mathcal{C}^2, f(A \cup B) \geq \max[f(A), f(B)], \quad (4.31)$$

$$\forall (A, B) \in \mathcal{C}^2, f(A \cap B) \leq \min[f(A), f(B)]. \quad (4.32)$$

Cette définition ne suppose aucune contrainte d'additivité. On pourrait les appeler plus simplement mesures non-additives, car le lien avec la théorie des ensembles flous présentée précédemment est relativement faible.

4.2.2 Exemples de mesures floues

Plusieurs familles de mesures floues peuvent être trouvées dans la littérature, les plus fréquentes étant :

- les mesures de probabilités ;
- les λ -mesures floues, obtenues en relâchant la contrainte d'additivité des mesures de probabilités sous la forme :

$$\forall (A, B) \in \mathcal{C}^2, A \cap B = \emptyset \Rightarrow f(A \cup B) = f(A) + f(B) + \lambda f(A)f(B) \quad (4.33)$$

avec $\lambda > -1$;

- les fonctions de croyance et de plausibilité de la théorie des fonctions de croyance [302] (voir chapitre 8) ;
- les mesures de possibilités [361], qui seront introduites dans la section 4.3.

Les liens entre plusieurs types de mesures floues peuvent être trouvés dans [17] ou encore [120].

4.2.3 Intégrales floues

Les intégrales floues [327, 167] sont la contre-partie des intégrales au sens de Lebesgue lorsque l'intégration est effectuée relativement à une mesure floue. Deux types d'intégrales floues peuvent être distinguées.

L'intégrale de Sugeno d'une fonction mesurable f , définie de \mathcal{S} dans $[0, 1]$, relativement à une mesure floue μ est définie par :

$$S_\mu(f) = \int f \circ \mu = \sup_{\alpha \in [0,1]} \min[\alpha, \mu(\{x \in \mathcal{S}, f(x) > \alpha\})]. \quad (4.34)$$

Dans le cas fini ($|\mathcal{S}| = N$), cette expression est équivalente à :

$$S_\mu(f) = \int f \circ \mu = \max_{i=1}^N \min[f(x_{p(i)}), \mu(A_i)], \quad (4.35)$$

où p est une permutation sur $\{1, 2...N\}$ telle que :

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

et où $A_i = \{x_{p(1)}, \dots, x_{p(N)}\}$.

L'intégrale de Choquet d'une fonction mesurable f , définie de \mathcal{S} dans \mathbb{R}^+ , relativement à une mesure floue μ est définie par :

$$C_\mu(f) = \int f d\mu = \int_0^{+\infty} \mu(\{x, f(x) > \alpha\}) d\alpha. \quad (4.36)$$

Dans le cas fini, on obtient :

$$C_\mu(f) = \int f d\mu = \sum_{i=1}^N [f(x_{p(i)}) - f(x_{p(i-1)})] \mu(A_i), \quad (4.37)$$

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

Les propriétés de ces intégrales sont détaillées dans [327, 250, 167]. Les principales dans le cas fini sont les suivantes :

- pour la mesure μ_{\min} définie par $\forall A \subset \mathcal{S}, A \neq \mathcal{S}, \mu_{\min}(A) = 0$ et $\mu_{\min}(\mathcal{S}) = 1$, les intégrales $S_{\mu_{\min}}(f)$ et $C_{\mu_{\min}}(f)$ sont égales au minimum des valeurs prises par f ;
- pour la mesure μ_{\max} définie par $\forall A \subseteq \mathcal{S}, A \neq \emptyset, \mu_{\max}(A) = 1$ et $\mu_{\max}(\emptyset) = 0$, les intégrales $S_{\mu_{\max}}(f)$ et $C_{\mu_{\max}}(f)$ sont égales au maximum des valeurs prises par f ;
- quelles que soient les deux fonctions mesurables f et f' et pour toute mesure floue μ , on a la propriété de monotonie suivante :

$$(\forall x \in \mathcal{S}, f(x) \leq f'(x)) \Rightarrow \begin{cases} S_\mu(f) &\leq S_\mu(f') \\ C_\mu(f) &\leq C_\mu(f') \end{cases} \quad (4.38)$$

qui est vraie également dans le cas infini ;

- pour toute fonction mesurable f et toutes mesures floues μ et μ' , on a la propriété de monotonie suivante :

$$(\forall A \subseteq \mathcal{S}, \mu(A) \leq \mu'(A)) \Rightarrow \begin{cases} S_\mu(f) & \leq S_{\mu'}(f) \\ C_\mu(f) & \leq C_{\mu'}(f) \end{cases} \quad (4.39)$$

qui est vraie aussi dans le cas infini ;

- on déduit des propriétés précédentes les inégalités suivantes pour tout fonction mesurable f et toute mesure floue μ :

$$\min_{i=1}^N f(x_i) \leq S_\mu(f) \leq \max_{i=1}^N f(x_i); \quad (4.40)$$

$$\min_{i=1}^N f(x_i) \leq C_\mu(f) \leq \max_{i=1}^N f(x_i); \quad (4.41)$$

- pour toute mesure additive (ou σ -additive dans le cas infini), l'intégrale de Choquet coïncide avec l'intégrale de Lebesgue ; en ce sens, les intégrales floues peuvent être considérées comme une extension des intégrales de Lebesgue ;
- pour toute mesure floue μ , les intégrales de Sugeno et de Choquet satisfont la propriété de continuité suivante : pour toute suite de fonctions mesurables f_n sur \mathcal{S} telle que :

$$\lim_{n \rightarrow +\infty} f_n = f$$

on a :

$$\lim_{n \rightarrow +\infty} S_\mu(f_n) = S_\mu(f), \quad (4.42)$$

$$\lim_{n \rightarrow +\infty} C_\mu(f_n) = C_\mu(f). \quad (4.43)$$

Les intégrales floues sont appliquées en particulier en agrégation multi-critères, mais aussi en fusion de données et en reconnaissance de formes.

4.2.4 Mesures d'ensembles flous

Dans ce qui précède, les mesures ont été appliquées à des ensembles nets. Si l'on considère maintenant des ensembles flous, nous avons besoin de mesures fournissant des évaluations quantitatives de tels ensembles. Ces mesures sont appelées mesures d'ensembles flous [66] ou mesures d'évaluation [130]. Il n'y a pas vraiment de consensus sur la définition de telles mesures. Nous donnons ici la définition la moins contraignante, proposée dans [66].

Une mesure d'ensemble flou est une fonction M de \mathcal{F} dans \mathbb{R}^+ telle que :

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

D'autres contraintes peuvent être ajoutées selon les applications, par exemple :

- M prend ses valeurs dans $[0, 1]$;
- $M(\mathcal{S}) = 1$;
- $M(\mu) = 0 \Leftrightarrow \mu = \emptyset$;
- $M(\mu) = 1 \Leftrightarrow \mu = \mathcal{S}$.

Des exemples simples de mesures d'ensembles flous sont le cardinal flou, le cardinal du support de μ , le supremum de μ , etc. Dans la section suivante d'autres exemples sont donnés : les mesures de flou.

4.2.5 Mesures de flou

Une question importante portant sur l'évaluation d'un ensemble flou concerne le degré de flou de l'ensemble. De Luca et Termini [219] ont proposé de définir un degré de flou comme une fonction f de \mathcal{F} dans \mathbb{R}^+ telle que :

1. $\forall \mu \in \mathcal{F}, f(\mu) = 0 \Leftrightarrow \mu \in \mathcal{C}$ (les ensembles nets sont complètement non flous et sont les seuls à vérifier cette propriété) ;
2. $f(\mu)$ est maximum si et seulement si $\forall x \in \mathcal{S}, \mu(x) = 0.5$;
3. $\forall (\mu, \nu) \in \mathcal{F}^2, f(\mu) \geq f(\nu)$ si ν est plus contrasté que μ (plus proche d'un ensemble binaire), c'est-à-dire :

$$\forall x \in \mathcal{S}, \begin{cases} \nu(x) \geq \mu(x) & \text{si } \mu(x) \geq 0.5 \\ \nu(x) \leq \mu(x) & \text{si } \mu(x) \leq 0.5 \end{cases}$$

4. $\forall \mu \in \mathcal{F}, f(\mu) = f(\mu^C)$ (un ensemble flou et son complémentaire sont aussi flous l'un que l'autre).

De Luca et Termini ont également défini l'entropie d'un ensemble flou [219], comme degré de flou, dans le cas fini :

$$E(\mu) = H(\mu) + H(\mu^C), \quad (4.44)$$

où $H(\mu)$ est définie de manière similaire à l'entropie de Shannon :

$$H(\mu) = -K \sum_{i=1}^N \mu(x_i) \log \mu(x_i). \quad (4.45)$$

Il est facile de voir que E satisfait tous les axiomes du degré de flou. De plus, on a :

$$H(\max(\mu, \nu)) + H(\min(\mu, \nu)) = H(\mu) + H(\nu), \quad (4.46)$$

et

$$E(\max(\mu, \nu)) + E(\min(\mu, \nu)) = E(\mu) + E(\nu). \quad (4.47)$$

De nombreuses autres mesures de flou ont été proposées, avec des propriétés similaires. Mentionnons ici les principales :

- la distance de Hamming à l'ensemble binaire le plus proche, qui n'est autre que la 0.5-coupe [189] :

$$f(\mu) = \sum_{i=1}^N |\mu(x_i) - \mu_{1/2}(x_i)|; \quad (4.48)$$

- la distance de Hamming ou quadratique entre μ et son complémentaire [346] ou de manière plus générale :

$$f(\mu) = \left[\sum_{i=1}^N |\mu(x_i) - \mu^C(x_i)|^p \right]^{1/p} = \left[\sum_{i=1}^N |2\mu(x_i) - 1|^p \right]^{1/p}; \quad (4.49)$$

- la mesure proposée par Kosko [199], qui compare l'intersection de μ et μ^C avec leur réunion selon la formule :

$$\frac{|\min(\mu, \mu^C)|}{|\max(\mu, \mu^C)|}; \quad (4.50)$$

- l'entropie généralisée définie à partir d'une fonction génératrice, soit sous forme additive, soit sous forme multiplicative [25] :
 - la forme additive est donnée par :

$$f(\mu) = \sum_{i=1}^N g[\mu(x_i)] + g[1 - \mu(x_i)] \quad (4.51)$$

où g est une fonction de $[0, 1]$ dans \mathbb{R}^+ telle que :

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

Des exemples de fonctions génératrices sont $g(t) = te^{1-t}$, $g(t) = at - bt^2$ (avec $0 < b < a$), $g(t) = -t \log t$ (cette dernière forme donne l'entropie floue de [219]).

- la forme multiplicative est donnée par :

$$f(\mu) = \sum_{i=1}^N g[\mu(x_i)]g[1 - \mu(x_i)] \quad (4.52)$$

où g est une fonction de $[0, 1]$ dans \mathbb{R}^+ telle que :

$$\forall t \in [0, 1], g'(t) > 0 \text{ et } g''(t) < 0.$$

Des exemples de fonctions génératrices sont $g(t) = te^{1-t}$, $g(t) = t^\alpha$.

4.3 Éléments de théorie des possibilités

La théorie des possibilités, dérivée de celle des ensembles flous, a été introduite par L. Zadeh dans [361], et ensuite été développée par plusieurs chercheurs, en particulier D. Dubois et H. Prade en France [120, 126].

4.3.1 Nécessité et possibilité

Une mesure de possibilité est une fonction Π de \mathcal{C} dans $[0, 1]$ telle que :

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

Dans le cas fini, une mesure de possibilité est une mesure floue. Elle correspond au cas limite de l'équation 4.31, déduite de la monotonie d'une mesure floue.

Par dualité, une mesure de nécessité est définie comme une fonction N de \mathcal{C} dans $[0, 1]$ telle que :

$$\forall A \subseteq \mathcal{S}, N(A) = 1 - \Pi(A^C). \quad (4.53)$$

Cette dualité signifie que si un événement est nécessaire, son contraire est impossible.

Une mesure de nécessité vérifie les propriétés suivantes :

1. $N(\emptyset) = 0$,
2. $N(\mathcal{S}) = 1$,

$$3. \forall I \subset \mathbb{N}, \forall A_i \subseteq \mathcal{S} (i \in I), N(\cap_{i \in I} A_i) = \inf_{i \in I} N(A_i).$$

Inversement, de toute mesure satisfaisant ces propriétés, on peut déduire une mesure de possibilité par dualité.

Les mesures de possibilité et de nécessité ont de plus les propriétés suivantes :

- $\forall A \subseteq \mathcal{S}, \max(\Pi(A), \Pi(A^C)) = 1$, qui exprime le fait que l'un des deux ensembles A et A^C est complètement possible ;
- $\forall A \subseteq \mathcal{S}, \min(N(A), N(A^C)) = 0$, qui exprime que deux événements contraires ne peuvent pas être simultanément nécessaires ;
- $\forall A \subseteq \mathcal{S}, \Pi(A) \geq N(A)$ (un événement doit être possible avant d'être nécessaire) ;
- $\forall A \subseteq \mathcal{S}, N(A) > 0 \Rightarrow \Pi(A) = 1$ (puisque $N(A) > 0 \Rightarrow \Pi(A^C) < 1$, et $\max(\Pi(A), \Pi(A^C)) = 1$) ;
- $\forall A \subseteq \mathcal{S}, \Pi(A) < 1 \Rightarrow N(A) = 0$;
- $\forall A \subseteq \mathcal{S}, N(A) + N(A^C) \leq 1$;
- $\forall A \subseteq \mathcal{S}, \Pi(A) + \Pi(A^C) \geq 1$.

Les deux dernières propriétés expriment la non additivité. La donnée de $\Pi(A)$ n'est pas suffisante pour déterminer complètement $\Pi(A^C)$, contrairement aux mesures de probabilité. L'incertitude attachée à un événement est exprimée par deux nombres, et non plus par un seul.

4.3.2 Distribution de possibilité

Une distribution de possibilité est une fonction π de \mathcal{S} dans $[0, 1]$ avec la condition de normalisation suivante :

$$\sup_{x \in \mathcal{S}} \pi(x) = 1. \quad (4.54)$$

Cette condition correspond à une hypothèse de monde fermé, dans lequel au moins un élément de \mathcal{S} est complètement possible. Cette condition peut être relâchée dans une hypothèse de monde ouvert.

Dans le cas fini, une distribution de possibilité permet de construire une mesure de possibilité par la formule :

$$\forall A \in \mathcal{C}, \Pi(A) = \sup\{\pi(x), x \in A\}. \quad (4.55)$$

Inversement, une mesure de possibilité induit une distribution de possibilité :

$$\forall x \in \mathcal{S}, \pi(x) = \Pi(\{x\}). \quad (4.56)$$

Par dualité, une mesure de nécessité est définie à partir d'une distribution de possibilité par :

$$\forall A \in \mathcal{C}, N(A) = 1 - \sup\{\pi(x), x \notin A\} = \inf\{1 - \pi(x), x \in A^C\}. \quad (4.57)$$

Dans le cas non normalisé, on n'a plus $\Pi(\mathcal{S}) = 1$. De même les propriétés $N(A) > 0 \Rightarrow \Pi(A) = 1$ et $\Pi(A) < 1 \Rightarrow N(A) = 0$ ne sont plus satisfaites.

Ces définitions ont une interprétation simple si on considère le problème de la représentation de la valeur prise par une variable. Alors \mathcal{S} représente le domaine de variation de cette variable. Une distribution de possibilité sur \mathcal{S} décrit les degrés avec lesquels la variable peut prendre chaque valeur possible. C'est en fait l'ensemble flou des valeurs possibles pour cette

variable. Le degré d'appartenance de chaque valeur à cet ensemble correspond au degré de possibilité que la variable prenne cette valeur. Une distribution de possibilité peut donc représenter l'imprécision attachée à la valeur exacte de la variable. Typiquement, un nombre flou est une distribution de possibilité décrivant les valeurs possibles que peut prendre ce nombre.

Considérons par exemple un problème de classification en traitement d'images. Voici quelques exemples (non exhaustifs) de distributions de possibilité.

- Soit \mathcal{S} l'ensemble des classes. Une distribution de possibilité sur \mathcal{S} , définie pour chaque objet à classer (point, région, etc.), peut représenter les degrés avec lesquels cet objet peut appartenir à chacune des classes.
- Soit \mathcal{S} un espace de caractéristiques (par exemple l'échelle des niveaux de gris). Une distribution de possibilité sur \mathcal{S} peut être définie pour chaque classe et représenter, pour chaque niveau de gris, la possibilité pour que cette classe apparaisse dans l'image avec ce niveau de gris.
- Soit \mathcal{S} l'espace de l'image. Une distribution de possibilité sur \mathcal{S} peut être définie pour chaque classe, et donner pour chaque point de l'image son degré de possibilité d'appartenir à cette classe.

Dans la définition présentée ici, nous avons toujours considéré la possibilité et la nécessité d'un sous-ensemble net de \mathcal{S} . Maintenant, considérons un ensemble flou μ de \mathcal{S} ($\mu \in \mathcal{F}$). La notion de possibilité doit alors être étendue [361] :

$$\Pi(\mu) = \sup_{x \in \mathcal{S}} \min(\mu(x), \pi(x)). \quad (4.58)$$

Cela correspond à l'interprétation suivante : étant donné une distribution de possibilité π sur \mathcal{S} , associée à une variable X prenant ses valeurs dans \mathcal{S} , on veut évaluer dans quelle mesure « X est μ ». La possibilité de μ combine donc le degré avec lequel la variable X prend la valeur x et le degré d'appartenance de x à l'ensemble flou.

4.3.3 Sémantique

Les fonctions d'appartenance et distributions de possibilités peuvent avoir différentes sémantiques, dont les principales sont :

- une sémantique de degré de similarité (notion de distance),
- une sémantique de degré de plausibilité qu'un objet dont on ne connaît qu'une description imprécise soit bien celui qu'on cherche à identifier,
- une sémantique de degré de préférence (une classe floue est alors l'ensemble des « bons » choix), cette interprétation étant alors proche de la notion de fonction d'utilité.

Ces trois types de sémantiques sont utilisés en traitement d'images.

4.4 Opérateurs

Après les travaux initiaux de L. Zadeh [358], beaucoup d'opérateurs ont été proposés dans la communauté floue, pour combiner des fonctions d'appartenance ou des distributions de possibilités³. Ces opérateurs sont aussi appelés connecteurs, opérateurs de combinaison

³Il faut noter que comme une distribution de possibilité et une fonction d'appartenance ont des expressions mathématiques similaires et qu'il existe des liens entre les deux, les opérateurs peuvent s'appliquer à l'une ou

ou d'agrégation. Les grandes classes d'opérateurs sont décrites dans [124, 126, 133, 351, 36]. Parmi les principaux opérateurs on trouve en particulier les t-normes, les t-conormes [241, 299], les moyennes [350, 166], les sommes symétriques, et des opérateurs prenant en compte des mesures de conflit ou encore de fiabilité des sources [129, 112]. Dans ce chapitre, nous donnons les principales définitions. Les interprétations en termes d'opérations ensemblistes et de fusion d'informations seront reprises dans les chapitres suivants.

Puisque la plupart des opérateurs agissent point par point (combinent les degrés d'appartenance ou les degrés de possibilités au même point de \mathcal{S}), il est suffisant de les définir sur les valeurs que peuvent prendre les fonctions d'appartenance ou distributions de possibilité. Les opérateurs sont donc définis comme des fonctions sur $[0, 1]$ ou de $[0, 1] \times [0, 1]$ dans $[0, 1]$. Dans ce qui suit, les lettres $x, y, \text{etc.}$ désignent les valeurs à combiner, valeurs dans $[0, 1]$ représentant des degrés d'appartenance ou de possibilité.

4.4.1 Complémentation floue

Une complémentation floue est une fonction c de $[0, 1]$ dans $[0, 1]$ telle que :

1. $c(0) = 1$,
2. $c(1) = 0$,
3. c est involutive : $\forall x \in [0, 1], c(c(x)) = x$,
4. c est strictement décroissante.

L'exemple le plus simple est celui introduit dans la section 4.1 :

$$\forall x \in [0, 1], c(x) = 1 - x. \quad (4.59)$$

Comme il est difficile de construire directement des fonctions involutives, il est intéressant de les caractériser par une forme générale de construction plus simple. Ainsi, les complémentations continues ont la forme générale suivante :

$$\forall x \in [0, 1], c(x) = \varphi^{-1}[1 - \varphi(x)], \quad (4.60)$$

avec $\varphi : [0, 1] \rightarrow [0, 1]$, telle que

- $\varphi(0) = 0$,
- $\varphi(1) = 1$,
- φ est strictement croissante.

Il existe de multiples fonctions φ vérifiant ces propriétés et il est facile d'en exhiber. L'exemple le plus simple est :

$$\forall x \in [0, 1], \varphi(x) = x^n, \quad (4.61)$$

qui permet de construire la complémentation

$$\forall x \in [0, 1], c(x) = (1 - x^n)^{1/n}. \quad (4.62)$$

On retrouve la forme la plus utilisée $c(x) = 1 - x$, obtenue pour $n = 1$. Plus n augmente (pour $n > 1$) ou diminue (pour $n < 1$), plus la forme obtenue devient binaire. Dans le premier cas,

la plupart des valeurs, sauf celles proches de 1, ont un complémentaire proche de 1, et dans à l'autre. Cependant, elles ont des origines, significations et sémantiques différentes, ce qui ne doit pas être négligé.

le second cas, la plupart des valeurs, sauf celles proches de 0, ont un complémentaire proche de 0.

Si, pour un réel a dans $]0, 1]$, φ prend la forme suivante :

$$\forall x \in [0, 1], \varphi(x) = \frac{ax}{(1-a)x + 1}, \quad (4.63)$$

alors la complémentation correspondante est :

$$\forall x \in [0, 1], c(x) = \frac{1-x}{1+a^2x}. \quad (4.64)$$

Donnons un autre exemple, dépendant de quatre paramètres a, b, c et n tels que $0 \leq a < b < c \leq 1$:

$$\forall x \in [0, 1], c(x) = \begin{cases} 1 & \text{si } 0 \leq x \leq a \\ 1 - \frac{1}{2}[\frac{x-a}{b-a}]^n & \text{si } a \leq x \leq b \\ \frac{1}{2}[\frac{c-x}{c-b}]^n & \text{si } b \leq x \leq c \\ 0 & \text{si } c \leq x \leq 1 \end{cases} \quad (4.65)$$

Quelques exemples sont illustrés sur la figure 4.2.

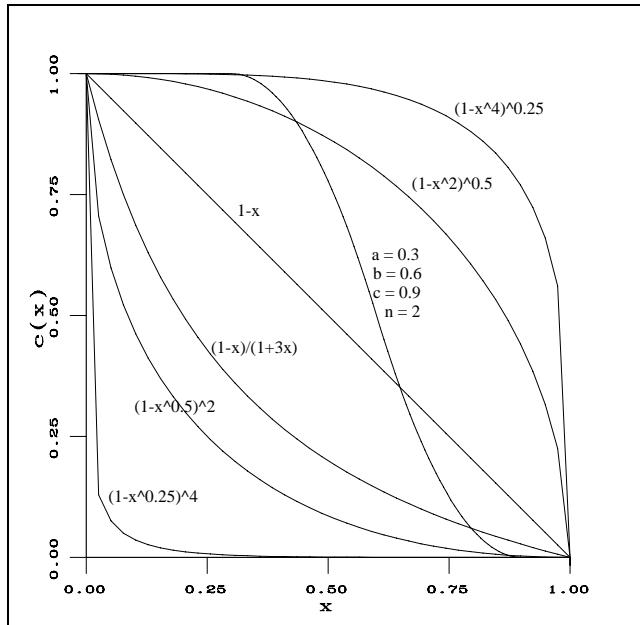


FIG. 4.2 – Quelques exemples de complémentation floue.

4.4.2 Normes et conormes triangulaires

Dans le contexte de la géométrie stochastique ([241, 299]), une norme triangulaire ou t -norme est une fonction $t : [0, 1] \times [0, 1] \rightarrow [0, 1]$ telle que :

1. t est commutative : $\forall (x, y) \in [0, 1]^2, t(x, y) = t(y, x)$;
2. t est associative : $\forall (x, y, z) \in [0, 1]^3, t[t(x, y), z] = t[x, t(y, z)]$;
3. 1 est élément neutre : $\forall x \in [0, 1], t(x, 1) = t(1, x) = x$;

4. t est croissante par rapport aux deux variables :

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

De plus, on a : $t(0, 1) = t(0, 0) = t(1, 0) = 0$, $t(1, 1) = 1$, et 0 est élément absorbant ($\forall x \in [0, 1], t(x, 0) = 0$).

Une propriété de continuité est souvent ajoutée à cet ensemble de propriétés.

Les opérateurs $\min(x, y)$, xy , $\max(0, x + y - 1)$ sont des exemples de t-normes, de loin les plus usitées.

Les t-normes généralisent la notion d'intersection aux ensembles flous ainsi que le « et » logique.

Il est facile de montrer le résultat suivant : pour toute t-norme t , on a :

$$\forall(x, y) \in [0, 1]^2, t(x, y) \leq \min(x, y). \quad (4.66)$$

Cela montre que le « min » est la plus grande t-norme et que tout t-norme a donc un comportement conjonctif.

À l'opposé, toute t-norme est toujours plus grande que t_0 , qui est la plus petite t-norme, définie par :

$$\forall(x, y) \in [0, 1]^2, t_0(x, y) = \begin{cases} x & \text{si } y = 1 \\ y & \text{si } x = 1 \\ 0 & \text{sinon} \end{cases} \quad (4.67)$$

De plus, les t-normes mentionnées plus haut vérifient⁴ :

$$\forall(x, y) \in [0, 1]^2, t_0(x, y) \leq \max(0, x + y - 1) \leq xy \leq \min(x, y). \quad (4.68)$$

Des formes paramétriques permettent des variations entre certains de ces opérateurs. Par exemple, la t-norme définie dans [347] par :

$$\forall(x, y) \in [0, 1]^2, t(x, y) = 1 - \min[1, [(1 - x)^p + (1 - y)^p]^{1/p}] \quad (4.69)$$

varie de la t-norme de Lukasiewicz $\max(0, x + y - 1)$ pour $p = 1$ au min pour $p = +\infty$.

Des exemples de t-normes sont illustrés dans la figure 4.3.

À partir d'une t-norme t et d'une complémentation c , un autre opérateur T appelé t-conorme peut être défini par dualité :

$$\forall(x, y) \in [0, 1]^2, T(x, y) = c[t(c(x), c(y))]. \quad (4.70)$$

Une t-conorme est donc une fonction $T : [0, 1] \times [0, 1] \rightarrow [0, 1]$ telle que :

1. T est commutative : $\forall(x, y) \in [0, 1]^2, T(x, y) = T(y, x)$;
2. T est associative : $\forall(x, y, z) \in [0, 1]^3, T[T(x, y), z] = T[x, T(y, z)]$;
3. 0 est élément neutre : $\forall x \in [0, 1], T(x, 0) = T(0, x) = x$;

⁴Notons cependant qu'il n'y pas d'ordre total sur l'ensemble de toutes les t-normes

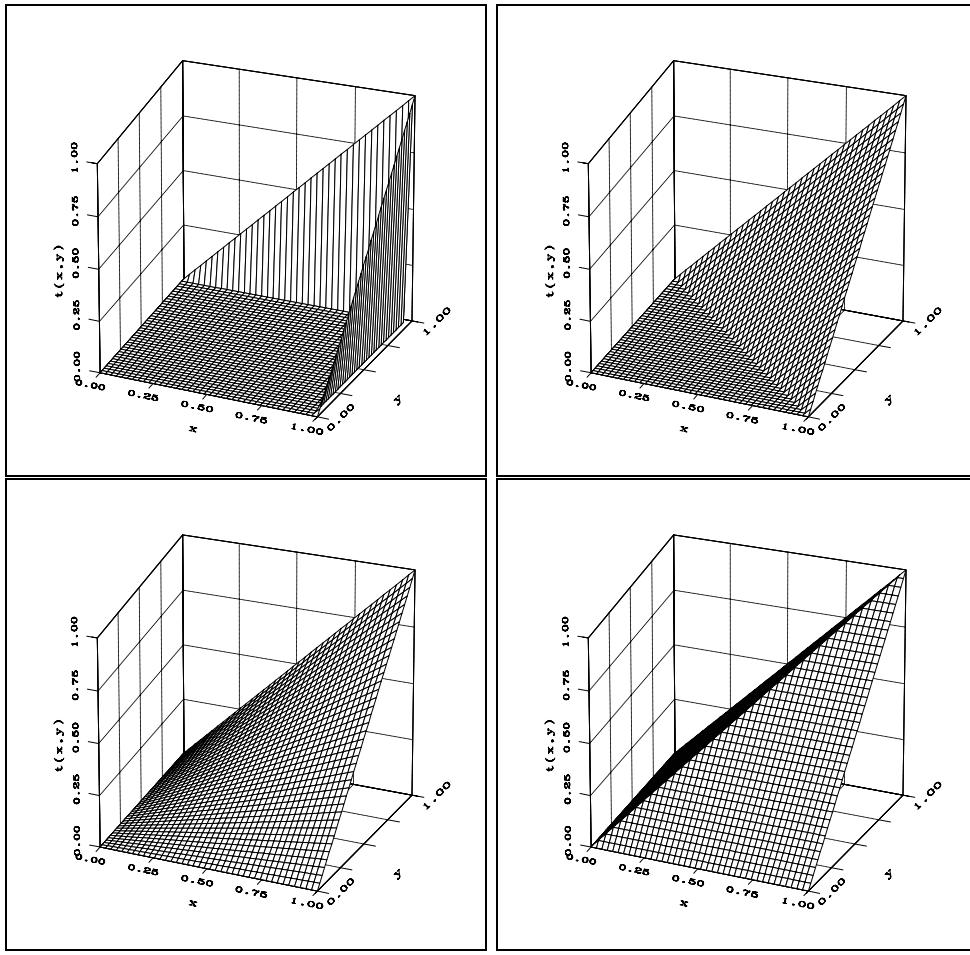


FIG. 4.3 – Quatre exemples de t-normes. Première ligne : t_0 (plus petite t-norme) et t-norme de Lukasiewicz. Seconde ligne : produit et minimum (plus grande t-norme).

4. T est croissante par rapport aux deux variables :

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

De plus, on a : $T(0, 1) = T(1, 1) = T(1, 0) = 1$, $T(0, 0) = 0$, et 1 est élément absorbant ($\forall x \in [0, 1], T(x, 1) = 1$).

Les exemples les plus classiques de t-conormes sont : $\max(x, y)$, $x + y - xy$, $\min(1, x + y)$.

Les t-conormes généralisent aux ensembles flous la notion de réunion ou de « ou » logique.

Pour toute t-conorme, on a ;

$$\forall(x, y) \in [0, 1]^2, T(x, y) \geq \max(x, y). \quad (4.71)$$

Cela montre que le « max » est la plus petite t-conorme et que toute t-conorme a un comportement disjonctif.

À l'opposé, toute t-conorme est plus petite que T_0 , qui est la plus grande t-conorme, définie par :

$$\forall(x, y) \in [0, 1]^2, T_0(x, y) = \begin{cases} x & \text{si } y = 0 \\ y & \text{si } x = 0 \\ 1 & \text{sinon} \end{cases} \quad (4.72)$$

De plus, on a les inégalités suivantes entre les t-conormes les plus classiques⁵ :

$$\forall (x, y) \in [0, 1]^2, T_0(x, y) \geq \min(1, x + y) \geq x + y - xy \geq \max(x, y). \quad (4.73)$$

Des exemples de t-conormes sont montrés sur la figure 4.4.

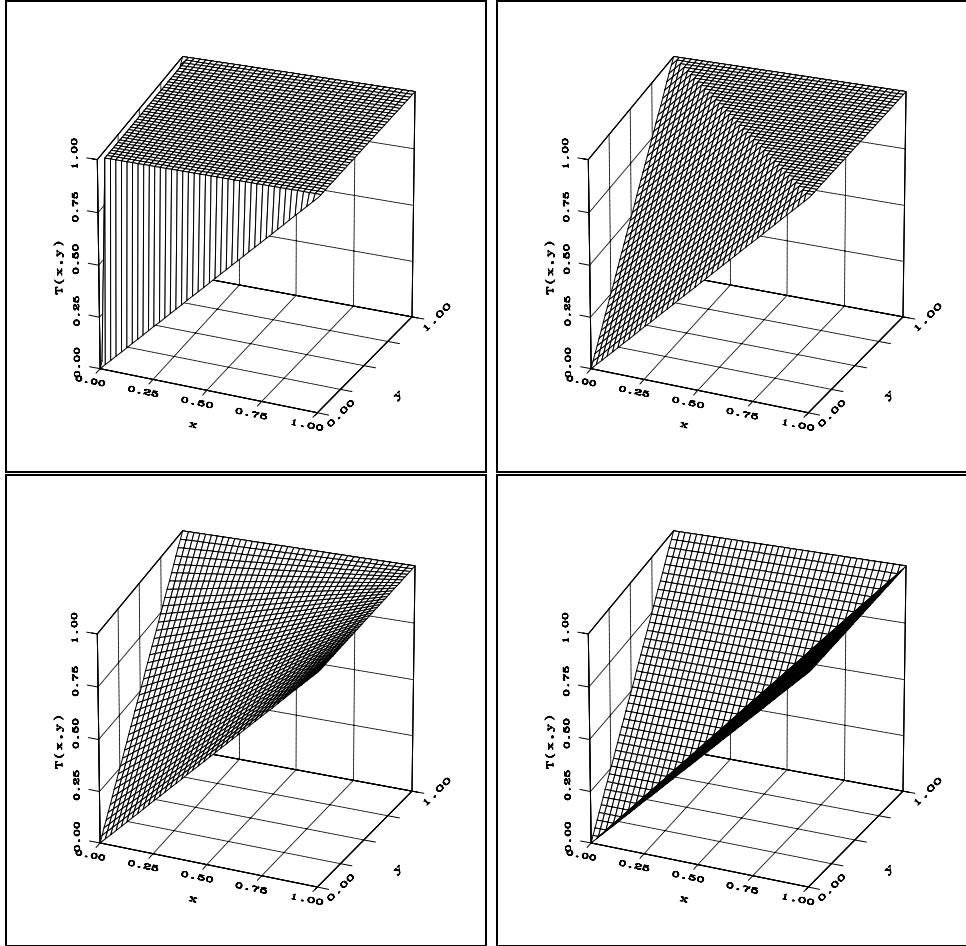


FIG. 4.4 – Quatre exemples de t-conormes. Première ligne : T_0 (plus grande t-conorme) et t-conorme de Lukasiewicz. Seconde ligne : somme algébrique et maximum (plus petite t-conorme).

Voici quelques autres propriétés utiles de ces opérateurs :

- toute t-norme ou t-conorme est distributive par rapport au min et au max, et on a donc des égalités du type :

$$\forall (x, y, z) \in [0, 1]^3, t[x, \min(y, z)] = \min[t(x, y), t(x, z)]; \quad (4.74)$$

- les seules t-normes et t-conormes qui sont mutuellement distributives sont le min et le max ;
- la seule t-norme idempotente est le min, et la seule t-conorme idempotente est le max ;
- à partir de toute t-norme t et de toute fonction continue strictement croissante h de $[0, 1]$ dans $[0, 1]$ telle que $h(0) = 0$ et $h(1) = 1$, on peut construire une autre t-norme t'

⁵Là encore, il n'existe pas d'ordre total sur l'ensemble de toutes les t-conormes.

par la formule [298] :

$$\forall(x, y) \in [0, 1]^2, t'(x, y) = h^{-1}[t(h(x), h(y))]. \quad (4.75)$$

Cela donne un moyen de générer des familles de t-normes à partir d'un exemple.

Il existe des formes génératrices pour des t-normes et des t-conormes ayant des propriétés particulières [124], dont nous détaillons maintenant les deux familles les plus utiles : les t-normes et t-conormes archimédiennes et nilpotentes.

Une t-norme archimédienne strictement monotone t vérifie :

$$\forall x \in [0, 1], t(x, x) < x, \quad (4.76)$$

et

$$\forall(x, y, y') \in [0, 1]^3, y < y' \Rightarrow t(x, y) < t(x, y'). \quad (4.77)$$

De même, une t-conorme archimédienne strictement monotone T vérifie les deux propriétés suivantes :

$$\forall x \in [0, 1], T(x, x) > x, \quad (4.78)$$

$$\forall(x, y, y') \in [0, 1]^3, y < y' \Rightarrow T(x, y) < T(x, y'). \quad (4.79)$$

Tout t-norme archimédienne strictement monotone t peut être exprimée sous la forme suivante :

$$\forall(x, y) \in [0, 1]^2, t(x, y) = f^{-1}[f(x) + f(y)], \quad (4.80)$$

où f , appelée fonction génératrice, est une bijection continue et décroissante de $[0, 1]$ dans $[0, +\infty]$ telle que $f(0) = +\infty$ et $f(1) = 0$.

Les t-conormes associées ont la forme suivante :

$$\forall(x, y) \in [0, 1]^2, T(x, y) = \varphi^{-1}[\varphi(x) + \varphi(y)], \quad (4.81)$$

où la fonction génératrice φ est une bijection continue et croissante de $[0, 1]$ dans $[0, +\infty]$ telle que $\varphi(0) = 0$ et $\varphi(1) = +\infty$.

De telles t-normes et t-conormes ne satisfont jamais les lois de non-contradiction et du tiers-exclu. Ces lois s'expriment par :

$$\forall x \in [0, 1], t[x, c(x)] = 0, \quad (4.82)$$

et

$$\forall x \in [0, 1], T[x, c(x)] = 1. \quad (4.83)$$

Ces deux équations ne sont en général pas vérifiées par les t-normes et t-conormes archimédiennes strictement monotones.

Toute t-norme (ou t-conorme) archimédienne strictement monotone peut également être définie à partir de générateurs multiplicatifs, par équivalence avec les générateurs additifs [81] :

$$\forall(x, y) \in [0, 1]^2, t(x, y) = h^{-1}[h(x)h(y)], \quad (4.84)$$

où h est une fonction strictement croissante de $[0, 1]$ dans $[0, 1]$ telle que $h(0) = 0$ et $h(1) = 1$. L'équivalence avec la forme additive est obtenue simplement en posant :

$$h = e^{-f} \quad (4.85)$$

où f est une fonction génératrice additive.

Les t-norme et t-conorme les plus courantes dans cette classe sont le produit et la somme algébrique :

$$\forall(x, y) \in [0, 1]^2, \quad t(x, y) = xy, \quad T(x, y) = x + y - xy. \quad (4.86)$$

Les seules t-normes rationnelles de cette classe sont les t-normes de Hamacher définies par [172] :

$$\forall(x, y) \in [0, 1]^2, \quad \frac{xy}{\gamma + (1 - \gamma)(x + y - xy)}, \quad (4.87)$$

où γ est un paramètre positif (pour $\gamma = 1$ on retrouve le produit). Elles sont illustrées sur la figure 4.5.

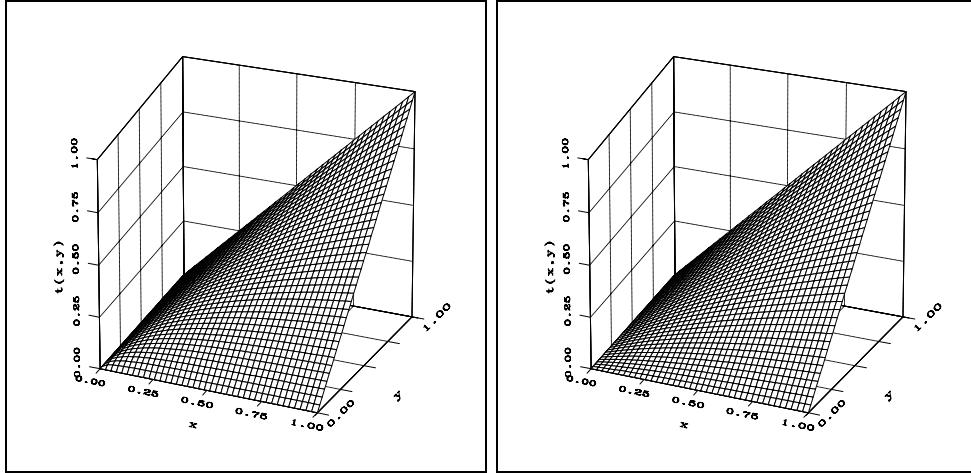


FIG. 4.5 – Deux exemples de t-normes de Hamacher, pour $\gamma = 0$ (à gauche) et $\gamma = 0.4$ (à droite).

Une autre famille paramétrique de celle classe est constituée des fonctions de Frank, définies par [144] :

$$\forall(x, y) \in [0, 1]^2, \quad t(x, y) = \log_s[1 + \frac{(s^x - 1)(s^y - 1)}{s - 1}], \quad (4.88)$$

où s est un paramètre strictement positif. Ces t-normes et leurs t-conormes duales satisfont l'égalité remarquable suivante (et ce sont les seules t-normes et t-conormes qui satisfont cette relation) :

$$\forall(x, y) \in [0, 1]^2, \quad t(x, y) + T(x, y) = x + y. \quad (4.89)$$

Des exemples de t-normes de Frank sont montrés dans la figure 4.6. Si s est petit et tend vers 0, la t-norme tend vers le minimum. Si s tend vers $+\infty$, la t-norme tend vers la t-norme de Lukasiewicz. Si $s = 1$, on retrouve le produit.

La deuxième famille utile de t-normes et t-conormes est constituée des opérateurs nilpotents, qui ont la forme générale suivante :

$$\forall(x, y) \in [0, 1]^2, \quad t(x, y) = f^*[f(x) + f(y)], \quad (4.90)$$

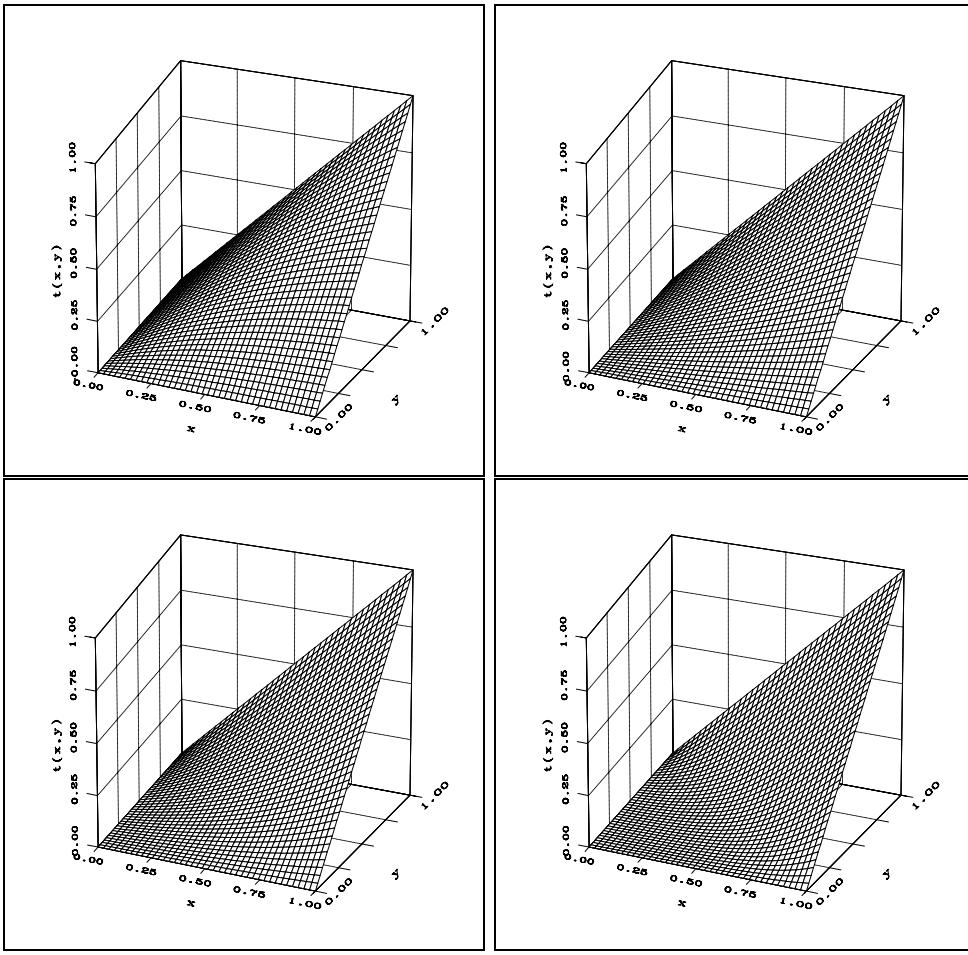


FIG. 4.6 – Quatre exemples de t-normes de Frank. Première ligne : $s = 0.1$ et $s = 2$. Seconde ligne : $s = 10$ et $s = 1000$.

où f est une bijection décroissante de $[0,1]$ dans $[0,1]$, telle que $f(0) = 1$, $f(1) = 0$, et $f^*(x) = f^{-1}(x)$ si $x \in [0,1]$, avec $f^*(x) = 0$ si $x \geq 1$. La forme générale des t-conormes nilpotentes en est déduite par dualité.

Ces opérateurs satisfont les lois du tiers-exclu et de non-contradiction.

Les t-norme et t-conorme les plus courantes dans cette classe sont les opérateurs de Lukasiewicz :

$$\forall (x, y) \in [0, 1]^2, \quad t(x, y) = \max(0, x + y - 1), \quad T(x, y) = \min(1, x + y). \quad (4.91)$$

Des exemples de fonctions génératrices f ont été proposées par Schweizer et Sklar [298] ou encore par Yager [347].

Des opérateurs combinant t-normes et t-conormes ont été proposés, par exemple dans [366] :

$$\forall (x, y) \in [0, 1]^2, \quad C_\gamma(x, y) = t(x, y)^{1-\gamma} T(x, y)^\gamma, \quad (4.92)$$

où γ est un paramètre dans $[0, 1]$.

4.4.3 Opérateurs de moyenne

Un opérateur de moyenne est une fonction $m : [0, 1] \times [0, 1] \rightarrow [0, 1]$ telle que :

1. le résultat de la combinaison est toujours compris entre le min et le max : $\forall(x, y) \in [0, 1]^2$, $\min(x, y) \leq m(x, y) \leq \max(x, y)$, mais $m \neq \min$ et $m \neq \max$;
2. m est commutative ;
3. m est croissante par rapport aux deux variables :

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

La première propriété entraîne que m est toujours une opération idempotente :

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

Ces opérateurs ne sont en général pas associatifs. Les seules qui le soient sont les médianes, où $m(x, y)$ est la valeur médiane de x, y et un paramètre α dans $[0, 1]$:

$$\forall(x, y) \in [0, 1]^2, m(x, y) = med(x, y, \alpha) = \begin{cases} x & \text{si } y \leq x \leq \alpha \text{ ou } \alpha \leq x \leq y \\ y & \text{si } x \leq y \leq \alpha \text{ ou } \alpha \leq y \leq x \\ \alpha & \text{si } y \leq \alpha \leq x \text{ ou } x \leq \alpha \leq y \end{cases} \quad (4.93)$$

La figure 4.7 illustre quelques opérateurs de médiane.

Une propriété plus faible que l'associativité est la bissymétrie :

$$\forall(x, y, z, t) \in [0, 1]^4, m[m(x, y), m(z, t)] = m[m(x, z), m(y, t)]. \quad (4.94)$$

Les moyennes qui vérifient cette propriété, qui sont continues et strictement croissantes ont la forme générale suivante :

$$\forall(x, y) \in [0, 1]^2, m(x, y) = k^{-1}[\frac{k(x) + k(y)}{2}], \quad (4.95)$$

où k est une fonction continue strictement croissante de $[0, 1]$ dans $[0, 1]$. La fonction k peut s'interpréter comme un changement d'échelle ou de dynamique des valeurs à combiner. Celles-ci, une fois transformées par k , sont alors combinées par une simple moyenne arithmétique, puis on ramène le résultat à l'échelle initiale.

Les moyennes les plus classiques sont obtenues pour des fonctions k du type :

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

avec $\alpha \in \mathbb{R}$. La moyenne arithmétique $\frac{x+y}{2}$ est obtenue pour $\alpha = 1$, la moyenne quadratique $\sqrt{\frac{x^2+y^2}{2}}$ pour $\alpha = 2$, la moyenne harmonique $\frac{2xy}{x+y}$ pour $\alpha = -1$, la moyenne géométrique \sqrt{xy} pour $\alpha = 0$. A la limite, quand α tend vers $-\infty$ ou $+\infty$, m tend vers le min ou le max.

La table 4.1 regroupe ces résultats.

La figure 4.8 montre quelques exemples de moyennes.

On trouve également dans la classe des opérateurs de moyenne les moyennes pondérées, les OWA (« ordered weighted average ») [350] et les intégrales floues [166].

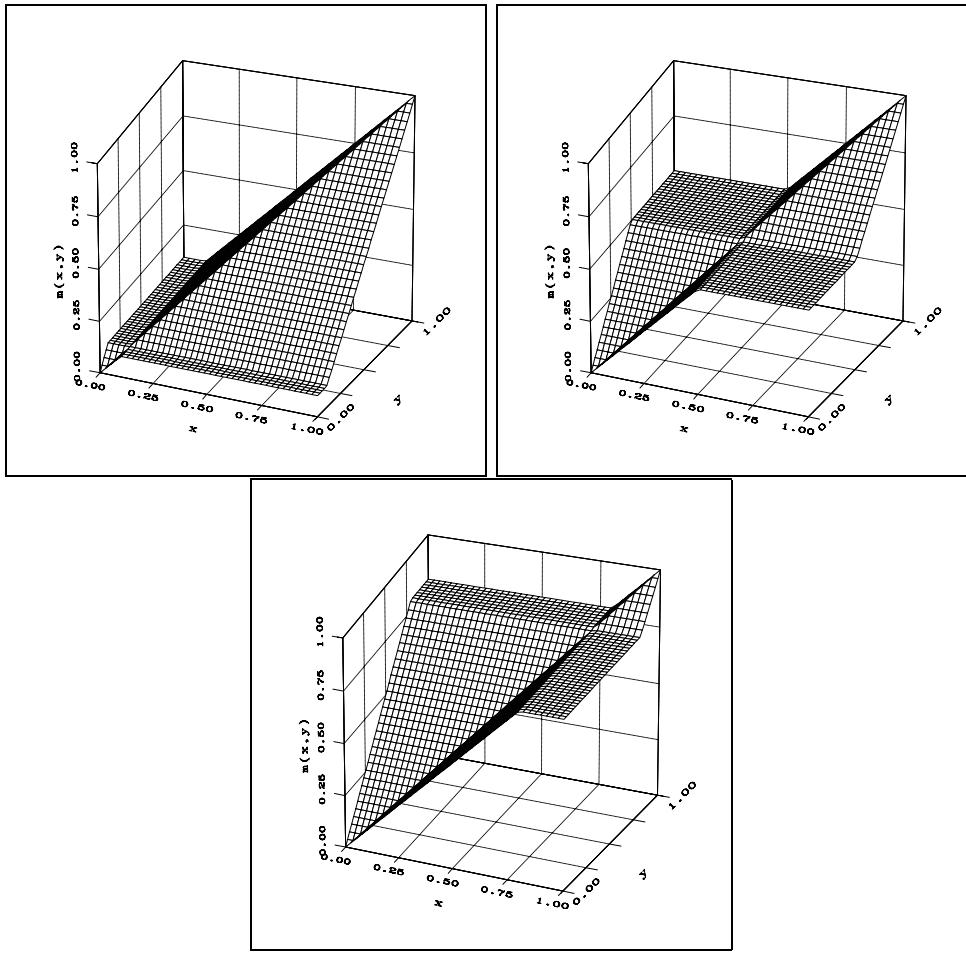


FIG. 4.7 – Trois exemples de médianes, pour α égal à 0.1, 0.5 et 0.8 respectivement.

α	$m(x, y)$	commentaire
$-\infty$	$\min(x, y)$	valeur limite
-1	$\frac{2xy}{x+y}$	moyenne harmonique
0	$(xy)^{-1/2}$	moyenne géométrique
+1	$\frac{x+y}{2}$	moyenne arithmétique
+2	$\sqrt{\frac{x^2+y^2}{2}}$	moyenne quadratique
$+\infty$	$\max(x, y)$	valeur limite

TAB. 4.1 – Exemples de moyennes bisymétriques, continues et strictement croissantes. Pour la moyenne harmonique, on pose par convention $m(0, 0) = 0$.

Dans les OWA, les poids sont définis par le rang des valeurs à combiner. Soient a_1, a_2, \dots, a_n ces valeurs. Elles sont ordonnées en une suite $a_{j_1}, a_{j_2}, \dots, a_{j_n}$ telle que :

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

Alors, pour un ensemble de poids w_i vérifiant :

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

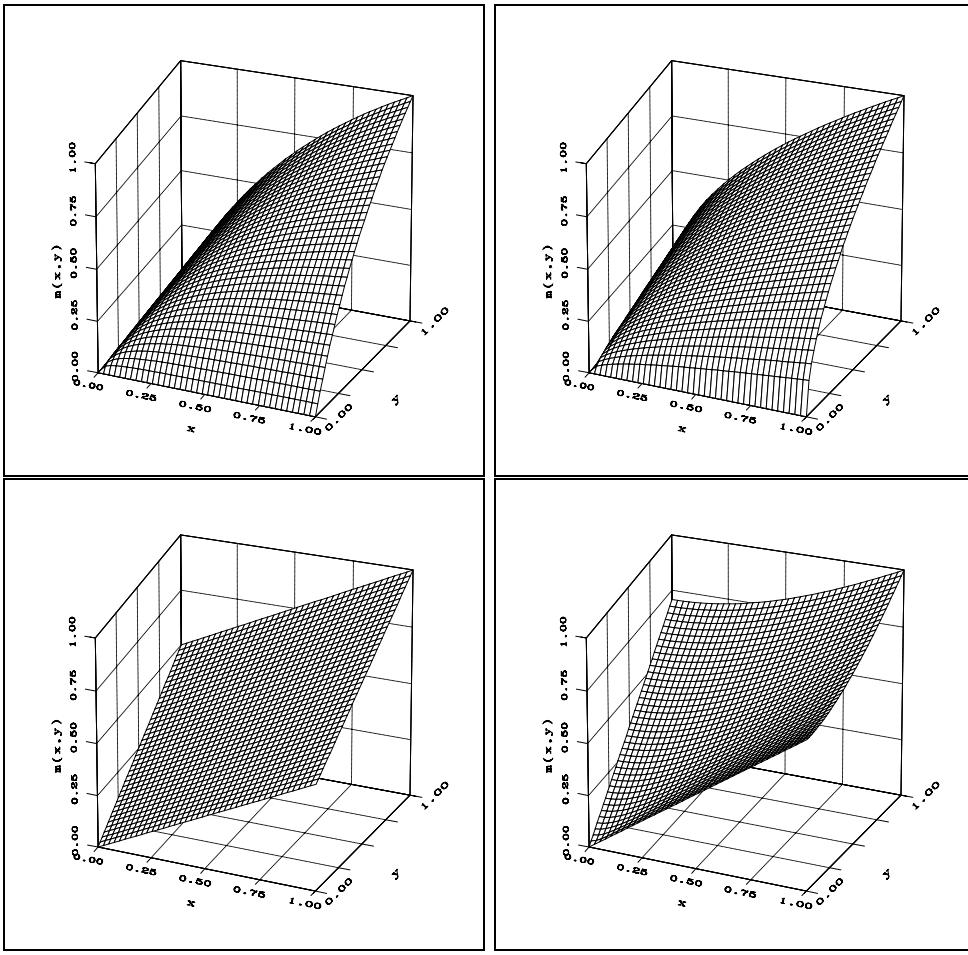


FIG. 4.8 – Quatre exemples d’opérateurs de moyenne. Première ligne : moyennes harmonique et géométrique. Seconde ligne : moyennes arithmétique et quadratique.

l’opérateur OWA est défini par l’expression :

$$OWA(a_1, a_2, \dots, a_n) = \sum_{i=1}^n w_i a_{j_i}. \quad (4.96)$$

On peut également considérer que se trouvent dans cette classe d’opérateurs les intégrales floues [166]. En effet, les intégrales de Choquet et de Sugeno sont idempotentes, continues, croissantes, et comprises entre le minimum et le maximum. On y trouve comme cas particulier les statistiques d’ordre, et donc le minimum, le maximum et la médiane. Les intégrales de Choquet définies par rapport à une mesure additive μ sont équivalentes à une moyenne arithmétique pondérée, dans laquelle les poids w_i affectés aux valeurs x_i sont égaux à $\mu(\{x_i\})$.

Les OWA peuvent aussi être vus comme une classe particulière d’intégrales de Choquet, où la mesure floue est définie par :

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

Inversement, toute intégrale de Choquet commutative est telle que $\mu(A)$ ne dépend que de $|A|$ et est égale à un OWA dont les poids sont donnés par :

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

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

où A_i désigne tout sous-ensemble tel que $|A_i| = i$.

Une étude plus détaillée des propriétés de tels opérateurs peut être trouvée dans [167, 166].

4.4.4 Sommes symétriques

Les sommes symétriques sont définies par une propriété d'auto-dualité, qui correspond à l'invariance du résultat de l'opération par inversion de l'échelle des valeurs à combiner. Plus précisément, une somme symétrique est une fonction $\sigma : [0, 1] \times [0, 1] \rightarrow [0, 1]$ telle que :

1. $\sigma(0, 0) = 0$;
2. σ est commutative;
3. σ est croissante par rapport aux deux variables;
4. σ est continue;
5. σ est auto-duale : $\forall (x, y) \in [0, 1]^2, \sigma(x, y) = 1 - \sigma(1 - x, 1 - y)$.

Notons que l'auto-dualité s'oppose à la dualité mentionnée entre les t-normes et les t-conormes. Pour ces opérateurs, inverser l'échelle entraîne le changement de type d'opérateur. Ici, l'échelle des valeurs peut être inversée sans changer la manière de les combiner. Cette propriété a été utilisée en particulier pour la combinaison de jugements d'experts. Elle pourrait également être exprimée avec d'autres complémentations.

De ces propriétés de base, on déduit que :

- $\sigma(1, 1) = 1$,
- $\forall x \in]0, 1[, \sigma(x, 1 - x) = \frac{1}{2}$,
- la seule somme symétrique qui soit à la fois associative et une moyenne est la médiane de paramètre $\frac{1}{2}$.

La forme générale des sommes symétriques est donnée par :

$$\sigma(x, y) = \frac{g(x, y)}{g(x, y) + g(1 - x, 1 - y)}, \quad (4.97)$$

où g est une fonction de $[0, 1] \times [0, 1]$ dans $[0, 1]$, croissante, positive, continue telle que $g(0, 0) = 0$. Typiquement on peut prendre pour g une t-norme ou une t-conorme continue.

Si $\forall x \in [0, 1], g(0, x) = 0$, $\sigma(0, 1)$ n'est pas défini, sinon $\sigma(0, 1) = \frac{1}{2}$.

La forme générale des sommes symétriques associatives, strictement croissantes est donnée par :

$$\forall (x, y) \in [0, 1]^2, \sigma(x, y) = \psi^{-1}[\psi(x) + \psi(y)], \quad (4.98)$$

où ψ est une fonction strictement monotone telle que $\psi(0)$ et $\psi(1)$ sont non bornés et $\forall x \in [0, 1], \psi(1 - x) + \psi(x) = 0$. On en déduit que 0 et 1 sont éléments neutres, et que $\frac{1}{2}$ est élément absorbant.

Quelques exemples typiques de sommes symétriques sont données dans la table 4.2. Elles sont obtenues en prenant diverses t-normes et t-conormes comme fonction génératrice g .

$g(x, y)$	$\sigma(x, y)$	propriété
xy	$\sigma_0(x, y) = \frac{xy}{1-x-y+2xy}$	associative
$x + y - xy$	$\sigma_+(x, y) = \frac{x+y-xy}{1+x+y-2xy}$	non associative
$\min(x, y)$	$\sigma_{\min}(x, y) = \frac{\min(x, y)}{1- x-y }$	moyenne
$\max(x, y)$	$\sigma_{\max}(x, y) = \frac{\max(x, y)}{1+ x-y }$	moyenne

TAB. 4.2 – Exemples de sommes symétriques, définies à partir de t-normes et de t-conormes. Pour σ_0 , on pose par convention $\sigma_0(0, 1) = \sigma_0(1, 0) = 0$, et pour σ_{\min} , on pose $\sigma_{\min}(0, 1) = \sigma_{\min}(1, 0) = 0$.

Ces opérations sont ordonnées de deux manières suivant les valeurs à combiner :

$$\forall (x, y) \in [0, 1]^2, x + y \leq 1 \Rightarrow \sigma_0(x, y) \leq \sigma_{\min}(x, y) \leq \sigma_{\max}(x, y) \leq \sigma_+(x, y), \quad (4.99)$$

$$\forall (x, y) \in [0, 1]^2, x + y \geq 1 \Rightarrow \sigma_0(x, y) \geq \sigma_{\min}(x, y) \geq \sigma_{\max}(x, y) \geq \sigma_+(x, y). \quad (4.100)$$

Les quatre exemples de la table 4.2 sont illustrés figure 4.9.

4.4.5 Opérateurs adaptatifs

Il existe dans la littérature beaucoup d'autres opérateurs, que nous ne décrivons pas ici. Mentionnons seulement les opérateurs adaptatifs en fonction du conflit entre des distributions de possibilité [129], qui se comportent comme un min si les distributions sont consonantes, et comme un max si elles présentent un fort conflit. Soient π_1 et π_2 deux distributions de possibilité que l'on souhaite combiner en une distribution globale π' . Elles peuvent représenter par exemple l'imprécision sur une variable estimée de deux manières différentes, et on souhaite en déduire une estimation globale. Soit $1 - h(\pi_1, \pi_2)$ une mesure du conflit entre les deux distributions, et t une t-norme. Voici quelques exemples de formules possibles pour π' :

$$\pi'(s) = \max \left[\frac{t[\pi_1(s), \pi_2(s)]}{h(\pi_1, \pi_2)}, 1 - h(\pi_1, \pi_2) \right], \quad (4.101)$$

$$\pi'(s) = \min \left[1, \frac{t[\pi_1(s), \pi_2(s)]}{h(\pi_1, \pi_2)} + 1 - h(\pi_1, \pi_2) \right], \quad (4.102)$$

$$\pi'(s) = t[\pi_1(s), \pi_2(s)] + 1 - h(\pi_1, \pi_2), \quad (4.103)$$

$$\pi'(s) = \max \left[\frac{\min(\pi_1, \pi_2)}{h}, \min[\max(\pi_1, \pi_2), 1 - h] \right]. \quad (4.104)$$

Dans [129], la mesure de conflit proposée est définie par $1 - h$ avec :

$$h = \sup_s t[\pi_1(s), \pi_2(s)]. \quad (4.105)$$

La dernière formule (équation 4.104) est illustrée sur la figure 4.10 pour deux distributions de possibilité avec un conflit croissant.

Nous reviendrons sur ce type d'opérateurs dans le chapitre 8.

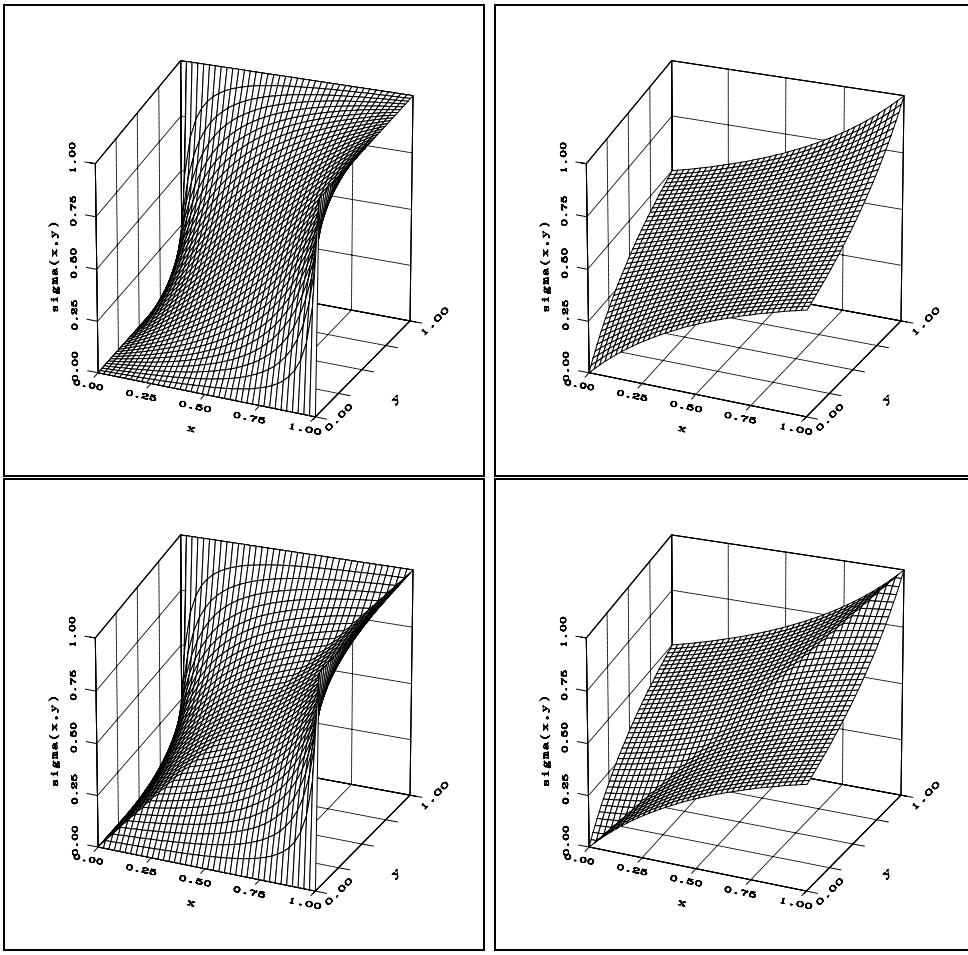


FIG. 4.9 – Quatre exemples de sommes symétriques. Première ligne : σ_0 et σ_+ . Seconde ligne : σ_{\min} et σ_{\max} .

4.5 Variables linguistiques

Il arrive souvent que les représentations numériques ne soient pas adaptées pour décrire une situation. Par exemple, si une variable a un grand domaine de variation, il peut être difficile d'affecter une valeur précise à chaque situation spécifique et on préférera alors utiliser des termes plus qualitatifs issus du langage naturel pour regrouper plus ou moins grossièrement des sous-ensembles typiques de situations intéressantes. Par exemple, pour décrire la taille d'un objet, il peut être plus facile et approprié d'utiliser seulement quelques termes de frontières souples, tels que petit, moyen, grand. Cela correspond à une certaine granularité de l'information. Selon [363], le concept de « granule » est le point de départ des théories de « calcul avec des mots » (*computing with words*). Zadeh définit un granule comme *a fuzzy set of points having the form of a clump of elements drawn together by similarity* [363]. Un mot est alors une étiquette de granule. Pour faire du calcul avec de telles représentations, des outils spécifiques doivent être développés. Le domaine du raisonnement approximatif en bénéficie particulièrement.

Ces types de représentations sont appelés variables linguistiques. Ce sont des variables dont les valeurs sont des mots, des groupes de mots ou des phrases [360]. Leur intérêt est

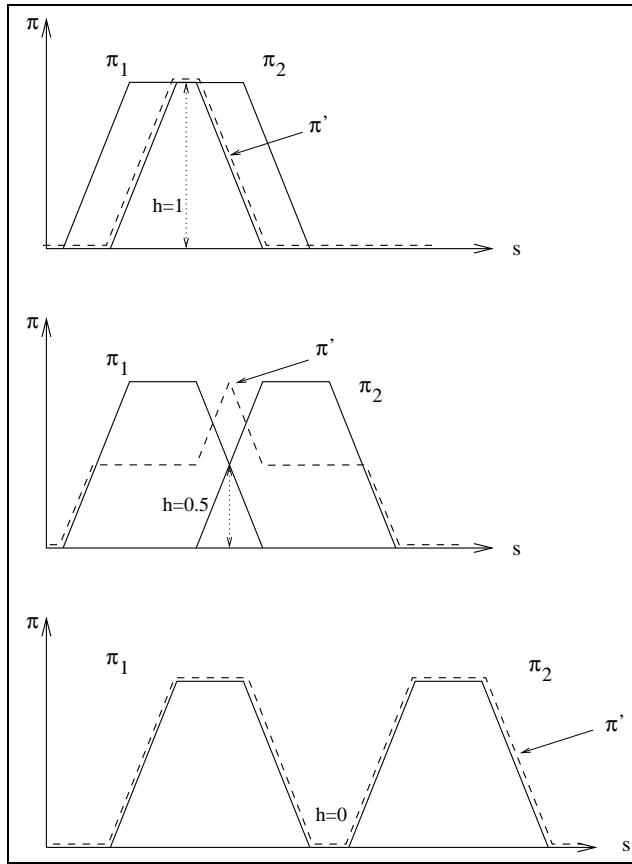


FIG. 4.10 – Exemple d’opérateurs adaptatifs variant du min au max lorsque le conflit entre les deux distributions augmente.

essentiellement que les caractérisations linguistiques peuvent être moins spécifiques que les caractérisations numériques, et nécessitent donc moins d’information pour être utilisées et manipulées dans des systèmes de raisonnement.

4.5.1 Définition

Formellement, une variable linguistique est définie par un quintuplet $(x, T(x), \mathcal{S}, G, M)$ où x est le nom de la variable, $T(x)$ l’ensemble des valeurs de x (appelées termes), \mathcal{S} est le domaine ou univers sur lequel les valeurs de la variable sont définies, G est une règle syntaxique permettant de générer le nom X de chaque valeur de x , et M est une règle sémantique, $M(X)$ étant l’ensemble flou défini sur \mathcal{S} représentant la signification de X [360, 120, 365].

Cette définition représente une conversion symbolique-numérique, et établit des liens entre le langage et les échelles numériques.

4.5.2 Exemple de variable linguistique

Considérons l’exemple de la taille d’un objet. En termes numériques, cette taille peut être exprimée par une valeur variant dans un domaine \mathcal{S} (typiquement \mathcal{S} est un sous-ensemble de

\mathbb{R}^+). En termes linguistiques, la taille peut être exprimée à l'aide de quelques termes tels que très petit, petit, moyen, grand, très grand, etc. La sémantique de ces termes est définie par des ensembles flous sur \mathcal{S} . La figure 4.11 illustre la notion de variable linguistique « taille ».

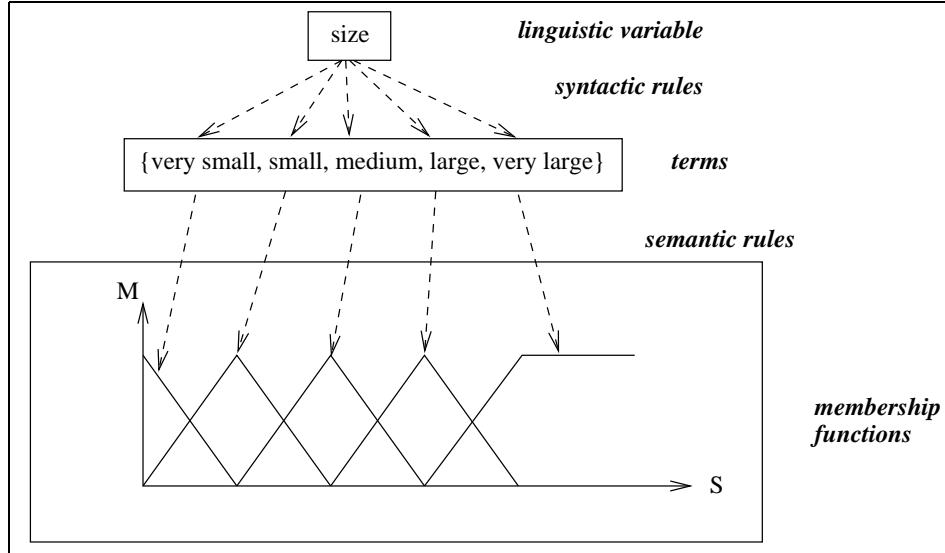


FIG. 4.11 – Illustration de la variable linguistique « taille », de ses termes et des ensembles flous associés. Les flèches allant de la variable linguistique à l'ensemble de termes représentent les règles syntaxiques. Le deuxième ensemble de flèches représente les règles sémantiques et traduisent les termes en fonctions d'appartenance.

4.5.3 Modificateurs

La signification d'un terme d'une variable linguistique peut être modulée par des opérateurs appelés modificateurs. Si A est un ensemble flou, alors le modificateur h permet de construire un terme composé $h(A)$ qui est un ensemble flou sur le même univers \mathcal{S} . Les opérateurs les plus courants sont les suivants :

- normalisation :

$$\mu_{norm(A)}(u) = \frac{\mu_A(u)}{\sup_{v \in \mathcal{S}} \mu_A(v)},$$

où μ_A désigne la fonction d'appartenance à A et u une valeur quelconque dans \mathcal{S} ;

- concentration : $\mu_{con(A)}(u) = [\mu_A(u)]^2$;
- dilatation⁶ : $\mu_{dil(A)}(u) = [\mu_A(u)]^{0.5}$;
- rehaussement de contraste :

$$\mu_{int(A)}(u) = \begin{cases} 2[\mu_A(u)]^2 & \text{si } \mu_A(u) \in [0, 0.5] \\ 1 - 2[1 - \mu_A(u)]^2 & \text{sinon.} \end{cases}$$

Ces fonctions sont illustrées dans la figure 4.12, pour une fonction d'appartenance simple, de forme triangulaire.

Des modificateurs typiques construits à partir de ces opérateurs sont [120] :

- *very A* = $con(A)$,

⁶Attention, ce n'est pas une dilatation au sens morphologique.

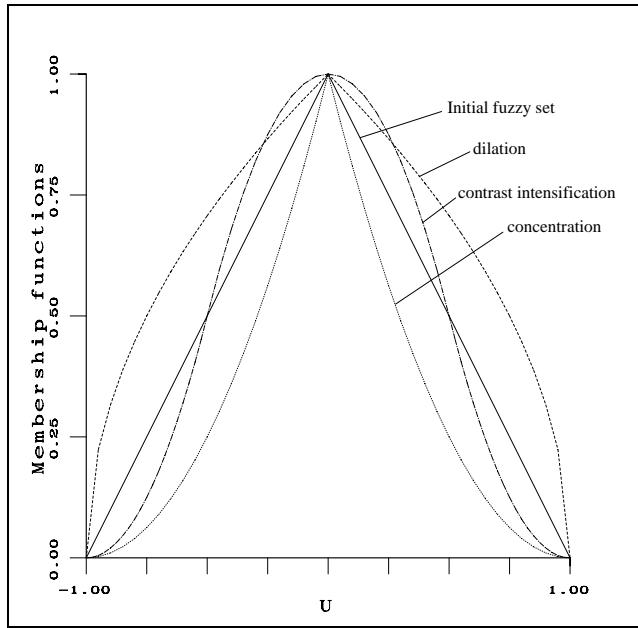


FIG. 4.12 – Illustration de quelques opérations utilisées dans la définition des modificateurs, sur un ensemble flou de forme triangulaire. En pointillés : concentration, en tiretés : dilatation, en pointillés-tiretés : rehaussement de contraste.

- *more or less A = dil(A),*
- *plus A = A^{1.25},*
- *slightly A = int[norm(plus A and not(very A))]* où « and » et « not » sont définis par une t-norme et une complémentation respectivement.

À titre anecdotique, mentionnons une explication du nombre d'or, proposée dans [198] comme application des modificateurs. L'idée développée par l'auteur est que le fait qu'une chose soit considérée comme plaisante dépend de l'optimum d'une variable. Par exemple, si quelqu'un met un peu de parfum, cela pourra être perçu comme plaisant. Mais un excès de parfum ne sentira pas très bon. Dans cet exemple comme dans beaucoup d'autres, il apparaît que le degré x avec lequel quelque chose est plaisant suit la règle suivante : s'il augmente trop, l'effet contraire sera obtenu. L'expression de cette règle en termes linguistiques donne :

$$\text{very}(x) = \text{not}(x)$$

Comme $\text{very}(x)$ est formalisé par x^2 et $\text{not}(x)$ par $1 - x$, on trouve que x doit satisfaire l'équation :

$$x^2 = 1 - x$$

dont la solution est le nombre d'or $\frac{\sqrt{5}-1}{2}$!

4.6 Relations floues

Le concept de relation a été étendu au flou dans [358, 359] pour modéliser les interactions entre des éléments. Cette extension enrichit la théorie des ensembles flous d'outils comparatifs et structurels, ce qui est important pour les applications en interprétation d'images.

4.6.1 Définitions

Soit \mathcal{S}_1 et \mathcal{S}_2 deux univers. Une relation floue R sur $\mathcal{S}_1 \times \mathcal{S}_2$ est un ensemble flou sur $\mathcal{S}_1 \times \mathcal{S}_2$:

$$R = \{((u_1, u_2), \mu_R(u_1, u_2)) | (u_1, u_2) \in \mathcal{S}_1 \times \mathcal{S}_2\}. \quad (4.106)$$

Cette définition de relation binaire s'étend directement à des relations n-aires.

Soient μ_1 et μ_2 deux ensembles flous définis respectivement sur \mathcal{S}_1 et \mathcal{S}_2 . Une relation floue sur $\mu_1 \times \mu_2$ est définie comme un ensemble flou sur $\mathcal{S}_1 \times \mathcal{S}_2$ tel que :

$$\forall (u_1, u_2) \in \mathcal{S}_1 \times \mathcal{S}_2, \mu_R(u_1, u_2) \leq \min[\mu_1(u_1), \mu_2(u_2)]. \quad (4.107)$$

La projection d'une relation floue R sur \mathcal{S}_1 est définie comme un sous-ensemble flou de \mathcal{S}_1 par :

$$\text{proj}[R, \mathcal{S}_1] = \{(u_1, \sup_{u_2 \in \mathcal{S}_2} \mu_R(u_1, u_2)) | u_1 \in \mathcal{S}_1\}. \quad (4.108)$$

La projection sur \mathcal{S}_2 est définie de manière similaire.

Dans le cas de relations n-aires, définies sur $\mathcal{S}_1 \times \dots \times \mathcal{S}_n$, la projection peut être définie sur un espace produit de certains des \mathcal{S}_i , comme une relation floue sur cet espace produit.

L'extension cylindrique d'une relation floue sur un espace produit $\mathcal{S}_{i_1} \times \dots \times \mathcal{S}_{i_k}$, où $\{i_1 \dots i_k\} \subseteq \{1 \dots n\}$, est définie comme la « plus grande » relation floue sur $\mathcal{S}_1 \times \dots \times \mathcal{S}_n$:

$$c(R) = \{((u_1, \dots, u_n), \mu_R(u_{i_1}, \dots, u_{i_k})) | (u_1, \dots, u_n) \in \mathcal{S}_1 \times \dots \times \mathcal{S}_n\}. \quad (4.109)$$

Une relation floue est appelée restriction floue si elle représente une contrainte élastique sur les valeurs (dans $\mathcal{S}_1 \times \mathcal{S}_2$) qui peuvent être affectées à une variable v définie sur $\mathcal{S}_1 \times \mathcal{S}_2$.

4.6.2 Propriétés d'une relation floue

Les propriétés des relations peuvent en général être étendues de plusieurs manières au cas flou. Nous présentons ici les plus courantes, pour une relation binaire sur $\mathcal{S} \times \mathcal{S}$.

Réflexivité

Une relation floue R dans $\mathcal{S} \times \mathcal{S}$ est réflexive si [359] :

$$\forall x \in \mathcal{S}, \mu_R(x, x) = 1. \quad (4.110)$$

Elle est ε -réflexive si [354] :

$$\forall x \in \mathcal{S}, \mu_R(x, x) \geq \varepsilon. \quad (4.111)$$

Elle est faiblement réflexive si [354] :

$$\forall x \in \mathcal{S}, \forall y \in \mathcal{S}, \mu_R(x, x) \geq \mu_R(x, y). \quad (4.112)$$

Symétrie et anti-symétrie

Une relation floue R dans $\mathcal{S} \times \mathcal{S}$ est symétrique si :

$$\forall x \in \mathcal{S}, \forall y \in \mathcal{S}, \mu_R(x, y) = \mu_R(y, x). \quad (4.113)$$

Une relation floue R dans $\mathcal{S} \times \mathcal{S}$ est anti-symétrique [189] si :

$$\forall x \in \mathcal{S}, \forall y \in \mathcal{S}, x \neq y \Rightarrow \begin{cases} \text{soit } \mu_R(x, y) \neq \mu_R(y, x) \\ \text{soit } \mu_R(x, y) = \mu_R(y, x) = 0 \end{cases} \quad (4.114)$$

Une relation floue R dans $\mathcal{S} \times \mathcal{S}$ est parfaitement anti-symétrique [359] si :

$$\forall x \in \mathcal{S}, \forall y \in \mathcal{S}, x \neq y \text{ et } \mu_R(x, y) > 0 \Rightarrow \mu_R(y, x) = 0. \quad (4.115)$$

Les deux définitions de l'anti-symétrie sont liées par l'implication suivante : l'anti-symétrie parfaite implique l'anti-symétrie.

Transitivité

La transitivité exprime que la relation directe entre deux éléments doit être au moins aussi forte que des relations indirectes entre ces éléments (passant par d'autres éléments).

Cette idée se transpose en termes flous de la manière suivante : une relation floue R dans $\mathcal{S} \times \mathcal{S}$ est max-min transitive si :

$$\forall (x, y, z) \in \mathcal{S}^3, \mu_R(x, z) \geq \min[\mu_R(x, y), \mu_R(y, z)]. \quad (4.116)$$

Cette définition peut prendre une forme plus générale, appelée transitivité max-*, où $x * y$ est définie par l'une des expressions suivantes :

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 de relations

La manière la plus courante de composer deux relations R et S définies respectivement sur $\mathcal{S}_1 \times \mathcal{S}_2$ et $\mathcal{S}_2 \times \mathcal{S}_3$ est appelée composition max-min, et est donnée par :

$$\forall u_1 \in \mathcal{S}_1, \forall u_3 \in \mathcal{S}_3, \mu_{R \circ S}(u_1, u_3) = \sup_{u_2 \in \mathcal{S}_2} \min[\mu_R(u_1, u_2), \mu_S(u_2, u_3)]. \quad (4.117)$$

La composition $R \circ S$ est une relation floue sur $\mathcal{S}_1 \times \mathcal{S}_3$.

À la place du min, on peut utiliser un autre opérateur, et définir ainsi une composition max-*. L'opérateur * peut être par exemple le produit ou la moyenne arithmétique [281].

Si l'opérateur $*$ est associatif et croissant par rapport aux deux arguments, alors la composition max-* est associative, distributive sur la réunion et croissante par rapport à l'inclusion.

Comme la composition est obtenue en considérant la « force » de tous les chemins entre u_1 et u_3 , cette notion est reliée à celle de transitivité max-min : une relation floue R sur $\mathcal{S} \times \mathcal{S}$ est max-min transitive si et seulement si $R \circ R \subseteq R$.

On a une équivalence similaire entre la transitivité max-* et la composition max-*.

4.6.4 Relations de similarité

Les relations de similarité constituent une classe particulière de relations floues, particulièrement utiles en reconnaissance des formes, pour la comparaison d'objets, de relations, d'objets et de modèles, pour le raisonnement à partir de cas, pour la reconnaissance à partir de modèles ou d'exemples, etc.

Une relation de similarité est une relation floue S sur \mathcal{S} qui est [359] :

1. réflexive,
2. symétrique,
3. max-min transitive.

La réflexivité peut être trop restrictive, et remplacée par la ε -réflexivité, ou la réflexivité faible. D'autres variantes utilisent la transitivité max-* au lieu de la transitivité max-min.

La transitivité étant souvent difficile à garantir et pas toujours indispensable suivant les applications, les relations plus faibles de tolérance ou de proximité sont intéressantes et ne doivent satisfaire que la réflexivité et la symétrie.

À partir d'une relation de similarité S , on peut déduire une relation de dissimilarité D par :

$$\forall(x, y) \in \mathcal{S}^2, \mu_D(x, y) = 1 - \mu_S(x, y). \quad (4.118)$$

Une relation de dissimilarité satisfait les propriétés suivantes :

1. anti-réflexivité :

$$\forall x \in \mathcal{S}, \mu_D(x, x) = 0,$$

2. symétrie,

3. min-max transitivité :

$$\forall(x, y, z) \in \mathcal{S}^3, \mu_D(x, z) \leq \max[\mu_D(x, y), \mu_D(y, z)].$$

La fonction d'appartenance $\mu_D(x, y)$ représente une fonction de distance sur \mathcal{S} , qui est de plus une ultramétrique, grâce à la dernière propriété.

Zadeh a montré [359] qu'une relation de similarité S peut être représentée par un arbre, appelé arbre de partitions ou arbre de similarité. Chaque niveau de l'arbre correspond à la partition de \mathcal{S} induite par une α -coupe de S . Si Π_α désigne la partition induite par S_α , alors Π_β est un raffinement de Π_α si $\beta \geq \alpha$. Ainsi les partitions obtenues aux différents niveaux de coupe sont emboîtées.

Prenons l'exemple de [359], où $\mathcal{S} = \{x_1, \dots, x_6\}$, et la relation de similarité est définie par la matrice suivante :

$$\mu_S = \begin{pmatrix} 1 & 0.2 & 1 & 0.6 & 0.2 & 0.6 \\ 0.2 & 1 & 0.2 & 0.2 & 0.8 & 0.2 \\ 1 & 0.2 & 1 & 0.6 & 0.2 & 0.6 \\ 0.6 & 0.2 & 0.6 & 1 & 0.2 & 0.8 \\ 0.2 & 0.8 & 0.2 & 0.2 & 1 & 0.2 \\ 0.6 & 0.2 & 0.6 & 0.8 & 0.2 & 1 \end{pmatrix}. \quad (4.119)$$

L'arbre de partitions correspondant à cette relation de similarité est donné Figure 4.13.

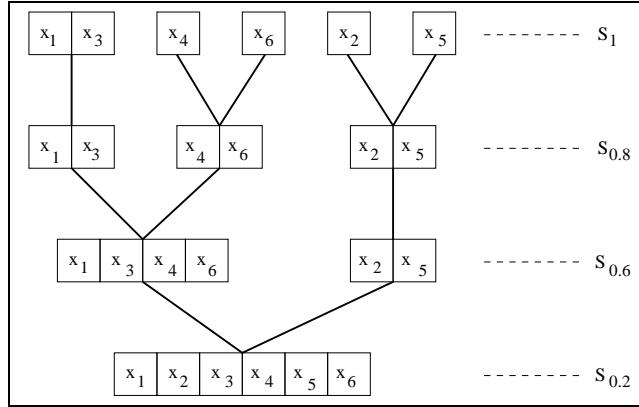


FIG. 4.13 – Arbre de partitions correspondant à la relation de similarité définie par l'équation 4.119 (extrait de [359]). Pour $\alpha = 0.2$, l' α -coupe $S_{0.2}$ donne une seule classe, car tous les x_i sont similaires au degré 0.2. Pour $\alpha = 0.6$, on obtient deux classes, etc.

4.6.5 Relations d'ordre

Contrairement aux relations de similarité, les relations d'ordre ne sont pas symétriques. Plusieurs types d'ordres flous peuvent être définis.

Une relation de pré-ordre floue est une relation floue qui est :

1. réflexive,
2. max-min transitive.

Une relation d'ordre floue est une relation floue qui est :

1. réflexive,
2. anti-symétrique,
3. max-min transitive.

Une relation d'ordre floue parfaite [189] ou relation d'ordre partiel floue [359, 120] est une relation floue qui est :

1. réflexive,
2. parfaitement anti-symétrique,
3. max-min transitive.

Une relation d'ordre total floue [189] ou ordre linéaire [359, 120] est une relation d'ordre partiel floue R telle que :

$$\forall(x, y) \in \mathcal{S}^2, \text{ soit } \mu_R(x, y) > 0 \text{ soit } \mu_R(y, x) > 0. \quad (4.120)$$

Une relation d'ordre partiel floue R sur un univers fini \mathcal{S} peut être représentée de manière équivalente par une matrice triangulaire et par un diagramme de Hasse (graphe orienté dont les arcs sont valués par μ_R). L'antisymétrie et la transitivité de R conduisent à un graphe sans cycle. Illustrons ces représentations avec un exemple issu de [359], pour un univers \mathcal{S} de 6 éléments. La matrice définissant la relation d'ordre partiel floue est donnée par :

$$\mu_R = \begin{pmatrix} 1 & 0.8 & 0.2 & 0.6 & 0.6 & 0.4 \\ 0 & 1 & 0 & 0 & 0.6 & 0 \\ 0 & 0 & 1 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 1 & 0.6 & 0.4 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (4.121)$$

Le diagramme de Hasse correspondant à cet ordre partiel est illustré sur la figure 4.14.

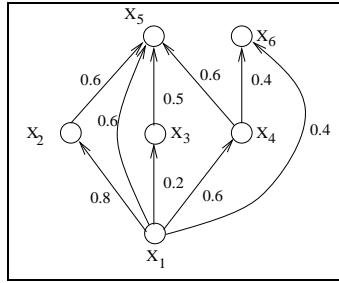


FIG. 4.14 – Diagramme de Hasse correspondant à la relation d'ordre partiel floue définie par l'équation 4.121 (extrait de [359]).

4.7 Logique floue et possibiliste

Le développement de la logique floue est directement lié aux spécificités du raisonnement humain : plus souple que la logique propositionnelle classique, il tolère l'imprécision et permet d'effectuer des déductions même en présence de données et de connaissances imparfaites. Il est capable de manipuler des prédictats graduels, provenant soit de l'utilisation de référentiels continus, soit de notions de typicalité (une situation peut-être plus ou moins typique de situations connues et il faut être capable de prendre cela en compte, dans des raisonnements par analogie).

Dans le cadre du raisonnement sur des propositions, l'incertitude (au sens large) correspond à l'incapacité à dire si une proposition est vraie ou fausse,

- soit parce que l'information est incomplète, vague, imprécise,
- soit parce que l'information est contradictoire ou fluctuante.

Dans le premier cas une modélisation de type possibiliste permet de prendre en compte ce type d'incertitude, alors que dans le deuxième cas, une modélisation de type probabiliste est bien adaptée.

Une autre distinction importante est celle qui existe entre degré de certitude et degré de vérité. La logique floue traite de propositions affectées de degrés de vérité. La logique possibiliste traite plutôt de propositions affectées de degrés d'incertitude.

4.7.1 Logique floue

En logique floue, on raisonne sur des propositions floues élémentaires du type :

$$X \text{ est } P \quad (4.122)$$

où X est une variable prenant ses valeurs dans l'espace de référence \mathcal{S} , et P est un sous-ensemble flou de \mathcal{S} , de fonction d'appartenance μ_P .

Les degrés de vérité de telles propositions sont définis comme des valeurs dans $[0, 1]$ à partir de μ_P .

Les connecteurs logiques sont définis de manière très simple, en utilisant les mêmes opérateurs que leurs équivalents ensemblistes. Par exemple, le degré de vérité d'une conjonction du type :

$$X \text{ est } A \text{ et } Y \text{ est } B$$

est défini à partir d'une t-norme t par :

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

De même, une disjonction telle que :

$$X \text{ est } A \text{ ou } Y \text{ est } B$$

a un degré de vérité défini à partir d'une t-conorme T :

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

et une négation a un degré de vérité défini par une complémentation floue c :

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

Dans le cas de variables à valeurs dans un espace produit, X à valeurs dans \mathcal{S} , et Y à valeurs dans \mathcal{V} , alors la conjonction s'interprète comme un produit cartésien. Le degré de vérité de

$$X \text{ est } A \text{ et } Y \text{ est } B$$

s'écrit alors :

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

Regardons maintenant l'implication. En logique classique, on a :

$$A \Rightarrow B \Leftrightarrow (B \text{ ou non } A), \quad (4.123)$$

et donc l'implication s'exprime à partir d'une disjonction et d'une négation. En utilisant la même équivalence dans le cas flou, on définit une implication floue à partir d'une t-conorme

(disjonction) et d'une complémentation (négation). Soit A et B non flous. Le degré avec lequel A implique B est défini par :

$$Imp(A, B) = T[c(A), B] \quad (4.124)$$

où T est une t-conorme et c une complémentation.

Dans le cas où A et B sont flous, on a :

$$Imp(A, B) = \inf_x T[c(\mu_A(x)), \mu_B(x)]. \quad (4.125)$$

Le tableau suivant résume les principales implications floues utilisées dans la littérature pour le raisonnement approximatif :

$T(x, y) = \max(x, y)$	$\max(1 - a, b)$	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

Dans tous les cas, on retrouve bien la table d'implication classique rappelée ci-dessous dans les cas extrêmes de propositions vraies (1) ou fausses (0), donc prenant des degrés de vérité binaires.

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

Ces quelques définitions permettent maintenant de définir les équivalents flous des principaux modes de raisonnement : modus ponens, modus tollens, syllogisme, contraposition. Prenons l'exemple du modus ponens. En logique classique, il s'écrit :

$$(A \wedge (A \Rightarrow B)) \Rightarrow B. \quad (4.126)$$

Son équivalent flou est défini de la manière suivante :

1. Soit la règle :

$$\text{si } X \text{ est } A \text{ alors } Y \text{ est } B$$

2. et la connaissance :

$$X \text{ est } A'$$

où A' est une approximation de A ,

3. alors on en déduit la conclusion :

$$Y \text{ est } B'$$

où B' est une approximation de B , avec le degré :

$$\mu_{B'}(y) = \sup_x t[\mu_{A \Rightarrow B}(x, y), \mu_{A'}(x)]. \quad (4.127)$$

Nous pouvons maintenant modéliser et manipuler des systèmes de règles floues. Soit par exemple la règle :

SI $(x \text{ est } A \text{ ET } y \text{ est } B) \text{ ALORS } z \text{ est } C$

et α le degré de vérité de x est A , β le degré de vérité de y est B , γ le degré de vérité de z est C . Le degré de vérité (ou de satisfaction) de la règle est obtenu par combinaison des connecteurs flous définis plus hauts :

$$Imp(t(\alpha, \beta), \gamma), \quad (4.128)$$

soit

$$T[c(T(\alpha, \beta)), \gamma]. \quad (4.129)$$

De même pour la règle :

SI $(x$ est A OU y est $B)$ ALORS z est C

son degré de satisfaction est :

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

Ces règles peuvent être utilisées pour décrire de manière qualitative le graphe d'une fonction floue par un petit nombre de règles. Par exemple une fonction du type de celle de la figure 4.15 peut être décrite, à un niveau de granularité relativement grossier, par :

SI X est petit ALORS Y est petit

SI X est moyen ALORS Y est grand

SI X est grand ALORS Y est petit

Ces règles font appel à la notion de variable linguistique vue plus haut, et la sémantique des valeurs « petit », « moyen », « grand » est définie par des ensembles flous sur les domaines de définition de X et Y .

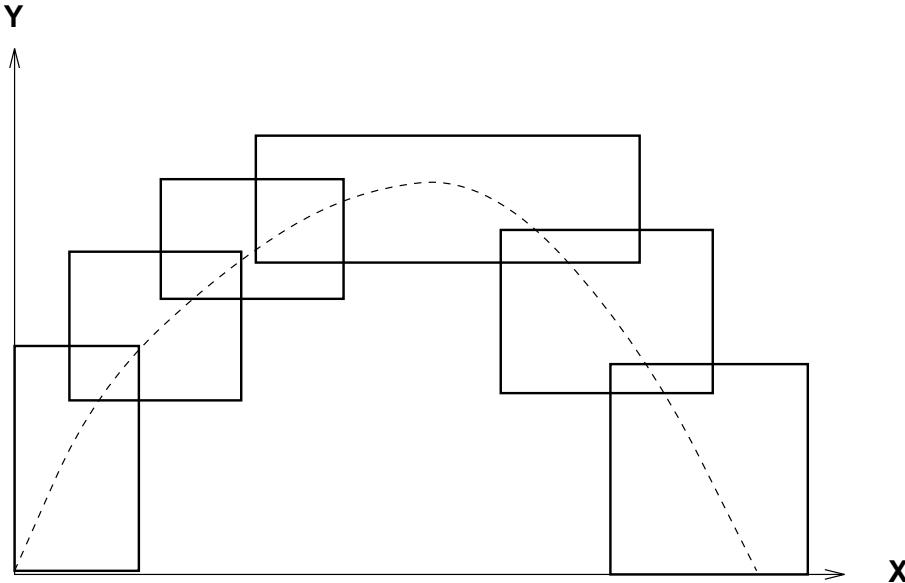


FIG. 4.15 – Exemple du graphe d'une fonction.

Les systèmes de règles floues ont été utilisés dans de nombreux domaines, en contrôle flou essentiellement, mais aussi pour le raisonnement approximatif, pour la modélisation de critères flexibles en traitement d'images, etc.

4.7.2 Logique possibiliste

La logique possibiliste repose sur la définition d'une mesure de possibilité Π sur une algèbre de Boole B de formules :

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

telle que :

- $\Pi(\perp) = 0$,
- $\Pi(\top) = 1$,
- $\forall \varphi, \psi, \Pi(\varphi \vee \psi) = \max(\Pi(\varphi), \Pi(\psi))$,
- $\forall \varphi, \Pi(\exists x \varphi) = \sup\{\Pi(\varphi[a|x]), a \in D(x)\}$ (où $D(x)$ est le domaine de la variable x , et $\varphi[a|x]$ est obtenue en remplaçant les occurrences de x dans φ par a).

Soit maintenant Ω l'ensemble des interprétations et π une distribution de possibilité normalisée :

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

telle que :

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

La possibilité d'une formule s'exprime alors par :

$$\Pi(\varphi) = \sup\{\pi(\omega), \omega \models \varphi\} \quad (4.130)$$

($\omega \models \varphi$ se lit ω est un modèle de φ et signifie que φ est satisfaite dans le monde ω).

De même que sur des ensembles, on définit par dualité une mesure de nécessité sur des formules par :

$$N(\varphi) = 1 - \Pi(\neg\varphi). \quad (4.131)$$

On a alors la propriété suivante :

$$\forall \varphi, \psi, N(\varphi \wedge \psi) = \min(N(\varphi), N(\psi)). \quad (4.132)$$

Ce formalisme permet de traiter de nombreuses situations en les modélisant très simplement. Par exemple, une règle par défaut du type « si A alors B », pouvant avoir des exceptions, s'exprime simplement par :

$$\Pi(A \wedge B) \geq \Pi(A \wedge \neg B) \quad (4.133)$$

De même un raisonnement par modus ponens possibiliste peut être modélisé par :

1. Si on a la règle :

$$N(A \Rightarrow B) = \alpha$$

2. et une connaissance qui s'écrit :

$$N(A) = \beta$$

3. alors la conclusion peut être exprimée par :

$$\min(\alpha, \beta) \leq N(B) \leq \alpha.$$

Le formalisme de la logique possibiliste est utilisé dans de nombreux domaines, par exemple pour représenter des modèles de préférences ou d'utilités sous forme de bases de connaissances ordonnées, puis pour raisonner sur ces bases de connaissances.

Soit une base de connaissance ordonnée notée sous la forme :

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

où α_i est un degré de certitude ou de priorité associé à la formule φ_i (représentant un élément de connaissance).

La satisfaction de cet ensemble de formules en chaque monde est représentée par une distribution de possibilité définie de la manière suivante. Dans le cas où la base de connaissance ne comporte qu'une formule, on a :

$$\pi_{(\varphi, \alpha)}(\omega) = \begin{cases} 1 & \text{si } \omega \models \varphi \\ 1 - \alpha & \text{sinon} \end{cases} \quad (4.134)$$

Plus généralement, pour un ensemble de connaissances avec priorités on a :

$$\pi_{KB}(\omega) = \min_{i=1 \dots n} \{1 - \alpha_i, \omega \models \neg \varphi_i\} = \min_{i=1 \dots n} \max(1 - \alpha_i, \varphi_i(\omega)). \quad (4.135)$$

Cette formule s'interprète de la manière suivante : si une formule est importante (α_i proche de 1), c'est le degré de satisfaction de cette formule dans le monde ω qui est pris en compte. Si au contraire elle n'est pas importante (α_i proche de 0), alors elle n'interviendra pas dans l'évaluation globale de la base de connaissance. Le min correspond au fait que l'on cherche à savoir dans quelle mesure les formules de la base de connaissances sont simultanément satisfaites dans ω .

On peut mesurer le degré d'inconsistance de la base KB par l'expression :

$$1 - \max_{\omega} \pi_{KB}(\omega). \quad (4.136)$$

Une base est dite complète si elle permet de déduire pour toute formule si elle est vraie ou si c'est son contraire qui est vrai : soit $KB \vdash \varphi$ (KB permet de déduire φ), soit $KB \vdash \neg \varphi$ (KB permet de déduire $\neg \varphi$).

Lorsqu'une base n'est pas complète, elle conduit à une ignorance sur certaines formules φ : $KB \not\vdash \varphi$ et $KB \not\vdash \neg \varphi$. La logique possibiliste permet de représenter cette situation de manière très simple :

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

alors qu'il n'existe pas de modèle aussi simple en probabilités par exemple.

4.8 Principes généraux de construction d'opérations floues à partir d'opérations binaires

Dans cette partie, nous présentons les principales méthodes génériques qui permettent de définir une opération ou une relation floue à partir des opérations ou relations binaires correspondantes. Ces méthodes peuvent être catégorisées en trois classes, reposant respectivement sur le principe d'extension introduit dans [360], sur des décompositions des ensembles flous en α -coupes, et sur des traductions formelles de concepts binaires en concepts flous.

4.8.1 Principe d'extension

Définition [360]

Considérons tout d'abord une fonction f de \mathcal{S} dans \mathcal{V} . Soit μ un ensemble flou défini sur \mathcal{S} . L'extension de f à un ensemble flou est un ensemble flou μ' défini sur \mathcal{V} . Il est construit de la manière suivante :

$$\forall y \in \mathcal{V}, \mu'(y) = \begin{cases} 0 & \text{si } f^{-1}(y) = \emptyset, \\ \sup_{x \in \mathcal{S} | y=f(x)} \mu(x) & \text{sinon.} \end{cases} \quad (4.137)$$

Si f est injective, cette équation se réduit à :

$$\forall y \in \mathcal{V}, \mu'(y) = \begin{cases} 0 & \text{si } f^{-1}(y) = \emptyset, \\ \mu[f^{-1}(y)] & \text{sinon.} \end{cases} \quad (4.138)$$

Considérons maintenant le cas plus général d'une fonction f définie sur un espace produit $\mathcal{S}_1 \times \mathcal{S}_2 \times \dots \times \mathcal{S}_n$. Soient μ_1, \dots, μ_n les n ensembles flous définis respectivement sur $\mathcal{S}_1, \dots, \mathcal{S}_n$. L'extension de f aux μ_i donne un ensemble flou sur \mathcal{V} défini par :

$$\forall y \in \mathcal{V}, \mu'(y) = \begin{cases} 0 & \text{si } f^{-1}(y) = \emptyset, \\ \sup_{(x_1, \dots, x_n) \in \mathcal{S}_1 \times \dots \times \mathcal{S}_n | y=f(x_1, \dots, x_n)} \min[\mu_1(x_1), \dots, \mu_n(x_n)] & \text{sinon.} \end{cases} \quad (4.139)$$

Si le supremum est atteint, c'est-à-dire si :

$$\forall y \in \mathcal{V}, \exists (x_1, \dots, x_n) \in \mathcal{S}_1 \times \dots \times \mathcal{S}_n \mid \mu'(y) = \min[\mu_1(x_1), \dots, \mu_n(x_n)], \quad (4.140)$$

alors l'extension floue d'une fonction commute avec les α -coupes, c'est-à-dire :

$$\forall \alpha \in [0, 1], [f(\mu_1, \dots, \mu_n)]_\alpha = f[(\mu_1)_\alpha, \dots, (\mu_n)_\alpha]. \quad (4.141)$$

D'autres principes d'extension peuvent être définis, en utilisant par exemple le produit à la place du minimum [120].

Le principe d'extension est illustré sur la figure 4.16 pour une fonction injective définie sur un espace 1D.

Application à la compatibilité de deux ensembles flous

Un exemple typique de l'application du principe d'extension est la compatibilité entre deux ensembles flous [357]. Considérons un ensemble flou μ sur \mathcal{S} . La valeur $\mu(x)$ peut être interprétée comme un degré de compatibilité de x avec l'ensemble flou μ [120] (μ étant par exemple une valeur floue, ou une variable linguistique). La compatibilité d'un ensemble flou μ' de \mathcal{S} avec μ peut être évaluée par le principe d'extension comme l'ensemble flou μ_{comp} sur $[0, 1]$ défini par :

$$\forall t \in [0, 1], \mu_{comp}(t) = \begin{cases} 0 & \text{si } \mu^{-1}(t) = \emptyset, \\ \sup_{x \in \mathcal{S} | t=\mu(x)} \mu'(x) & \text{sinon.} \end{cases} \quad (4.142)$$

La construction de l'ensemble flou de compatibilité est illustré dans la figure 4.17. Il faut noter que ce n'est pas une notion symétrique en μ et μ' .

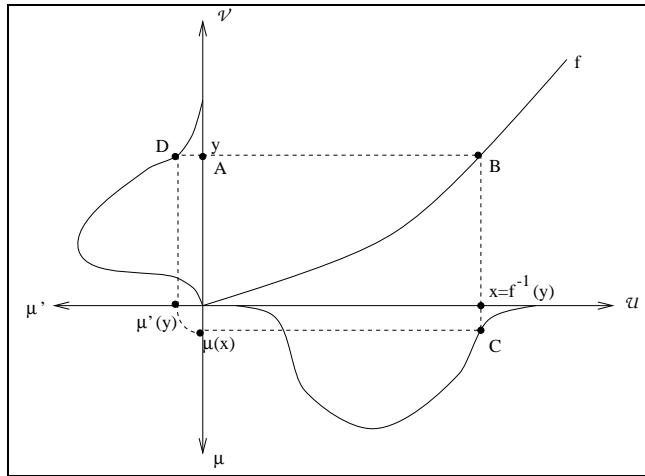


FIG. 4.16 – Principe d'extension. Partant d'un point A quelconque d'ordonnée y , B est le point correspondant sur le graphe de f (de même ordonnée) et donne $x = f^{-1}(y)$, dont l'appartenance $\mu(x)$ est représentée par le point C. L'appartenance de A de coordonnée y à l'ensemble flou image de μ par f est égale à $\mu(x)$.

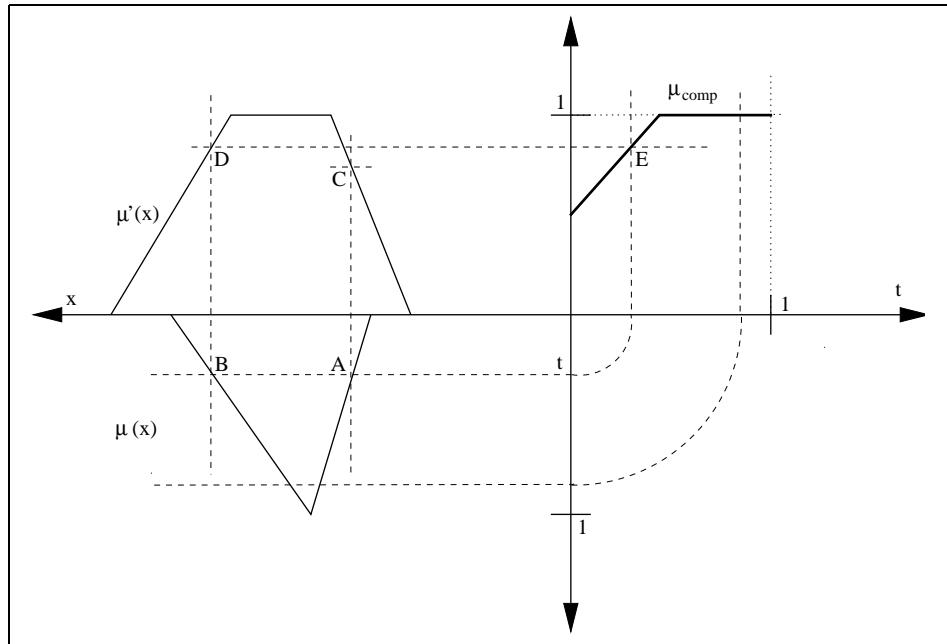


FIG. 4.17 – Construction de la fonction de compatibilité entre deux ensembles flous μ et μ' (le graphe de μ a été dessiné sur une échelle inversée pour faciliter la lecture).

Application aux nombres flous

Une des applications importantes du principe d'extension concerne les opérations sur des nombres flous, comme l'addition, la multiplication, etc. [120]. Si μ_A et μ_B sont deux nombres flous (ensembles flous convexes et normalisés sur \mathbb{R}), et $*$ est une opération quelconque sur

les nombres, l'extension de $*$ aux nombres flous est définie par :

$$\forall z \in \mathbb{R}, \mu_{A*B}(z) = \sup_{(x,y) \in \mathbb{R}^2 | x*y=z} \min[\mu_A(x), \mu_B(y)]. \quad (4.143)$$

L'extension d'une opération commutative (respectivement associative) est commutative (respectivement associative).

Le calcul de telles opérations peut être réalisé par des algorithmes simples dans le cas de nombres flous de type $L-R$ [120]. Par exemple, soient μ_1 et μ_2 deux nombres flous définis par :

$$\forall x \in \mathbb{R}, \mu_1(x) = \begin{cases} L\left(\frac{m_1-x}{\alpha_1}\right) & \text{pour } x \leq m_1, \\ R\left(\frac{x-m_1}{\beta_1}\right) & \text{pour } x \geq m_1. \end{cases} \quad (4.144)$$

$$\forall x \in \mathbb{R}, \mu_2(x) = \begin{cases} L\left(\frac{m_2-x}{\alpha_2}\right) & \text{pour } x \leq m_2, \\ R\left(\frac{x-m_2}{\beta_2}\right) & \text{pour } x \geq m_2. \end{cases} \quad (4.145)$$

La somme de μ_1 et μ_2 est un nombre flou $L-R$ de paramètres $m_1 + m_2$, $\alpha_1 + \alpha_2$ et $\beta_1 + \beta_2$.

La figure 4.18 illustre la différence et la somme de deux nombres flous obtenues par le principe d'extension.

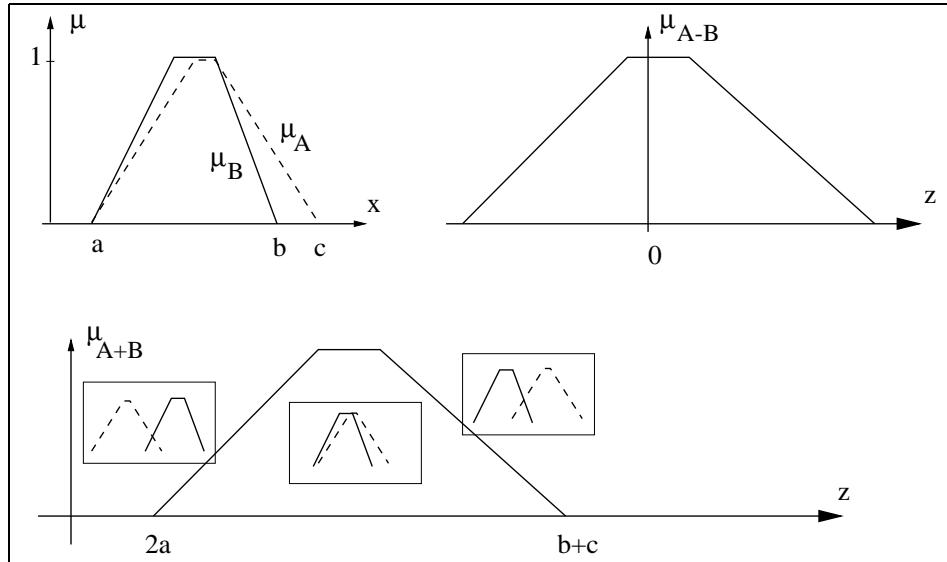


FIG. 4.18 – Application du principe d'extension au calcul de la différence (en haut à droite) et de la somme (en bas) de deux nombres flous.

4.8.2 Combinaison d'opérations sur les α -coupes

Reconstruction à partir des α -coupes

Un ensemble flou peut être vu comme l'empilement de ses α -coupes et on peut reconstruire un ensemble flou à partir de ses α -coupes (voir section 4.1.4). La construction d'une opération floue peut s'inspirer du même schéma, combinant les résultats de l'opération binaire équivalente appliquée sur les α -coupes.

Relations unaires. Soit μ la fonction d'appartenance d'un ensemble flou défini sur l'espace S . Considérons une fonction d'ensemble R_B prenant ses valeurs dans un espace \mathcal{V} . Son équivalent flou R , défini de \mathcal{F} dans \mathcal{V} , peut suivre un schéma de fuzzification par intégration [119, 202, 57] :

$$R(\mu) = \int_0^1 R_B(\mu_\alpha) d\alpha. \quad (4.146)$$

D'autres modes de fuzzification sont possibles, tels que [57, 158] :

$$R(\mu) = \sup_{\alpha \in [0,1]} \min(\alpha, R_B(\mu_\alpha)), \quad (4.147)$$

si la relation prend ses valeurs dans $[0, 1]$, ou :

$$R(\mu) = \sup_{\alpha \in [0,1]} (\alpha R_B(\mu_\alpha)). \quad (4.148)$$

Ces équations ne donnent pas forcément les mêmes résultats.

Relations binaires. Considérons maintenant une relation nette R_B de deux arguments. Son extension floue R , appliquée à μ et ν dans S , s'obtient par :

$$R(\mu, \nu) = \int_0^1 R_B(\mu_\alpha, \nu_\alpha) d\alpha, \quad (4.149)$$

ou par une double intégration :

$$R(\mu, \nu) = \int_0^1 \int_0^1 R_B(\mu_\alpha, \nu_\beta) d\alpha d\beta. \quad (4.150)$$

Les autres modes de fuzzification (équations 4.147 et 4.148) peuvent aussi être appliqués ici.

L'extension à des opérateurs n-aires est immédiate.

Principe d'extension fondé sur les α -coupes

Un autre moyen pour combiner les résultats sur les α -coupes, similaire au principe d'extension, permet d'obtenir des résultats sous forme d'ensemble flou sur \mathcal{V} . Par exemple si $\mathcal{V} = \mathbb{IR}$, la relation binaire fournit des valeurs réelles et son équivalent flou obtenu selon les méthodes précédentes donne également des nombres, alors que la méthode décrite ici donne des nombres flous. Pour une relation binaire, on a :

$$\forall n \in \mathcal{V}, R(\mu, \nu)(n) = \sup_{R_B(\mu_\alpha, \nu_\alpha)=n} \alpha. \quad (4.151)$$

La procédure est similaire dans le cas de relations unaires ou n-aires.

Si la relation à étendre prend seulement des valeurs binaires (0/1 ou faux/vrai), alors cette équation se réduit à :

$$R(\mu, \nu) = \sup_{R_B(\mu_\alpha, \nu_\alpha)=1} \alpha. \quad (4.152)$$

Le principe d'extension fondé sur les α -coupes peut être interprété de la manière suivante : choisir une α -coupe particulière revient à prendre une décision sur les limites de l'ensemble flou considéré. On regarde alors quelle est la valeur de la relation nette pour cette décision. Si différentes valeurs de α conduisent à la même valeur pour la relation, on ne retient que la plus grande de ces valeurs de α .

4.8.3 Traduction d'expressions binaires en expressions floues

La dernière classe de méthodes consiste à transposer des équations binaires en leur équivalent flou par une technique de « traduction » terme à terme. Cette approche diffère complètement de la précédente dans le sens où elle n'utilise pas explicitement la relation ou l'opération nette. En effet, aussi bien dans le principe d'extension que dans les approches fondées sur les α -coupes, l'opération floue est une fonction de l'opération nette correspondante. Dans l'approche décrite ici, l'opération floue est directement exprimée par une équation impliquant des concepts flous.

Ainsi, l'intersection est remplacée par une t-norme, la réunion par une t-conorme, les ensembles par des fonctions d'appartenance à des ensembles flous. Ce type de traduction est particulièrement simple à réaliser lorsque les relations nettes sont exprimées en termes logiques et ensemblistes. La table 4.3 regroupe les principaux concepts nets et leur équivalent flou.

concept net	concept flou équivalent
ensemble X	ensemble flou / fonction d'appartenance μ
complémentaire d'un ensemble	complémentation floue c
intersection \cap	t-norme t
réunion \cup	t-conorme T
existence \exists	supremum
symbole universel \forall	infimum

TAB. 4.3 – Dictionnaire des concepts nets et équivalents flous.

Les multiples possibilités pour traduire une réunion par exemple conduisent à autant de définitions possibles pour une opération floue, dépendant du choix des opérateurs utilisés (des t-conormes par exemple).

Illustrons cette méthode sur l'exemple de l'inclusion. Au début de ce chapitre, nous avons considéré qu'un ensemble flou μ était inclus dans un ensemble flou ν si :

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

Il s'agit donc d'une définition nette, en tout ou rien, de la notion d'inclusion. Mais si deux ensembles sont définis de manière imprécise, on peut s'attendre à ce que les relations qui les lient soient également imprécises, et l'inclusion devient alors une notion graduelle. Le principe de traduction exposé ci-dessus permet de définir un degré d'inclusion entre deux ensembles flous.

Dans le cas net, l'équation ensembliste exprimant l'inclusion d'un ensemble X dans un ensemble Y peut s'écrire :

$$X \subseteq Y \Leftrightarrow X^C \cup Y = \mathcal{S} \quad (4.153)$$

$$\Leftrightarrow \forall x \in \mathcal{S}, x \in X^C \cup Y. \quad (4.154)$$

En utilisant le dictionnaire de la table 4.3 on obtient les équivalences suivantes :

$$\forall x \in \mathcal{S} \Leftrightarrow \inf_{x \in \mathcal{S}}, \quad (4.155)$$

$$x \in X^C \Leftrightarrow c[\mu(x)], \quad (4.156)$$

$$x \in Y \leftrightarrow \nu(x), \quad (4.157)$$

$$X^C \cup Y \leftrightarrow T[c(\mu), \nu]. \quad (4.158)$$

Et finalement, le degré d'inclusion de μ dans ν est défini par :

$$\mathcal{I}(\mu, \nu) = \inf_{x \in S} T[c(\mu(x)), \nu(x)], \quad (4.159)$$

où T est une t-conorme et c une complémentation floue.

4.8.4 Comparaison

Le principe d'extension a été défini à l'origine pour des fonctions. Les approches présentées dans la section 4.8.2 s'occupent principalement d'opérateurs (opérateurs ensemblistes, relations entre ensembles, etc.). Il existe cependant des liens entre le principe d'extension et la combinaison des α -coupes par l'équation 4.147 [158]. Soit f une fonction de $S_1 \times \dots \times S_n$ dans V , et R_f un opérateur ensembliste défini par :

$$R_f(X_1, X_2, \dots, X_n) = \{f(x_1, x_2, \dots, x_n) | x_1 \in X_1, \dots, x_n \in X_n\}, \quad (4.160)$$

où X_1, \dots, X_n sont des sous-ensembles de S_1, \dots, S_n . Alors l'extension de R_f selon l'équation 4.147 coïncide avec l'extension de Zadeh de f (équation 4.139).

D'autres liens peuvent être établis entre les définitions de la section 4.8.2. Par exemple, si R_B est une relation nette prenant ses valeurs dans $\{0, 1\}$, son extension selon l'équation 4.152 est une valeur dans $[0, 1]$ et est équivalente aux deux procédures de fuzzification selon les équations 4.147 et 4.148.

Prenons l'exemple de l'extension de la réunion entre ensembles à l'aide des méthodes de la section 4.8.2. Soit μ et ν deux ensembles flous sur S . Par intégration sur les α -coupes, on obtient :

$$\begin{aligned} (\mu \cup \nu)(x) &= \int_0^1 (\mu_\alpha \cup \nu_\alpha)(x) d\alpha \\ &= \int_0^{\max[\mu(x), \nu(x)]} 1 d\alpha \\ &= \max[\mu(x), \nu(x)], \end{aligned}$$

puisque $(\mu_\alpha \cup \nu_\alpha)(x) = 1$ si et seulement si $\mu(x) \geq \alpha$ ou $\nu(x) \geq \alpha$. Cette extension donne exactement la réunion originale de Zadeh [358]. On obtient exactement le même résultat avec les autres méthodes de fuzzification, par exemple selon les équations 4.147 ou 4.152.

En utilisant le principe de traduction formelle décrit dans la section 4.8.3, on peut parfois obtenir les mêmes résultats qu'avec la combinaison des α -coupes. L'exemple de la connexité que nous verrons plus loin en est une illustration.

La question du choix de la méthode d'extension est encore ouverte. Le principe d'extension est bien adapté pour transposer des expressions analytiques, alors que la traduction formelle est appropriée à la transposition d'équations ensemblistes et logiques. Les propriétés des opérations floues obtenues jouent un grand rôle dans le choix d'une méthode, car elles peuvent être très différentes d'une méthode à l'autre. Par exemple, comme nous le verrons plus loin, une simple intégration sur les α -coupes (équation 4.149) pour étendre une distance entre ensembles flous permet de conserver les mêmes propriétés que la distance nette, alors qu'une double intégration (équation 4.150) conduit à la perte de certaines propriétés.

Chapitre 5

Théorie des ensembles flous et des possibilités en traitement d'images

5.1 Introduction

Imprécisions et incertitudes sont inhérentes aux images, et trouvent leur origine à différents niveaux : phénomènes observés, capteurs et acquisitions des données, processus numériques de reconstruction, nature des images et mode de représentation de leurs éléments constitutifs, etc. Par exemple, la représentation d'informations (symboliques) sous forme schématique (par des cartes ou des atlas) est source à la fois d'imprécision et d'incertitude. Imprécisions et incertitudes sont ensuite renforcées dans les primitives extraites des images et sur lesquelles s'appuient l'analyse et l'interprétation des images. Cet antagonisme entre précision et certitude a bien été identifié comme un trait caractéristique de la démarche en reconnaissance des formes [308]. Il se traduit par le fait que si les données ou les connaissances et leurs représentations sont précises, alors elles sont probablement incertaines ; si l'on renforce leur certitude, cela se fait souvent au prix de plus d'imprécision, rendant les données peu informatives si cette imprécision devient trop importante.

L'intérêt des ensembles flous pour le traitement de l'information en image et vision [200] peut se décliner en particulier selon les quatre aspects suivants :

- la capacité des ensembles flous à représenter l'information spatiale dans les images ainsi que son imprécision, à différents niveaux (local, régional ou global), et sous différentes formes (numérique, symbolique, quantitative, qualitative),
- la possibilité de représenter des informations très hétérogènes, extraites directement des images ou issues de connaissances externes, comme des connaissances expertes ou génériques sur un domaine ou un problème,
- la possibilité de généraliser aux ensembles flous des opérations pour manipuler l'information spatiale,
- la souplesse des opérateurs de combinaison permettant de fusionner des informations de multiples natures dans des situations très variées.

Ce chapitre se propose d'illustrer quelques uns de ces aspects. Il ne prétend pas être exhaustif. Dans la partie 5.2, nous définissons des objets flous spatiaux, permettant de manipuler des objets avec leur imprécision spatiale dans les images. Des objets géométriques particuliers sont définis, ainsi que des mesures et des transformations géométriques. Dans

la partie 5.3 nous nous intéressons aux opérations de morphologie mathématique floue, que nous illustrons brièvement sur une application médicale. Dans la partie 5.4, nous étendons des outils topologiques simples au cas d'objets flous (connexité, frontière, adjacence). Dans les parties 5.5 et 5.6, nous présentons des relations spatiales métriques, de distance puis de position relative directionnelle. Les principales méthodes de classification floue utilisées en traitement d'images sont décrites dans la partie 5.7. Nous présentons ensuite quelques systèmes flous d'amélioration d'image et de détection de contours dans la partie 5.8. Les principaux aspects de la fusion d'images par des techniques floues seront abordées dans le chapitre 8.

Dans toute la suite, on désignera par \mathcal{S} le domaine de l'image, typiquement \mathbb{Z}^2 ou \mathbb{Z}^3 pour des images discrètes 2D ou 3D, et \mathbb{R}^2 ou \mathbb{R}^3 dans le cas continu.

5.2 Représentation de l'information spatiale

5.2.1 Objets flous spatiaux

Nous appelons objet flou spatial un ensemble flou défini dans le domaine de l'image. Sa fonction d'appartenance μ (fonction de \mathcal{S} dans $[0, 1]$) représente l'imprécision sur la définition spatiale de l'objet (ses position, taille, forme, limites, etc.). Pour tout point x de \mathcal{S} (pixel ou voxel), $\mu(x)$ représente le degré avec lequel x appartient à l'objet flou.

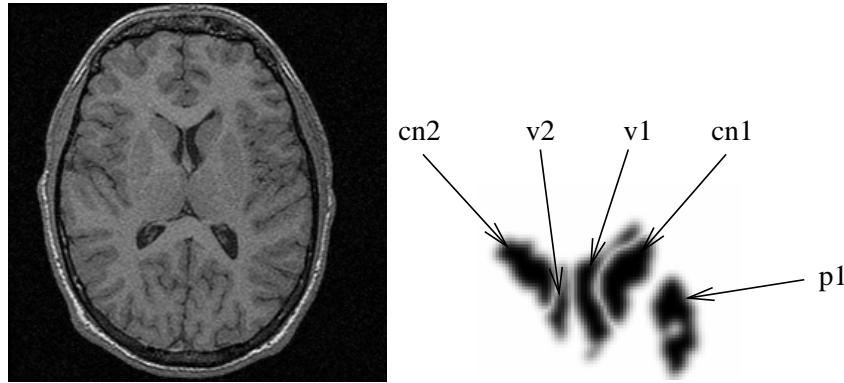


FIG. 5.1 – Image IRM de cerveau (une coupe extraite du volume 3D) et 5 objets flous représentant des structures internes du cerveau extraits de cette image (les valeurs d'appartenance varient de 0 à 1, du blanc au noir).

À titre d'exemple, une coupe d'une image de cerveau acquise en imagerie par résonance magnétique (IRM) est montrée sur la figure 5.1. et quelques structures internes sont représentées comme des objets spatiaux flous. Le flou peut représenter différents types d'imprécision, sur la frontière des objets (à cause par exemple de la résolution spatiale et du volume partiel), sur la variabilité inter-individuelle de ces structures, etc.

Une de ces structures ($v1$: ventricule latéral gauche) est représentée sur la figure 5.2, avec son support, son α -coupe de niveau 0,5, et sa fonction d'appartenance sous forme de graphe 3D.

La construction de ces ensembles flous spatiaux peut être effectuée de plusieurs manières. Une première méthode consiste à définir la fonction d'appartenance d'un objet flou à partir

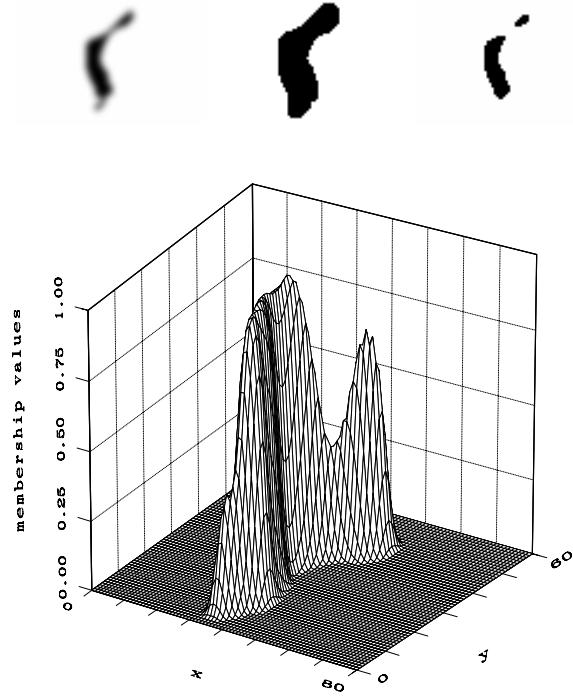


FIG. 5.2 – Un objet flou (v1), son support, son α -coupe de niveau 0,5, et sa fonction d'appartenance sous forme de graphe 3D (deux axes représentent les coordonnées spatiales, et le troisième les degrés d'appartenance à la structure).

de la fonction d'intensité I (les niveaux de gris) de l'image : $\forall x \in \mathcal{S}, \mu(x) = F[(I(x))]$, où F est une fonction à déterminer en fonction du problème. Les fonctions les plus utilisées sont des fonctions de normalisation ou des fonctions en S [260] (ce qui revient à considérer que les parties claires de l'image ont une appartenance élevée à l'objet), des fonctions Π (monomodales, elles associent l'objet à une plage de niveaux de gris aux frontières imprécises), ou encore des fonctions multimodales. Ces fonctions sont souvent déterminées de manière supervisée, mais elles peuvent également être apprises, par exemple à partir d'algorithmes de classification automatique tels que les C-moyennes floues [23] ou les C-moyennes possibilistes [201] (voir par exemple [24] pour une revue des algorithmes de classification floue utilisés en traitement d'images, et la section 5.7). D'autres caractéristiques peuvent être utilisées. Par exemple, l'ensemble des contours d'une image peut être défini par un ensemble flou spatial dont la fonction d'appartenance est une fonction du gradient de l'image : $\forall x \in \mathcal{S}, \mu(x) = F[(\nabla I(x))]$, où F est une fonction décroissante du gradient. Si l'on dispose de détecteurs d'objets particuliers, la fonction d'appartenance à ces objets est alors définie comme une fonction de la réponse à ces détecteurs (le cas des contours fait partie de cette catégorie). Par exemple, un détecteur de routes peut fournir dans une image satellitaire une réponse d'amplitude croissante avec l'appartenance à la route. Enfin, des objets flous peuvent être définis à partir d'une détection préliminaire binaire des objets, afin de tenir compte typiquement des imprécisions possibles dans la délimitation de ces objets. On construit alors une fonction d'appartenance qui vaut 1 à l'intérieur de l'objet à une certaine distance des bords, 0 à l'extérieur de l'objet également à une certaine distance des bords, et décroissante entre ces deux limites. Par exemple, on peut modéliser une zone d'imprécision au bord de l'objet comme la zone comprise entre

l'érodé et le dilaté de cet objet, la taille de ces opérations dépendant de l'extension spatiale de l'imprécision que l'on veut représenter. Si O est l'objet binaire de départ, $E^n(O)$ son érodé de taille n et $D^m(O)$ son dilaté de taille m , la fonction d'appartenance à l'objet flou peut être définie par : $\mu(x) = 1$ si $x \in E^n(O)$, $\mu(x) = 0$ si $x \in S - D^m(O)$, et $\mu(x) = F[d(x, E^n(O))]$ sinon, où F est une fonction décroissante de la distance d de x à $E^n(O)$.

5.2.2 Opérations ensemblistes

Les opérations ensemblistes entre ensembles flous spatiaux (intersection ou réunion de deux objets flous) sont effectuées à l'aide des opérateurs flous classiques vus dans le chapitre 4 (t-normes et t-conormes [120, 124, 351]). Le résultat est alors encore un ensemble flou spatial. Une autre question qui peut se poser à propos des opérations ensemblistes est de savoir avec quel degré une relation ensembliste entre deux objets flous est satisfaite. Ce type d'évaluation intervient dans beaucoup d'opérations spatiales, comme on le verra dans les sections suivantes. Une des méthodes les plus simples, et qui conduit en général à de bonnes propriétés, consiste à transposer terme à terme des équations ensemblistes binaires pour les généraliser au flou [57]. Typiquement, un ensemble est remplacé par une fonction d'appartenance à un ensemble flou, une intersection par une t-norme, une réunion par une t-conorme, un symbole existentiel par un sup, etc. (voir la section 4.8 du chapitre 4). Par exemple, X et Y étant deux ensembles binaires, l'écriture $X \cap Y \neq \emptyset$ est équivalente à $\exists x \in S, x \in X \cap Y$. Une transposition directe de cette équation en termes flous permet de définir le degré d'intersection entre deux ensembles flous μ et ν par :

$$\mu_{int}(\mu, \nu) = \sup_{x \in S} t[\mu(x), \nu(x)], \quad (5.1)$$

où t est une t-norme. De manière similaire, le degré d'inclusion de μ dans ν peut être défini par :

$$\mathcal{I}(\mu, \nu) = \inf_{x \in S} T[c(\mu(x)), \nu(x)], \quad (5.2)$$

où T est une t-conorme et c une complémentation. L'inclusion floue a été formalisée par de nombreux auteurs, qui en proposent d'autres formes (par exemple [163, 184, 310, 356, 199]).

Les définitions à partir des t-normes et t-conormes ont de très bonnes propriétés algébriques, mais ont l'inconvénient de ne retenir qu'un point de l'espace, et ne sont donc pas toujours adaptées au traitement d'images. Par exemple pour le degré d'intersection, on peut obtenir la même valeur pour des objets flous ayant des recouvrements très différents. Cette information de recouvrement spatial peut être introduite en mesurant le volume de l'intersection des deux ensembles flous, et on obtient par exemple :

$$\mu_{int}(\mu, \nu) = \frac{V[t(\mu, \nu)]}{\min[V(\mu), V(\nu)]}, \quad (5.3)$$

où le volume d'un objet flou (ou sa surface en 2D) est calculé simplement par sa cardinalité floue :

$$V(\mu) = \sum_{x \in S} \mu(x), \quad (5.4)$$

pour le cas discret, et dans le cas continu :

$$V(\mu) = \int_{x \in S} \mu(x) dx. \quad (5.5)$$

Cette expression est plus discriminante, ce qui est souhaitable pour certaines applications. On en verra une utilisation dans la section 5.4 pour l'adjacence entre deux objets flous. La forme donnée par l'équation 5.1 sera utilisée pour les opérations morphologiques dans la section 5.3.

Pour les degrés d'inclusion, ceux proposés par exemple dans [163, 199] effectuent des sommes sur tout l'espace et prennent donc également en compte l'extension spatiale des objets.

Ainsi, pour des applications en image, la forme des mesures choisies est importante, et son choix s'ajoute à celui des t-normes et t-conormes.

5.2.3 Ensembles flous géométriques

Des objets flous particuliers peuvent être définis, tels que des points, droites, rectangles, disques flous. Nous en donnons ici les principales définitions.

Des points flous sont des extensions dans l'espace à n dimensions de nombres flous, considérés comme des points sur la droite réelle selon [365], c'est-à-dire des ensembles flous de fonction d'appartenance semi-continue, de valeur modale unique et convexe [71].

Des droites floues ont été définies dans [71] dans un espace bidimensionnel. Des extensions dans des espaces de dimension supérieure sont immédiates en suivant le même principe. Ce principe consiste à considérer une forme analytique de droite classique, et à considérer que ses paramètres sont des nombres flous et non plus des nombres classiques. La fonction d'appartenance d'une droite floue est alors calculée de manière similaire au principe d'extension. Ainsi, si une droite du plan est définie analytiquement sous la forme $ax + by = c$, on définit une droite floue par sa fonction d'appartenance μ_D :

$$\mu_D(x, y) = \sup\{\alpha \in [0, 1] \mid ax + by = c, a \in \mu_{A_\alpha}, b \in \mu_{B_\alpha}, c \in \mu_{C_\alpha}\}, \quad (5.6)$$

où μ_A , μ_B et μ_C sont les fonctions d'appartenance de trois nombres flous.

Ce type de calcul peut être appliqué à n'importe quelle forme analytique de droite. La figure 5.3 en montre un exemple. Il s'étend également à des variétés de dimension supérieure comme des plans, et à n'importe quelle structure géométrique dont on connaît une expression analytique paramétrique.

Des équivalents flous de la notion de rectangle ou plus généralement de polygones convexes ont été proposés dans [283]. Un rectangle flou est un ensemble flou séparable et convexe. La séparabilité signifie que la fonction d'appartenance s'exprime comme la conjonction de fonctions sur chaque axe. La figure 5.4 en montre un exemple.

Enfin, un disque flou est un ensemble flou pour lequel il existe un point P de \mathcal{S} tel que la fonction d'appartenance μ ne dépende que de la distance r à P , qui est alors le centre du disque flou. Le disque flou est convexe si et seulement si $\mu(r)$ est décroissante. En général on suppose que le support de μ est borné. Les α -coupes d'un disque flou convexe sont des disques concentriques. Si de plus $\mu(r)$ a une unique valeur modale (en $r = 0$), alors le disque flou convexe est un point flou. Deux exemples de disques flous, l'un convexe, l'autre non, sont montrés sur la figure 5.5.

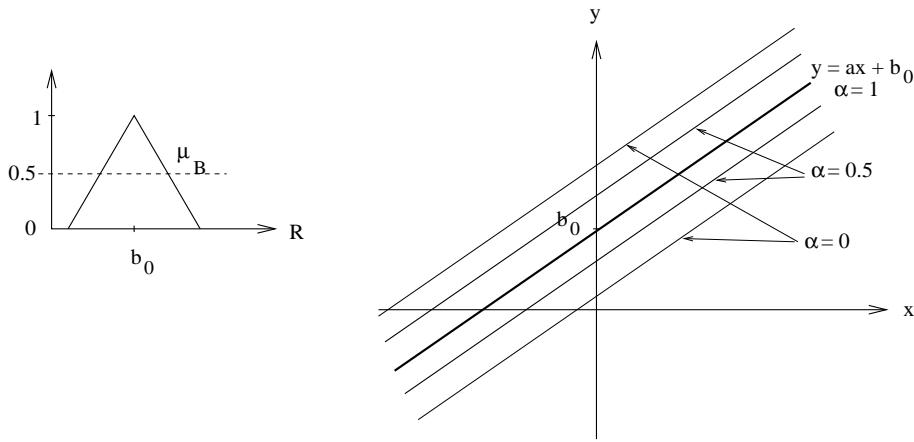


FIG. 5.3 – Exemple de droite floue dans le plan, dans laquelle seul le paramètre b est flou (et donné par μ_B à gauche) dans l'équation $y = ax + b$. Trois α -coupes sont montrées.

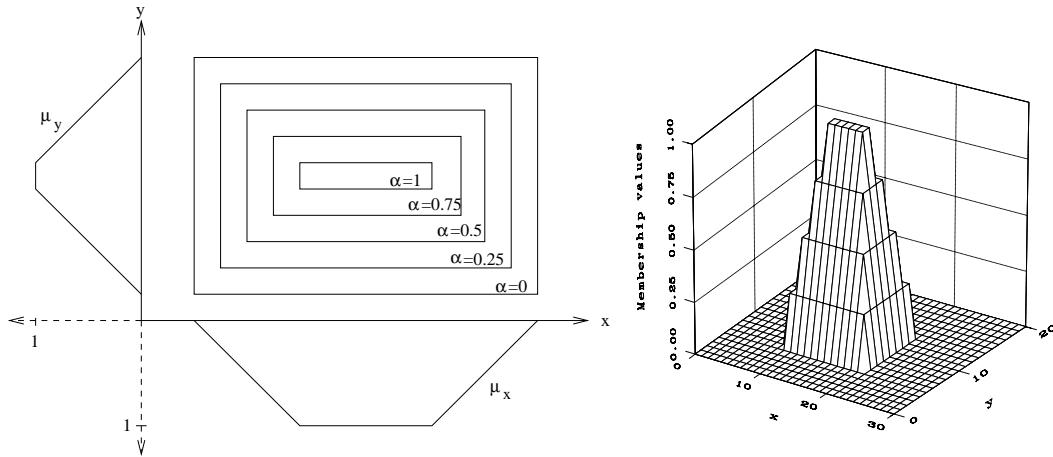


FIG. 5.4 – Exemple de rectangle flou dans le plan. À gauche : μ_x et μ_y et quelques α -coupes. À droite : représentation 3D du rectangle flou.

5.2.4 Mesures géométriques d'objets flous

Dans cette partie, nous donnons quelques exemples de mesures géométriques calculées sur des objets flous.

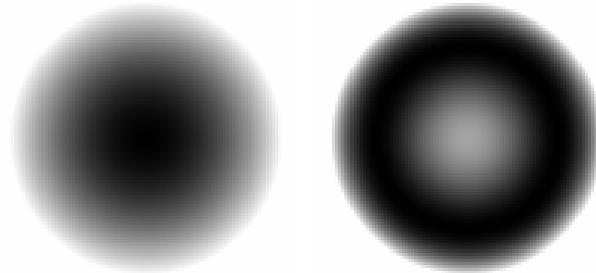


FIG. 5.5 – Deux exemples de disques flous dans le plan. Celui de gauche est convexe, celui de droite ne l'est pas.

L'exemple du **volume** a déjà été mentionné plus haut.

Le **périmètre** d'un objet flou de fonction d'appartenance μ est défini par [285] :

$$p(\mu) = \int_{x \in S} |\nabla \mu(x)| dx. \quad (5.7)$$

Dans le cas discret 2D, cette expression revient à calculer la somme pondérée des longueurs des courbes délimitant les α -coupes, les poids étant la différence des valeurs d'appartenance entre deux α -coupes successives.

La **compacité** en 2D est définie par [283] :

$$c(\mu) = \frac{V(\mu)}{p(\mu)^2}, \quad (5.8)$$

où V désigne le volume (ici en 2D, la surface) et p le périmètre. On peut montrer que pour des disques flous convexes, on a $c(\mu) \geq \frac{1}{4\pi}$, et donc la plus faible compacité est obtenue pour des disques nets.

La **hauteur** et la **largeur** d'un ensemble flou 2D sont définies comme le volume des projections sur chacun des axes. Par exemple la hauteur est définie en 2D par [283] :

$$h(\mu) = \int_y [\sup_x \mu(x, y)] dy. \quad (5.9)$$

Des notions simples de **parallélisme** entre droites floues ont été introduites dans [71] : le degré de parallélisme entre deux droites floues est égal à $1 - h$ où h est la hauteur de l'intersection entre les deux droites floues.

5.2.5 Mesures géométriques floues d'objets flous

Dans les définitions précédentes, les mesures géométriques sont définies comme des nombres nets. Or si les objets sont imprécis, on peut s'attendre à ce que leurs mesures soient également imprécises, et donc définies comme des nombres flous, qui peuvent être calculés selon le principe d'extension. Ainsi, pour une mesure M (surface, périmètre, etc.) sa forme floue s'exprime par :

$$\forall \lambda \in \mathbb{R}^+, M(\mu)(\lambda) = \sup_{M(\mu_\alpha)=\lambda} \alpha, \quad (5.10)$$

qui exprime le degré avec lequel la mesure M appliquée à μ vaut λ . Par exemple pour le volume, la définition obtenue ainsi est équivalente à la cardinalité d'un ensemble flou défini comme un nombre flou dans [120]. Comme les α -coupes sont emboîtées, leur volume est décroissant en α , et donc le volume flou a une fonction d'appartenance décroissante (voir figure 5.6). Cette propriété n'est pas vraie pour toutes les mesures (par exemple elle ne l'est pas pour le périmètre, comme le montre la figure 5.6 à droite).

Ce type de représentation est très utile lorsque, dans des étapes ultérieures de traitement, on souhaite évaluer des propositions de la forme « la mesure est grande, petite, etc. », où « grande », « petite » sont des valeurs de variables linguistiques. Ces évaluations peuvent être effectuées en comparant les fonctions d'appartenance de la mesure géométrique floue et celles de la variable linguistique, par des techniques de compatibilité, de similarité ou de « pattern matching ».

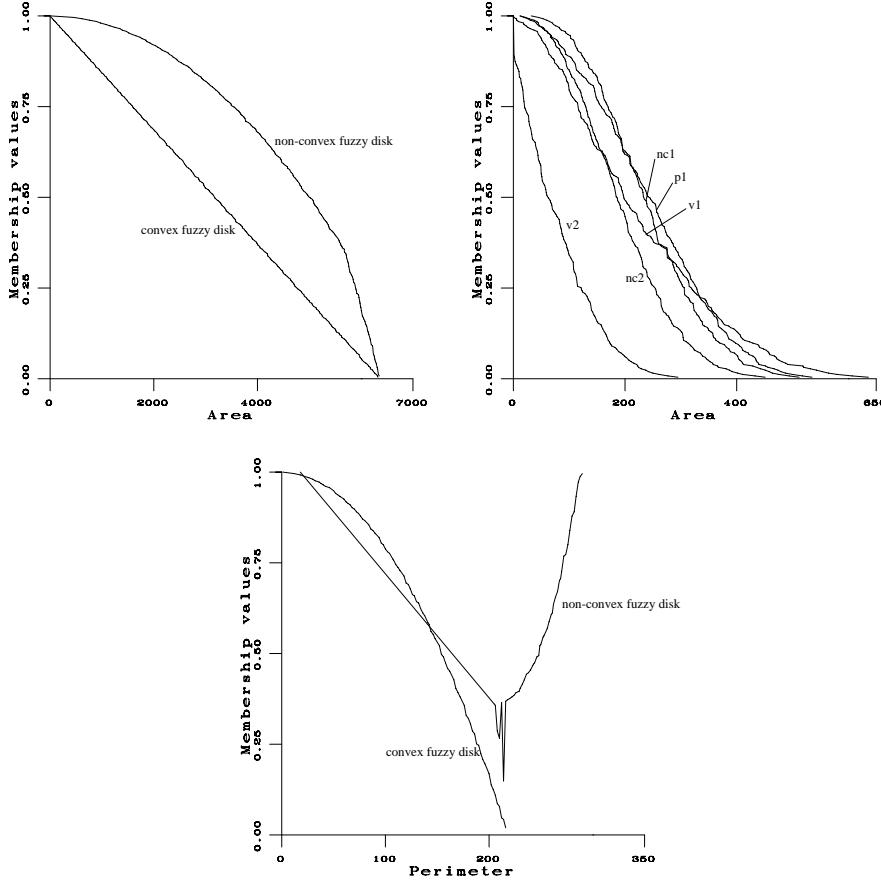


FIG. 5.6 – Surface exprimée comme un nombre flou, illustrée sur les disques flous de la figure 5.5 et sur les structures du cerveau de la figure 5.1. En bas, périmètre flou illustré sur les disques flous de la figure 5.5.

5.2.6 Transformations géométriques

Dans cette partie, nous décrivons comment calculer des transformations géométriques d'ensembles flous spatiaux, telles que des translations ou des rotations.

Transformations nettes. Tout d'abord, nous supposons que seul l'ensemble à transformer est flou. Soit \mathcal{T} une transformation géométrique quelconque (translation, rotation, symétrie, homothétie). Nous supposons que \mathcal{T} est bien définie, c'est-à-dire que les paramètres de la transformation sont des nombres usuels. Soit μ un ensemble flou dans \mathcal{S} (en dimension quelconque). La transformation de μ par \mathcal{T} est alors simplement définie par :

$$\forall x \in \mathcal{S}, \mathcal{T}(\mu)[\mathcal{T}(x)] = \mu(x), \quad (5.11)$$

ou de manière équivalente :

$$\forall x \in \mathcal{S}, \mathcal{T}(\mu)(x) = \mu[\mathcal{T}^{-1}(x)], \quad (5.12)$$

Dans le cas discret, $\mathcal{T}(x)$ peut ne pas appartenir à \mathcal{S} (il ne coïncide pas forcément avec

un point de la trame discrète). Ce problème est également rencontré en traitement d'images classique, et des méthodes similaires d'interpolation peuvent être utilisées.

Transformations floues. Nous considérons maintenant que la transformation est également imprécise, et que ses paramètres sont cette fois des nombres flous. Soit \mathcal{T} la transformation floue, dépendant de paramètres p_1, \dots, p_n , chacun d'eux étant flou. Soient μ_{p_i} les fonctions d'appartenance définissant les paramètres p_i . Nous notons $\mathcal{T}_{p_1, \dots, p_n}$ la transformation nette obtenue pour des valeurs précises des paramètres. La transformation d'un point y de \mathcal{S} par \mathcal{T} est un ensemble flou, dont la fonction d'appartenance est :

$$\forall z \in \mathcal{S}, \mu_{\mathcal{T}(y)}(z) = \sup_{p_1, \dots, p_n | \mathcal{T}_{p_1, \dots, p_n}(y)=z} t[\mu_{p_1}(p_1), \dots, \mu_{p_n}(p_n)], \quad (5.13)$$

où t est une t-norme.

Maintenant, étendons cette définition au cas d'un point appartenant à un ensemble flou μ . Les valeurs d'appartenance $\mu(y)$ doivent être combinées à celles de $\mu_{\mathcal{T}(y)}$. Nous définissons alors la transformation de μ par [61] :

$$\forall x \in \mathcal{S}, \mathcal{T}(\mu)(x) = \sup_{y \in \mathcal{S}} t[\mu(y), \mu_{\mathcal{T}(y)}(x)]. \quad (5.14)$$

En pratique, le calcul n'est pas toujours immédiat. Une solution consiste à appliquer directement le principe d'extension, mais le calcul peut en être très lourd. Une autre solution est d'exprimer la transformation sous forme analytique à partir de représentations paramétriques des paramètres flous, de manière analogue aux opérations sur les nombres flous de type L-R. Mais de tels calculs ne sont pas possibles pour toutes les transformations, ni pour des chaînes de traitement où les mêmes variables peuvent être impliquées plusieurs fois. Dans le cas discret, une autre solution a été proposée dans [61] : le domaine de définition de chaque paramètre peut être discrétilisé, et pour chaque valeur possible des paramètres, le point correspondant ainsi que son degré d'appartenance sont calculés par l'équation 5.14.

5.3 Morphologie mathématique floue

La morphologie mathématique [300, 301, 294] est une théorie largement employée en traitement d'images binaires et numériques. Elle propose des transformations non linéaires utilisées en filtrage et amélioration d'images, segmentation, mesures, reconnaissance des formes. Puisqu'elle repose essentiellement sur des concepts ensemblistes et des propriétés algébriques, la morphologie mathématique apparaît comme une suite logique dans la série des opérations manipulant l'information spatiale et de leur extension au flou. Un autre intérêt de l'extension de la morphologie mathématique au flou est que d'autres opérateurs peuvent en être déduits (adjacence ou distance par exemple) comme on le verra plus loin [45].

Les premiers travaux portant sur la morphologie mathématique floue datent des années 80 et considéraient simplement les niveaux de gris d'une image comme une fonction d'appartenance [160, 159, 205, 114, 115, 253] ou n'utilisaient que des éléments structurants binaires [283]. Ici, nous restreignons la présentation aux approches réellement floues, dans lesquelles des ensembles flous sont transformés à l'aide d'éléments structurants flous. Les développements initiaux peuvent être trouvés dans la définition de l'addition de Minkovski

floue [121, 190]. Puis ce problème a été abordé indépendamment par plusieurs auteurs [53, 56, 57, 309, 97, 16, 273, 311, 252]. Nous nous attachons ici principalement aux quatre opérations de base de la morphologie mathématique que sont l'érosion, la dilation, l'ouverture et la fermeture, à partir desquelles de nombreuses autres opérations peuvent être construites (les définitions et résultats sont donnés de manière plus détaillée dans [57] par exemple). On se restreindra également aux opérateurs qui sont invariants par translation, et qui peuvent donc s'exprimer à l'aide d'éléments structurants translatés en toutes les positions de l'image (on pourrait aussi considérer la définition la plus générale de la dilatation comme opération croissante qui commute avec la réunion ou le sup et l'étendre au flou, ou encore la notion d'adjonction [106]).

Rappelons tout d'abord les définitions des opérations élémentaires. On notera X (ou f) l'ensemble (ou la fonction) à transformer et B (ou g) l'élément structurant ensembliste (ou fonctionnel). La dilatation dans l'espace \mathcal{S} est définie par [300, 294] :

$$D_B(X) = \{x \in \mathcal{S} \mid B_x \cap X \neq \emptyset\}, \quad (5.15)$$

où B_x désigne le translaté de B au point x ,

$$D_B(f)(x) = \sup\{f(y) \mid y \in B_x\}, \quad (5.16)$$

$$D_g(f)(x) = \sup\{f(y) + g(y - x) \mid y \in \mathcal{S}\}. \quad (5.17)$$

L'érosion est définie par :

$$E_B(X) = \{x \in \mathcal{S} \mid B_x \subset X\}, \quad (5.18)$$

$$E_B(f)(x) = \inf\{f(y) \mid y \in B_x\}, \quad (5.19)$$

$$D_g(f)(x) = \inf\{f(y) - g(y - x) \mid y \in \mathcal{S}\}. \quad (5.20)$$

Dans les trois cas, l'ouverture et la fermeture sont définies respectivement par :

$$X_B = D_{\check{B}}[E_B(X)], \quad (5.21)$$

$$X^B = E_{\check{B}}[D_B(X)], \quad (5.22)$$

où \check{B} désigne le symétrique de B par rapport à l'origine.

Par l'intermédiaire de sa fonction d'appartenance, un ensemble flou peut être assimilé à une fonction qui prend ses valeurs dans $[0, 1]$. Cependant, le cas le plus général d'une opération sur un ensemble flou avec un élément structurant flou ne peut pas être simplement construit à partir de la morphologie fonctionnelle. En effet, une dilatation d'une fonction à valeurs dans $[0, 1]$ par une autre donne en général une fonction à valeur dans $[0, 2]$. En revanche, dans les cas où l'une des deux fonctions ne prend ses valeurs que dans $\{0, 1\}$, on peut demander à la morphologie floue de coïncider avec la morphologie classique.

5.3.1 Définitions

Les constructions de la morphologie mathématique floue reposent sur des principes assez classiques d'extension d'opérations binaires au cas flou, soit à partir du principe d'extension ou par combinaison d'opérations appliquées sur les α -coupes, soit par traduction directe d'équations ensemblistes en équations floues. Ces différentes constructions concernent surtout la dilatation et l'érosion, l'ouverture et la fermeture étant toujours obtenues par composition de ces deux opérations.

Morphologie floue à partir des α -coupes. Considérons d'abord la méthode reposant sur une combinaison par intégration des opérations binaires appliquées sur les α -coupes [56, 54]. Soient μ et ν deux ensembles flous définis sur l'espace \mathcal{S} . La dilatation de μ par ν s'obtient par :

$$\forall x \in \mathcal{S}, D_\nu(\mu)(x) = \int_0^1 \int_0^1 D_{\nu_\alpha}(\mu_\beta)(x) d\alpha d\beta = \int_0^1 \sup_{y \in (\nu_\alpha)_x} \mu(y) d\alpha. \quad (5.23)$$

De la même manière, l'érosion de μ par ν est définie par :

$$\forall x \in \mathcal{S}, E_\nu(\mu)(x) = \int_0^1 \inf_{y \in (\nu_\alpha)_x} \mu(y) d\alpha, \quad (5.24)$$

où $(\nu_\alpha)_x$ est la translation de ν_α (α -coupe de ν) au point x . Ces définitions garantissent que $D_\nu(\mu)$ et $E_\nu(\mu)$ sont les fonctions d'appartenance d'ensembles flous (prenant leurs valeurs dans $[0, 1]$). Par construction, si ν est binaire, ces définitions coïncident avec celles de la morphologie classique.

D'autres combinaisons sont possibles, de la même manière qu'un ensemble flou peut être reconstruit de multiples manières à partir de ses α -coupes. On peut ainsi définir la dilatation floue par :

$$\forall x \in \mathcal{S}, D_\nu(\mu)(x) = \sup_{y \in \mathcal{S}} \min[\mu(y), \nu(y - x)], \quad (5.25)$$

ou encore :

$$\forall x \in \mathcal{S}, D_\nu(\mu)(x) = \sup_{y \in \mathcal{S}} [\mu(y)\nu(y - x)], \quad (5.26)$$

et des équations similaires pour l'érosion, et en appliquant le principe d'extension, on obtient également la formule 5.25. Contrairement à celles obtenues par intégration, ces équations sont en fait des cas particuliers de celles obtenues par traduction formelle, définies dans le paragraphe suivant.

Morphologie floue par traduction formelle à partir de t-normes et t-conormes. Une autre méthode de construction consiste à remplacer formellement, dans les équations définissant les opérations, les symboles binaires ou logiques par leurs équivalents flous. La dilatation et l'érosion sont alors construites à partir de degrés d'intersection et d'inclusion, ainsi qu'ils ont été définis dans le paragraphe 5.2.2 (équations 5.1 et 5.2). Ainsi, la dilatation floue de μ par ν est définie par :

$$\forall x \in \mathcal{S}, D_\nu(\mu)(x) = \sup_{y \in \mathcal{S}} t[\nu(y - x), \mu(y)], \quad (5.27)$$

où t est une t-norme, et l'érosion floue de μ par ν est définie par :

$$\forall x \in \mathcal{S}, E_\nu(\mu)(x) = \inf_{y \in \mathcal{S}} T[c(\nu(y - x)), \mu(y)], \quad (5.28)$$

où T est la t-conorme duale de t par rapport à la complémentation c (la dualité permet de garantir la dualité entre D_ν et E_ν).

Ces formes de dilatation et érosion floues sont très générales, et plusieurs définitions de la littérature en sont des cas particuliers et correspondent aux équations 5.27 et 5.28 pour différents choix de t et T . Les définitions proposées dans des cas particuliers d'éléments structurants binaires sont aussi incluses dans ce cadre plus général.

La construction des ouvertures et des fermetures est simplement obtenue par composition d'une érosion et d'une dilatation utilisant des t-normes et t-conormes duales.

Utilisation de t-normes et t-conormes faibles. Si on remplace les t-normes et les t-conormes par des opérateurs plus faibles (en général non associatifs et qui n'admettent pas 1 ou 0 comme élément neutre), les équations 5.27 et 5.28 sont alors des généralisations des définitions proposées dans [310] :

$$\forall x \in \mathcal{S}, D_\nu(\mu)(x) = \sup_{y \in \mathcal{S}} \max[0, 1 - \lambda(\mu(y)) - \lambda(\nu(y - x))], \quad (5.29)$$

$$\forall x \in \mathcal{S}, E_\nu(\mu)(x) = \inf_{y \in \mathcal{S}} \min[1, \lambda(1 - \mu(y)) + \lambda(\nu(y - x))], \quad (5.30)$$

où λ est une fonction de $[0, 1]$ dans $[0, 1]$ satisfaisant les 6 conditions suivantes : $\lambda(z)$ est une fonction non croissante de z , $\lambda(0) = 1$, $\lambda(1) = 0$, l'équation $\lambda(z) = 0$ admet une solution unique, $\forall \alpha \in [0, 5, 1]$, l'équation $\lambda(z) = \alpha$ admet une solution unique, $\forall z \in [0, 1]$, $\lambda(z) + \lambda(1 - z) \geq 1$.

La t-norme et la t-conorme faibles utilisées dans ces définitions sont :

$$\forall (a, b) \in [0, 1]^2, t(a, b) = \max[0, 1 - \lambda(a) - \lambda(b)], \quad (5.31)$$

$$\forall (a, b) \in [0, 1]^2, T(a, b) = \min[1, \lambda(1 - a) + \lambda(1 - b)]. \quad (5.32)$$

Le cas $\lambda(z) = 1 - z = \lambda_0(z)$ conduit à de vraies t-norme et t-conorme, correspondant aux opérateurs de Lukasiewicz. Dans le cas général, ces définitions ont des propriétés beaucoup plus faibles que celles obtenues à partir de vraies t-normes et t-conormes.

Morphologie floue à partir d'implications résiduelles. Une autre manière de considérer l'inclusion floue est de la définir à partir d'une implication :

$$\mathcal{I}(\nu, \mu) = \inf_{x \in \mathcal{S}} Imp[\nu(x), \mu(x)]. \quad (5.33)$$

L'implication définie par [128] $Imp(a, b) = T[c(a), b]$ permet de retrouver les définitions de la morphologie floue à partir de t-normes et t-conormes. Les implications résiduelles définies par :

$$Imp(a, b) = \sup\{\varepsilon \in [0, 1], t(a, \varepsilon) \leq b\}, \quad (5.34)$$

fournissent d'autres formes d'érosion et de dilatation, fondées sur l'inclusion floue suivante :

$$\mathcal{I}(\nu, \mu) = \inf_{x \in \mathcal{S}} \sup\{\varepsilon \in [0, 1], t(\nu(x), \varepsilon) \leq \mu(x)\}. \quad (5.35)$$

L'intersection entre les deux types de formes est obtenue pour des t-normes archimédiennes ayant des éléments nilpotents, typiquement celle de Lukasiewicz. Des opérations de morphologie mathématique floue ont été construites selon ce principe dans [97]. Un de leurs avantages essentiels est qu'elles conduisent à des ouvertures et fermetures idempotentes.

5.3.2 Propriétés

Le détail des propriétés de ces diverses définitions peut être trouvé dans [57], et les preuves dans [28]. Ici, nous résumons simplement les principales propriétés. Les définitions à partir de t-normes et de t-conormes satisfont le plus de propriétés :

- l'érosion et la dilation (respectivement l'ouverture et la fermeture) sont duales par rapport à la complémentation c ;

- si l'élément structurant est binaire, on retrouve les définitions de la morphologie mathématique classique ;
- les opérations sont compatibles avec les translations ;
- elles satisfont la propriété de connaissance locale (la valeur du résultat d'une opération en un point ne dépend que d'un voisinage autour de ce point, dépendant du support de l'élément structurant) ;
- continuité si la t-norme est continue (ce qui est le cas le plus fréquent) ;
- croissance de toutes les opérations par rapport à l'inclusion ;
- extensivité de la dilatation et anti-extensivité de l'érosion si et seulement si $\nu(0) = 1$ (cette contrainte correspond à la condition que l'origine doit appartenir à l'élément structurant dans le cas binaire) ;
- extensivité de la fermeture, anti-extensivité de l'ouverture et idempotence de ces deux opérations si et seulement si $t[b, u(c(b), a)] \leq a$, ce qui est satisfait pour la t-norme et la t-conorme de Lukasiewicz ;
- la dilatation commute avec la réunion floue, et l'érosion avec l'intersection ;
- propriété d'itérativité de la dilatation (sorte d'associativité).

Les définitions obtenues par intégration sur les α -coupes satisfont moins de propriétés : l'extensivité de la fermeture et l'anti-extensivité de l'ouverture, l'idempotence de ces opérations, la commutativité de la dilatation avec la réunion et de l'érosion avec l'intersection et la propriété d'itérativité sont perdues. Les autres propriétés sont conservées.

Pour les définitions obtenues à partir de t-normes et de t-conormes faibles, la plupart des propriétés ne sont satisfaites que dans le cas où $\lambda = \lambda_0$, pour lequel on retrouve une t-norme et une t-conorme particulières.

Avec les définitions obtenues à partir d'une implication résiduelle, l'ouverture et la fermeture sont toujours idempotentes, l'ouverture est anti-extensive et la fermeture extensive.

5.3.3 Exemples

Quelques définitions d'érosions et dilatations floues sont illustrées sur la figure 5.7, sur des exemples synthétiques bidimensionnels.

Donnons maintenant une application réelle, où la dilatation floue est utilisée pour modéliser des imprécisions spatiales. Le problème est celui de la reconstruction 3D de vaisseaux sanguins par fusion de données angiographiques et échographiques [61, 268]. Le but de la fusion est d'éviter les limites inhérentes à chaque modalité : la reconstruction 3D à partir de deux projections angiographiques est imprécise à cause des hypothèses qu'elle nécessite, et la reconstruction à partir de coupes d'échographie endovasculaire est limitée par des facteurs physiques et nécessite également des hypothèses simplificatrices qui introduisent des imprécisions. La figure 5.8 illustre un exemple de données.

La fusion des deux modalités nécessite de recaler toutes les données dans un même repère. Par exemple, la position de chaque coupe échographique est définie par trois paramètres de translation et trois paramètres de rotation. L'estimation de ces paramètres est effectuée sur des radiographies de contrôle. La reconstruction des deux modalités dans le repère commun conduit à des contradictions lors de leur fusion, qui proviennent des imprécisions sur l'estimation de tous les paramètres. Au lieu de les considérer comme des nombres, nous les modélisons comme des nombres flous. Les imprécisions sur les paramètres en translation et en rotation sont finalement représentés par des volumes flous, qui définissent des éléments

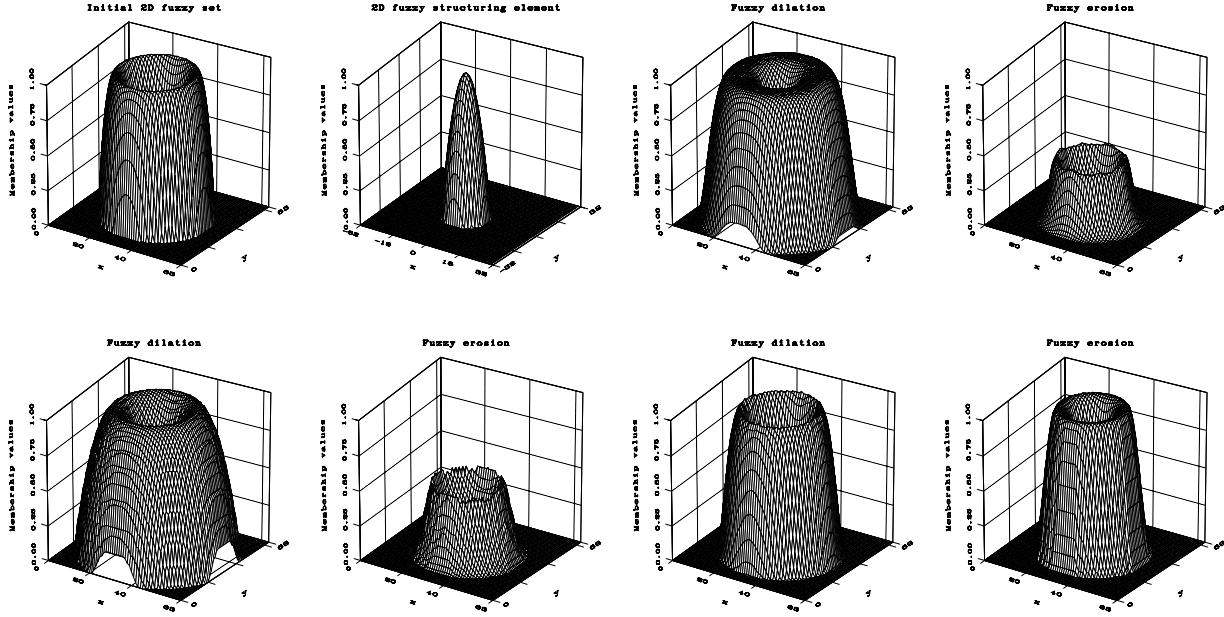


FIG. 5.7 – Illustrations de quelques définitions de dilatation et érosion floues sur un exemple à deux dimensions. Première ligne : ensemble flou et élément structurant définis sur \mathbb{R}^2 , dilatation et érosion par intégration sur les α -coupes ; deuxième ligne : dilatation et érosion avec la t-norme min et la t-conorme max, puis avec les opérateurs de Lukasiewicz.

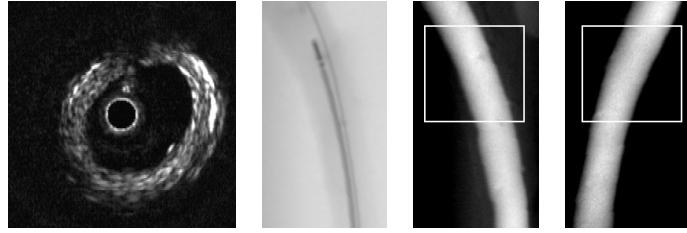


FIG. 5.8 – Images utilisées pour la reconstruction. Coupe échographique, radiographie de contrôle (permettant de localiser la position et l'orientation des échographies), angiographies gauche et droite. Le cadre indique la zone dans laquelle la reconstruction sera effectuée.

structurants flous, autour de chaque point de la surface du vaisseau. La dilatation floue est un bon outil pour introduire ces imprécisions de manière contrôlée, et en préservant de bonnes propriétés. Le fait que la dilatation commute avec la réunion permet de traiter globalement la surface du vaisseau considérée comme un ensemble de points, et la propriété d'itération permet d'introduire successivement les imprécisions en rotation et en translation. Si on appelle V_{bin} la surface binaire (sans imprécisions), et ν_1^x l'imprécision en rotation à la position x (ces imprécisions varient en effet suivant la position), un premier volume flou est obtenu par :

$$\mu_{V'_f}(x) = \sup\{\nu_1^y(x) \mid y \in V_{bin}\}.$$

Puis dans une deuxième étape, une dilatation par l'élément structurant ν_2 prenant en compte les imprécisions en translation (indépendant de la position cette fois) est effectuée :

$$V_f = \bigcup \{D_{D_{\nu_2}(\nu_1^x)}(\{x\}) \mid x \in V_{bin}\} = D_{\nu_2}(V'_f).$$

La figure 5.9 représente une coupe axiale du vaisseau reconstruit aux différentes étapes de la méthode (la méthode est en fait appliquée en 3D directement).

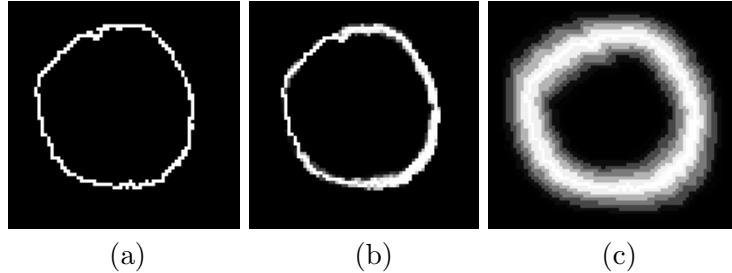


FIG. 5.9 – Coupe échographique axiale du vaisseau reconstruit. (a) Reconstruction binaire. (b) Reconstruction floue incluant les imprécisions en rotation. (c) Reconstruction floue après dilatation par ν_2 (cette reconstruction inclut toutes les imprécisions).

Une fois que cette procédure est appliquée sur les deux modalités, les deux reconstructions contiennent explicitement toutes les imprécisions et ne sont plus contradictoires. Une fusion conjonctive, typiquement par un min, peut alors être effectuée (voir chapitre 8), afin de trouver un consensus entre les deux modalités et de réduire les imprécisions. La décision binaire finale est obtenue par surface de partage des eaux (extension à 3D de la ligne de partage des eaux [294]), qui permet d'obtenir une surface ayant la topologie souhaitée (ici celle d'un cylindre) et passant pas les points d'appartenance maximale au volume fusionné. Le résultat est illustré sur la figure 5.10. Les contradictions entre les deux modalités sont résolues de manière satisfaisante : dans les régions où les deux reconstructions binaires étaient consensuelles, le même résultat est obtenu. Là où elles étaient en conflit, une position intermédiaire est obtenue.

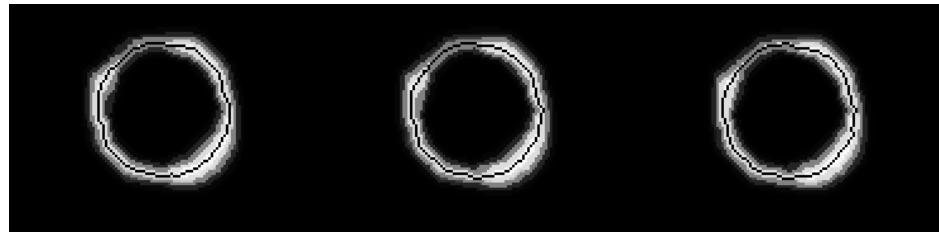


FIG. 5.10 – Superposition du volume flou après fusion et de la surface de partage des eaux (en noir) sur quelques coupes du segment vasculaire.

5.3.4 Conclusion et extensions

À partir des transformations de base présentées ici, beaucoup d'autres peuvent être construites, comme des transformations géodésiques [46], des filtres [311], des opérations par reconstruction, des mesures de convexité [272], etc. Des applications variées commencent à voir le jour, soit en traitement d'images, soit dans d'autres domaines. Des applications en analyse de données ont par exemple été développées dans [337]. Les propriétés de dualité des opérations morphologiques leur permettent d'être utilisées pour définir des opérations duales dans d'autres domaines, par exemple dans la théorie des fonctions de croyance [38],

dans la théorie des ensembles approximatifs (« rough sets ») [47], ou en logique [52, 49]. En traitement d’images et reconnaissance des formes structurelles, la morphologie floue fournit également des outils puissants pour définir des relations spatiales dans des situations imprécises [45], comme nous le verrons dans les sections suivantes.

5.4 Topologie floue

Cette partie est dédiée à l’extension au flou de notions de topologie simples, telles que la connexité ou l’adjacence, qui sont très importantes en traitement d’images, à la fois localement au voisinage d’un point, ou plus globalement quand il s’agit de relations spatiales entre objets.

5.4.1 Connexité floue et voisinage

La notion de connexité est une notion de base indispensable en traitement d’images, impliquée dans nombre de transformations, de mesures (qu’elles soient topologiques, géométriques ou morphologiques), de méthodes d’analyse et de reconnaissance, dont la généralisation au flou nécessite donc de définir une connexité floue. De plus, c’est un outil important dès qu’il s’agit d’aller d’un endroit à l’autre dans des ensembles flous, ce qui peut être exploité pour définir des distances géodésiques floues [32] et une morphologie mathématique floue géodésique [46].

Plusieurs définitions de la connexité floue ont été proposées dans la littérature, qui se différencient essentiellement par leur comportement vis-à-vis des plateaux de la fonction d’appartenance [283, 338, 339]. Nous retenons ici celle de Rosenfeld [283], qui est la plus utilisée, et surtout qui a les meilleures propriétés du point de vue de la reconnaissance des formes. On suppose que l’on travaille dans un espace \mathcal{S} discret et borné (typiquement le support d’une image), sur lequel une connexité discrète est définie. Rosenfeld définit le degré de connexité entre deux points x et y d’un ensemble flou μ par :

$$c_\mu(x, y) = \max_{L_{x,y}} \left[\min_{1 \leq i \leq n} \mu(x_i) \right], \quad (5.36)$$

où $L_{x,y} = x_1 \dots x_n$ est un chemin de $x_1 = x$ à $x_n = y$ dans \mathcal{S} , au sens de la connexité de base sur \mathcal{S} . Le degré de connexité prend ses valeurs dans $[0,1]$. En considérant les degrés d’appartenance comme un relief sur \mathcal{S} , $c_\mu(x, y)$ correspond à l’altitude minimale sur le chemin qui va de x à y en descendant le moins possible. On en déduit la notion de composante connexe floue associée à un point x comme l’ensemble flou de fonction d’appartenance :

$$\forall y \in \mathcal{S}, \Gamma_\mu^x(y) = c_\mu(x, y). \quad (5.37)$$

Le degré de connexité flou a les propriétés suivantes :

1. c_μ est faiblement réflexif : $\forall (x, y) \in \mathcal{S}^2, c_\mu(x, x) \geq c_\mu(x, y)$ et $c_\mu(x, x) = \mu(x)$,
2. c_μ est symétrique,
3. c_μ est transitif, au sens de la max-min transitivité :
 $\forall (x, z) \in \mathcal{S}^2, c_\mu(x, z) \geq \max_{y \in E} [\min(c_\mu(x, y), c_\mu(y, z))]$,
4. $c_\mu(x, y)$ est obtenu en descendant le moins possible dans les valeurs d’appartenance pour aller de x à y , et est lié aux valeurs d’appartenance de x et y , en particulier :
 $c_\mu(x, y) \leq \min[\mu(x), \mu(y)]$.

Toujours en considérant la fonction d'appartenance μ comme un relief, des notions topologiques peuvent être exprimées en termes topographiques. Ainsi, un plateau est défini comme un sous-ensemble Π de \mathcal{S} tel que [282] : Π est connexe (toujours au sens de la connexité discrète définie dans \mathcal{S}), la fonction d'appartenance est constante sur le plateau, et tout voisin d'un point du plateau a un degré d'appartenance différent de la valeur du plateau. Π est appelé « top » si sa valeur est un maximum régional, et « bottom » si sa valeur est un minimum régional. L'ensemble flou μ est connexe (au sens de la connexité floue) si et seulement si il a un et un seul plateau maximum régional. À partir de ces notions, Rosenfeld définit des points simples dans un ensemble flou : un point x de \mathcal{S} est simple si et seulement si il est simple dans l' α -coupe de niveau $\alpha = \mu(x)$.

Dans les notions précédentes, le voisinage utilisé est toujours celui de la topologie discrète associée à \mathcal{S} et à sa connexité élémentaire, donc un voisinage classique. On peut également définir des voisinages flous, et définir un degré n_{xy} avec lequel deux points x et y de \mathcal{S} sont voisins. Plusieurs définitions ont été proposées [104, 60], qui sont typiquement des fonctions décroissantes de la distance entre les deux points :

$$n_{xy} = \frac{1}{1 + d(x, y)}, \text{ ou } n_{xy} = \frac{1 + \exp(-b)}{1 + \exp b(\frac{d(x, y)-1}{S} - 1)}, \quad (5.38)$$

où b et S sont deux paramètres positifs qui contrôlent l'aplatissement, la pente et la largeur de la courbe. D'autres fonctions peuvent être utilisées, comme des fonctions en S par exemple.

5.4.2 Frontière d'un objet flou

La frontière d'un ensemble binaire est définie comme l'ensemble des points limites de l'objet (frontière interne) ou limites du complémentaire de l'objet (frontière externe). Si un objet est flou, il est naturel de considérer que sa frontière est également un ensemble flou. Sa définition peut être obtenue simplement en traduisant l'expression « un point appartient à la frontière (interne) si et seulement si il appartient à l'objet et il a au moins un voisin à l'extérieur de l'objet ». Ainsi, la frontière interne b_μ^i d'un objet flou μ est définie par [60] :

$$\forall x \in \mathcal{S}, b_\mu^i(x) = t[\mu(x), \sup_{z \in \mathcal{S}} t[c(\mu)(z), n_{xz}]], \quad (5.39)$$

où t est une t-norme, c une complémentation floue, et n_{xz} le degré de voisinage entre x et z . Des notions similaires de frontière entre deux objets flous ont été proposées dans [104]. Cette définition de la frontière floue peut également être exprimée en termes morphologiques, puisque les points de la frontière sont les points de l'ensemble qui ne sont pas dans son érodé (ou de manière équivalente qui sont dans le dilaté de son complémentaire). Cette constatation conduit à la définition suivante dans le cas flou :

$$b_\mu^i(x) = t[\mu(x), D_{B_c}(c(\mu))(x)], \quad (5.40)$$

où B_c est un élément structurant élémentaire, relatif soit à la connexité de base sur \mathcal{S} , soit à un voisinage flou.

De manière similaire, la frontière externe peut être définie par :

$$\forall x \in \mathcal{S}, b_\mu^e(x) = t[c(\mu(x)), \sup_{z \in \mathcal{S}} t[\mu(z), n_{xz}]], \quad (5.41)$$

ou en termes morphologiques par : $b_\mu^e(x) = t[D_{B_c}(\mu)(x), c(\mu)(x)]$.

5.4.3 Adjacence entre deux objets flous

Assez peu de travaux dans la littérature sont consacrés à la définition de l'adjacence entre deux ensembles flous. Une première approche consiste à définir l'adjacence entre deux ensembles binaires à partir de la notion de visibilité d'un ensemble depuis l'autre, puis à étendre cette notion au cas flou [286], mais présente l'inconvénient de ne pas être symétrique (la notion de visibilité ne l'étant pas). Une deuxième approche consiste à s'appuyer sur la notion de frontière des objets ou entre les objets [104, 60]. Nous donnons ici des exemples relevant de cette deuxième approche, en se restreignant au cas discret, qui est le plus directement utilisable en traitement d'images. On suppose donc que \mathcal{S} est muni d'une connexité discrète c , définie par la variable booléenne $n_c(x, y)$.

Rappelons d'abord les définitions de l'adjacence dans le cas binaire. Soient X et Y deux sous-ensembles de \mathcal{S} . X et Y sont adjacents au sens de la c -connexité si :

$$X \cap Y = \emptyset \text{ et } \exists x \in X, \exists y \in Y : n_c(x, y). \quad (5.42)$$

La frontière interne de X étant notée $\partial X = X - E_{B_c}(X)$ (où B_c est l'élément structurant élémentaire de la c -connexité), une conséquence de l'adjacence est que les points x et y qui satisfont $n_c(x, y)$ appartiennent nécessairement à ∂X et ∂Y respectivement. La définition de l'adjacence peut également être exprimée en termes morphologiques, de manière équivalente, par :

$$X \cap Y = \emptyset \text{ et } D_{B_c}(X) \cap Y \neq \emptyset, \quad D_{B_c}(Y) \cap X \neq \emptyset. \quad (5.43)$$

Notons que pour éviter les paradoxes topologiques dus à l'utilisation de la même connexité pour tous les objets dans certaines topologies discrètes, il peut être préférable de séparer les objets et leur frontière (et de les considérer avec des connexités différentes). Cela conduit à des définitions de l'adjacence un peu différentes, dont l'extension au flou suit les mêmes principes que ceux présentés ci-dessous. Elles sont détaillées dans [60] mais ne sont pas reprises ici.

L'extension de l'adjacence au cas flou peut donc être effectuée de trois manières différentes : soit en s'appuyant uniquement sur la contrainte de connexité, soit en utilisant également la contrainte sur la frontière, soit en s'appuyant sur la dilatation floue. Le principe de construction décrit ici suit la méthode de traduction formelle déjà utilisée pour les transformations de morphologie mathématique (paragraphe 5.3), et utilise les définitions de degrés d'intersection décrites dans le paragraphe 5.2.2, de voisinage flou (paragraphe 5.4.1) et de frontière d'objets flous (paragraphe 5.4.2).

Utilisation de la seule contrainte de connexité. Généraliser la définition de l'adjacence au flou revient à effectuer une conjonction entre un degré de non intersection entre les deux ensembles flous et un degré d'existence de deux points voisins. La conjonction est exprimée par une t-norme t , la non intersection par la complémentation c d'un degré d'intersection μ_{int} , l'existence par un sup. On aboutit ainsi à la définition suivante du degré d'adjacence entre deux sous-ensembles flous μ et ν de \mathcal{S} :

$$\mu_{adj}(\mu, \nu) = t[c(\mu_{int})(\mu, \nu), \sup_{x \in \mathcal{S}} \sup_{y \in \mathcal{S}} t[\mu(x), \nu(y), n_{xy}]]. \quad (5.44)$$

Cette définition est symétrique, coïncide avec la définition binaire si μ et ν sont binaires et si $n_{xy} = n_c(x, y)$, et est invariante par transformations géométriques (pour les homothéties, cela suppose que n_{xy} soit aussi invariant).

La figure 5.11 illustre le résultat obtenu avec cette définition dans le cas 1D. En utilisant comme degré d'intersection la hauteur de l'intersection (équation 5.1), on obtient $\mu_{adj}(\mu, \nu) = \alpha_1 = 0,36$ et $\mu_{adj}(\mu, \nu') = \alpha_3 = 0,35$, donc des valeurs très proches. En utilisant le volume de l'intersection (équation 5.3), la définition rend mieux compte de la différence de recouvrement des ensembles flous, et on obtient $\mu_{adj}(\mu, \nu) = \alpha_2 = 0,67$ et $\mu_{adj}(\mu, \nu') = \alpha_4 = 0,34$, ce qui correspond mieux à l'intuition.

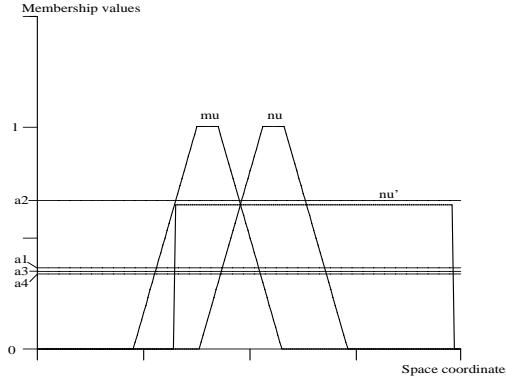


FIG. 5.11 – Illustration de l'adjacence floue, en utilisant différentes définitions du degré d'intersection.

La table 5.1 donne les degrés d'adjacence flous obtenus entre quelques structures de la figure 5.1. Les résultats sont en accord avec ce qu'on peut attendre, d'après un modèle du cerveau, donné par exemple par un atlas anatomique. Ces résultats montrent que l'adjacence floue est une des relations spatiales qui peut être utilisée dans un système de reconnaissance structurelle.

Objet flou 1	Objet flou 2	degré d'adjacence	adjacence dans le modèle (binaire)
v1	v2	0,368	1
v1	nc1	0,463	1
v1	t1	0,000	0
v1	nc2	0,035	0
v2	nc2	0,427	1
nc1	t1	0,035	0

TAB. 5.1 – Degrés d'adjacence flous entre quelques structures du cerveau. Les degrés les plus élevés sont obtenus entre les structures pour lesquelles l'adjacence est attendue, et des degrés très faibles correspondent à des structures non adjacentes.

Introduction des contraintes sur la frontière. Si l'on rajoute la contrainte sur la frontière, la définition de l'adjacence floue devient :

$$\mu_{adj}(\mu, \nu) = t[c(\mu_{int})(\mu, \nu), \sup_{x \in \mathcal{S}} \sup_{y \in \mathcal{S}} t[b_\mu^i(x), b_\nu^i(y), n_{xy}]], \quad (5.45)$$

obtenue simplement en remplaçant l'appartenance de x et y à μ et ν par leur appartenance aux frontières internes des deux ensembles flous. Cette définition a exactement les mêmes propriétés que la précédente.

Expression morphologique. Selon le même principe, la définition binaire de l'adjacence exprimée à partir de la dilatation se traduit par :

$$\mu_{adj}(\mu, \nu) = t[c(\mu_{int})(\mu, \nu), \mu_{int}(D_{B_c}(\mu), \nu), \mu_{int}(D_{B_c}(\nu), \mu)]. \quad (5.46)$$

Cette expression représente la conjonction entre un degré de non intersection entre μ et ν et un degré d'intersection entre un ensemble et le dilaté de l'autre ensemble. L'élément structurant B_c peut être choisi en fonction de la connexité, ou comme un ensemble flou, représentant par exemple l'imprécision spatiale (la distribution de possibilité de la position de chaque point). Les mêmes propriétés que celles des définitions précédentes sont vérifiées.

5.5 Distances

L'intérêt des distances en traitement d'images est multiple et couvre plusieurs aspects : distance d'un point à un ensemble, distance entre objets, distance entre deux points dans un ensemble (au sens géodésique). Le deuxième type de distance est largement traité dans la littérature sur les ensembles flous, avec une grande variété d'approches dont seulement certaines sont bien adaptées au traitement d'images. Les deux autres le sont beaucoup moins. Les distances d'un point à un ensemble sont utiles en classification, dans les méthodes de regroupement visant à former des régions spatialement cohérentes, en morphologie mathématique, dans des problèmes de recalage, et sont également impliquées dans certaines définitions du squelette. Les distances entre deux ensembles apportent une information structurelle sur l'image, en décrivant l'arrangement spatial métrique des objets. Les distances entre deux points d'un ensemble, ou distances géodésiques, sont les moins abordées dans le cas flou, mais sont utiles lorsque les opérations doivent être conditionnées à un ensemble. Avant d'aborder successivement ces trois types de distances, nous commençons par décrire les diverses représentations possibles de distances.

5.5.1 Représentations

Alors que dans le cas d'ensembles binaires, une distance est toujours représentée par un nombre de \mathbb{R}^+ , dans le cas flou, plusieurs représentations sont possibles. La représentation sous forme de nombre est la plus courante. Les valeurs sont prises soit dans \mathbb{R}^+ , soit seulement dans $[0, 1]$ lorsque les distances sont définies comme des dissimilarités floues par exemple.

Si les objets sont imprécis, on peut également considérer que les distances qui les séparent sont imprécises. Ainsi une distance peut être représentée sous forme de nombre flou [122, 284]. Dans [284], deux concepts sont introduits. Le premier est celui de densité de distance, notée $\delta(\mu, \nu)$, et le second est celui de distribution de distance, notée $\Delta(\mu, \nu)$, les deux étant des ensembles flous sur \mathbb{R}^+ . Les deux fonctions sont reliées par la relation :

$$\Delta(\mu, \nu)(n) = \int_0^n \delta(\mu, \nu)(n')dn'. \quad (5.47)$$

La valeur prise en n par la distribution de distance $\Delta(\mu, \nu)(n)$ représente le degré avec lequel la distance entre μ et ν est inférieure ou égale que n , alors que la valeur en n de la densité de distance $\delta(\mu, \nu)(n)$ représente le degré avec lequel la distance est égale à n .

Enfin, la distance peut être représentée par une variable linguistique. Cela suppose que l'espace des valeurs est quantifié, avec une granularité plus ou moins fine, en classes symboliques telles que « proche », « loin », etc. [22, 202, 27], dont la sémantique est définie par des ensembles flous de \mathbb{R}^+ .

Du point de vue des propriétés, une distance est rigoureusement une fonction positive, séparable, symétrique, et satisfaisant l'inégalité triangulaire. En traitement d'images, toutes ces propriétés ne sont pas forcément nécessaires. Par exemple pour la comparaison de formes au moyen de distances, ce sont surtout les propriétés de fonctions de similarité qui sont importantes. La propriété de l'inégalité triangulaire n'est pas toujours indispensable, par exemple lorsque l'on compare un objet à un modèle, et on peut alors se restreindre à des semi-métriques ou même des semi-pseudométriques. En classification, on peut exploiter des fonctions ne vérifiant pas la séparabilité pour regrouper des objets séparés par une distance nulle, et ce sont alors les propriétés d'indistingualité qui sont importantes. Pour simplifier, nous parlerons toujours de distance, même dans le cas de notions plus faibles.

5.5.2 Distance d'un point à un objet flou

Nous nous posons dans cette partie une question qui n'a été que très peu traitée dans la littérature : quelle est la distance d'un point x de \mathcal{S} à un ensemble flou μ défini sur \mathcal{S} ? Dans le cas binaire, la distance d'un point à un ensemble (compact) est classiquement définie comme la distance minimale entre x et les points de l'ensemble. Il est souhaitable que la distance floue correspondante dépende de la distance euclidienne d_E définie sur \mathcal{S} , afin de prendre en compte l'information spatiale dans le calcul de la distance. Des définitions de telles distances ont été proposées dans [31, 43].

Une première solution consiste à pondérer la distance entre deux points, de telle sorte que les points d'appartenance faible aient une importance moins grande dans le calcul du minimum, par exemple :

$$d(x, \mu) = \min_{y \in \mathcal{S}} [d_E(x, y)f(\mu(y))], \quad (5.48)$$

où $f(t)$ est une fonction décroissante de t .

Une deuxième solution consiste à intégrer les valeurs des distances de x aux α -coupes de μ :

$$d(x, \mu) = \int_0^1 \min_{y \in \mu_\alpha} d_E(x, y)d\alpha, \quad (5.49)$$

ou à pondérer ces distances par α :

$$d(x, \mu) = \sup_{\alpha \in]0,1]} [\alpha d_B(x, \mu_\alpha)]. \quad (5.50)$$

Cette deuxième forme a de moins bonnes propriétés, puisque la distance obtenue est toujours celle de x au noyau de μ .

Une autre approche consiste à définir une distance floue (qui ne soit pas seulement un scalaire), à partir de la dilatation floue. L'avantage de cette approche est qu'elle utilise une

formulation ensembliste (et non analytique) de la distance, qui est plus facile à transposer au flou par un simple principe de traduction terme à terme.

Dans le cas binaire et dans un espace discret, la distance d'un point x à un ensemble X est donnée en fonction de la dilatation par :

$$d_B(x, X) = 0 \Leftrightarrow x \in X \quad (5.51)$$

$$d_B(x, X) = n \Leftrightarrow x \in D^n(X) \text{ et } \forall n' < n, x \notin D^{n'}(X) \quad (5.52)$$

où D^n est la dilatation par une boule de rayon n centrée à l'origine de \mathcal{S} (et $D^0(X) = X$) (voir par exemple [64] pour une étude des boules discrètes et des distances discrètes dans le cas binaire). Pour ces éléments structurants, la dilatation est extensive, et $\forall n' < n, x \notin D^{n'}(X)$ est équivalent à $x \notin D^{n-1}(X)$. L'équation 5.52 est donc équivalente à :

$$x \in D^n(X) \cap [D^{n-1}(X)]^C, \quad (5.53)$$

où A^C est le complémentaire de A dans \mathcal{S} . Cette expression purement ensembliste peut maintenant être traduite en termes flous, ce qui conduit à la définition suivante du degré $\delta_{(x,\mu)}(n)$ avec lequel $d(x, \mu)$ est égal à n :

$$\delta_{(x,\mu)}(0) = \mu(x), \quad (5.54)$$

$$\delta_{(x,\mu)}(n) = t[D_\nu^n(\mu)(x), c[D_\nu^{n-1}(\mu)(x)]], \quad (5.55)$$

où t est une t-norme, c une complémentation floue, et ν un élément structurant flou, pour lequel plusieurs choix sont possibles. Ce peut être simplement la boule unité, ou un ensemble flou représentant la plus petite entité spatiale significative dans l'image avec son imprécision. Dans ce cas, ν doit valoir 1 à l'origine, afin de garder la propriété d'extensivité [57].

Les propriétés de cette définition sont les suivantes [31] : si $\mu(x) = 1$, $\delta_{(x,\mu)}(0) = 1$ et $\forall n > 0, \delta_{(x,\mu)}(n) = 0$, et donc la distance est un nombre classique. Si μ et ν sont binaires, la définition coïncide avec la forme binaire. L'ensemble flou $\delta_{(x,\mu)}$ peut être interprété comme une densité de distance, dont on peut déduire une distribution de distance par intégration. Enfin, $\delta_{(x,\mu)}$ est un nombre flou non normalisé (dans le cas discret fini).

La figure 5.12 présente un exemple de nombre flou $\delta_{(x,\mu)}(n)$ pour deux points et un ensemble flou en dimension 1.

5.5.3 Distance entre deux objets flous

Plusieurs distances entre ensembles flous ont été proposées dans la littérature. Nous en donnons une présentation qui, bien que s'inspirant directement de celle proposée dans [367], est adaptée aux intérêts du traitement d'images et repose sur les types d'informations pris en compte dans les différentes définitions. Une première classe d'approches ne concerne que la comparaison de fonctions d'appartenance, tandis qu'une seconde classe introduit également des distances dans le domaine spatial. Dans la suite, on note μ et ν les fonctions d'appartenance de deux ensembles flous définis sur l'espace \mathcal{S} (considéré discret dans tout ce qui suit). On note $d(\mu, \nu)$ la distance entre ces deux ensembles flous et $d_E(x, y)$ la distance spatiale euclidienne entre deux points x et y de \mathcal{S} (indépendamment de leurs degrés d'appartenance aux ensembles flous considérés). Les différentes formes de distances ainsi que leurs propriétés sont détaillées dans [43].

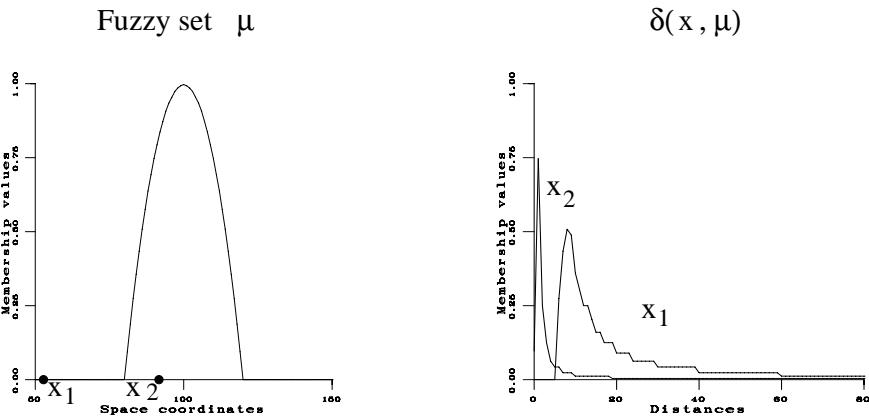


FIG. 5.12 – Nombres flous représentant $\delta_{(x,\mu)}$ (μ est représenté à gauche) pour deux points x différents.

Comparaison de fonctions d'appartenance. On trouve dans la première classe, ne comparant que les fonctions d'appartenance, quatre approches différentes : fonctionnelle, reposant sur la théorie de l'information, ensembliste, et s'inspirant de la reconnaissance des formes.

L'approche fonctionnelle est certainement la plus répandue. Elle est fondée sur le choix d'une norme L_p entre μ et ν , conduisant à la définition générique suivante :

$$d_p(\mu, \nu) = \left[\sum_{x \in S} |\mu(x) - \nu(x)|^p \right]^{1/p}, \quad d_\infty(\mu, \nu) = \max_{x \in S} |\mu(x) - \nu(x)|. \quad (5.56)$$

D'autres définitions peuvent être déduites de la notion de similarité floue et de celle de t -indistingualibilité. Étant donnée une telle fonction de similarité s , la fonction $1 - s$ est une ultra-métrique. Ces notions connaissent un grand développement en intelligence artificielle (par exemple pour le raisonnement par analogie) (voir par exemple [66] pour une revue complète de mesures de comparaison). En traitement d'images, elles ont été utilisées par exemple selon une approche pyramidale dans [216].

Les premières définitions fondées sur **la théorie de l'information** reposent sur la définition de l'entropie floue $\mathcal{E}(\mu)$, dont une pseudo-distance est déduite par [219] : $d(\mu, \nu) = |\mathcal{E}(\mu) - \mathcal{E}(\nu)|$. Cette définition ne satisfait pas la condition de séparabilité. Cette propriété peut être obtenue en travaillant dans l'espace quotient obtenu par la relation d'équivalence $\mu \sim \nu \Leftrightarrow \mathcal{E}(\mu) = \mathcal{E}(\nu)$. Cependant une telle démarche n'est pas bien adaptée au traitement d'images. En effet, puisque l'entropie d'un ensemble précis (binaire) est toujours égale à zéro, deux structures binaires dans une image appartiennent à la même classe d'équivalence, même si elles sont complètement différentes et quelle que soit leur distance dans S . Des mesures d'entropie conditionnelles à une mesure de similarité ont été proposées, qui combinent cette approche avec l'approche fonctionnelle [352]. Elles ont été appliquées dans des problèmes de décision (en particulier pour l'étude de questionnaires), mais, à notre connaissance, pas en traitement d'images. Dans le même ordre d'idées, la notion de divergence floue [26] (qui peut être interprétée comme une distance) a été introduite comme un concept inspiré de celui d'entropie de Kullback [203] :

$$d(\mu, \nu) = \frac{1}{|S|} \sum_{x \in S} [D_x(\mu, \nu) + D_x(\nu, \mu)] \quad (5.57)$$

avec :

$$D_x(\mu, \nu) = \mu(x) \log \frac{\mu(x)}{\nu(x)} + (1 - \mu(x)) \log \frac{1 - \mu(x)}{1 - \nu(x)}. \quad (5.58)$$

Cette distance est positive, symétrique, satisfait la propriété de séparabilité, mais pas l'inégalité triangulaire.

Dans **l'approche ensembliste**, une distance entre ensembles flous est considérée comme une fonction de dissimilarité entre ensembles, exprimée à partir de réunions et d'intersections floues. On en trouve quelques exemples dans [367]. L'idée qui préside à la construction de ces fonctions est que la distance doit être plus importante si les ensembles s'intersectent faiblement. Mentionnons les exemples suivants :

$$d(\mu, \nu) = 1 - \frac{\sum_{x \in \mathcal{S}} \min[\mu(x), \nu(x)]}{\sum_{x \in \mathcal{S}} \max[\mu(x), \nu(x)]}, \quad (5.59)$$

$$d(\mu, \nu) = 1 - \max_{x \in \mathcal{S}} \min[\mu(x), \nu(x)], \quad (5.60)$$

$$d(\mu, \nu) = \max_{x \in \mathcal{S}} \max[\min(\mu(x), 1 - \nu(x)), \min(1 - \mu(x), \nu(x))]. \quad (5.61)$$

Notons que les deux dernières définitions sont respectivement équivalentes aux fonctions $1 - \Pi(\mu; \nu)$ et $1 - \max[N(\mu; \nu), N(\nu; \mu)]$ (où Π et N sont des fonctions de possibilité et de nécessité) qui sont utilisées en marquage adapté flou (« fuzzy pattern matching ») [132]. Elles ont un vaste domaine d'application, y compris en traitement d'images. Elles sont de plus reliées à la morphologie mathématique floue, puisque $\Pi(\mu; \nu)$ correspond à la valeur de la dilatation de μ par ν à l'origine, alors que $N(\mu; \nu)$ correspond à la valeur de l'érosion de μ par ν à l'origine.

L'approche de reconnaissance des formes consiste à représenter dans un premier temps chaque ensemble flou dans un espace de paramètres flous (cardinal, moments, obliquité par exemple) et à calculer dans un deuxième temps des distances entre vecteurs de paramètres. Cette approche peut exploiter les avantages de certaines des définitions précédentes, par exemple en incluant des mesures d'entropie ou de similarité dans l'ensemble de paramètres.

Combinaison de comparaisons des fonctions d'appartenance et d'informations spatiales. Dans la seconde classe de méthodes, on cherche à inclure la distance spatiale d_E dans la définition de la distance entre μ et ν . On trouve dans cette deuxième classe de distances des approches géométriques, morphologiques, et reposant sur la théorie des graphes.

L'approche géométrique consiste à généraliser l'une des distances classiquement utilisées entre ensembles binaires, telles que la distance minimale entre points les plus proches, la distance moyenne entre points les plus proches, ou la distance de Hausdorff.

Ces généralisations peuvent suivre quatre principes. Le premier consiste à considérer des ensembles flous définis dans un espace de dimension n comme des ensembles binaires de dimension $n+1$ et utiliser alors des distances classiques binaires dans cet espace de dimension supérieure. Cependant, cette démarche est souvent peu satisfaisante en traitement d'images car les n dimensions de l'espace \mathcal{S} et la dimension des degrés d'appartenance (prenant leurs valeurs dans $[0,1]$) ont des significations complètement différentes, et les traiter de la même manière semble fortement criticable.

Le deuxième principe est un principe de fuzzification : soit D une distance entre ensembles précis, alors son équivalent flou est défini par :

$$d(\mu, \nu) = \int_0^1 D(\mu_\alpha, \nu_\alpha) d\alpha, \quad (5.62)$$

ou par une somme discrète si les fonctions d'appartenance sont constantes par morceaux. Par ce procédé, $d(\mu, \nu)$ hérite des propriétés de la distance binaire choisie.

Le troisième principe consiste à pondérer les distances par les fonctions d'appartenance. Pour la distance moyenne, cela conduit à la définition suivante :

$$d(\mu, \nu) = \frac{\sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} d_E(x, y) \min[\mu(x), \nu(y)]}{\sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} \min[\mu(x), \nu(y)]}. \quad (5.63)$$

Enfin, la dernière approche consiste à définir une distance floue comme un ensemble flou sur \mathbb{R}^+ au lieu d'un nombre précis. Pour la distance minimale entre points les plus proches, cela donne [284] :

$$d(\mu, \nu)(r) = \sup_{x, y, d_E(x, y) \leq r} \min[\mu(x), \nu(y)]. \quad (5.64)$$

Plusieurs généralisations de la distance de Hausdorff ont également été proposées selon le même principe [122].

Ces distances partagent la plupart des avantages et des inconvénients de la distance binaire sous-jacente : les coûts de calcul peuvent être très élevés (c'est déjà le cas pour plusieurs distances binaires) ; de plus, l'interprétation et la robustesse dépendent beaucoup de la distance choisie, par exemple la distance de Hausdorff est très sensible au bruit (alors que la distance moyenne ne l'est pas) mais a des avantages en particulier pour son utilité en morphologie mathématique.

Une **approche par tolérance**, proposée dans [217], combine les fonctions d'appartenance et des distances spatiales locales. Une différence locale entre μ et ν en un point x de \mathcal{S} est définie par :

$$d_x^\tau(\mu, \nu) = \inf_{y, z \in B(x, \tau)} |\mu(y) - \nu(z)|, \quad (5.65)$$

où $B(x, \tau)$ est la boule fermée centrée en x et de rayon τ . Ensuite les fonctions de l'approche fonctionnelle sont redéfinies pour introduire la tolérance spatiale τ . Par exemple, pour la norme L_p , on a :

$$d_p^\tau(\mu, \nu) = [\int_{\mathcal{S}} [d_x^\tau(\mu, \nu)]^p dx]^{1/p}. \quad (5.66)$$

Plusieurs résultats sont démontrés dans [217], en particulier des propriétés de convergence. Cette approche a une interprétation morphologique puisque le voisinage considéré autour de chaque point peut être interprété comme un élément structurant.

Une **approche morphologique**, proposée dans [31, 43], consiste à définir une distance comme un ensemble flou sur \mathbb{R}^+ , fondée sur la dilatation floue. Elle est issue de l'observation que certaines distances binaires peuvent être exprimées en termes de dilatation morphologique [300]. Nous donnons ici les exemples de la distance minimum et de la distance de Hausdorff. Dans le cas binaire, pour $n > 0$, la distance minimum d_N entre deux ensembles X et Y peut être exprimée par :

$$d_N(X, Y) = n \Leftrightarrow D^n(X) \cap Y \neq \emptyset \text{ et } D^{n-1}(X) \cap Y = \emptyset \quad (5.67)$$

et l'expression symétrique. Pour $n = 0$ on a $d_N(X, Y) = 0 \Leftrightarrow X \cap Y \neq \emptyset$. La traduction de ces équivalences pour des ensembles flous μ et μ' donne, pour $n > 0$, la densité de distance suivante :

$$\delta_N(\mu, \mu')(n) = t[\sup_{x \in S} t[\mu'(x), D_\nu^n(\mu)(x)], c[\sup_{x \in S} t[\mu'(x), D_\nu^{n-1}(\mu)(x)]]] \quad (5.68)$$

et $\delta_N(\mu, \mu')(0) = \sup_{x \in S} t[\mu(x), \mu'(x)]$. Cette expression montre comment les valeurs d'appartenance à μ' sont prises en compte, sans utiliser le principe d'extension. ν désigne l'élément structurant utilisé pour la dilatation.

La distance de Hausdorff est définie par :

$$d_H(X, Y) = \max[\sup_{x \in X} d(x, Y), \sup_{y \in Y} d(y, X)]. \quad (5.69)$$

Elle s'exprime en termes morphologiques par :

$$d_H(X, Y) = \inf\{n, X \subseteq D^n(Y) \text{ et } Y \subseteq D^n(X)\}. \quad (5.70)$$

À partir de l'équation 5.70, on peut définir une distribution de distance à partir de la dilatation floue :

$$\Delta_H(\mu, \mu')(n) = t[\inf_{x \in S} T[D_\nu^n(\mu)(x), c(\mu'(x))], \inf_{x \in S} T[D_\nu^n(\mu')(x), c(\mu(x))]], \quad (5.71)$$

où c est une complémentation, t une t-norme et T une t-conorme. Une densité de distance peut être déduite implicitement de cette distribution. Mais on peut aussi obtenir une densité de distance en traduisant les expressions suivantes en termes flous :

$$d_H(X, Y) = 0 \Leftrightarrow X = Y, \quad (5.72)$$

et pour $n > 0$:

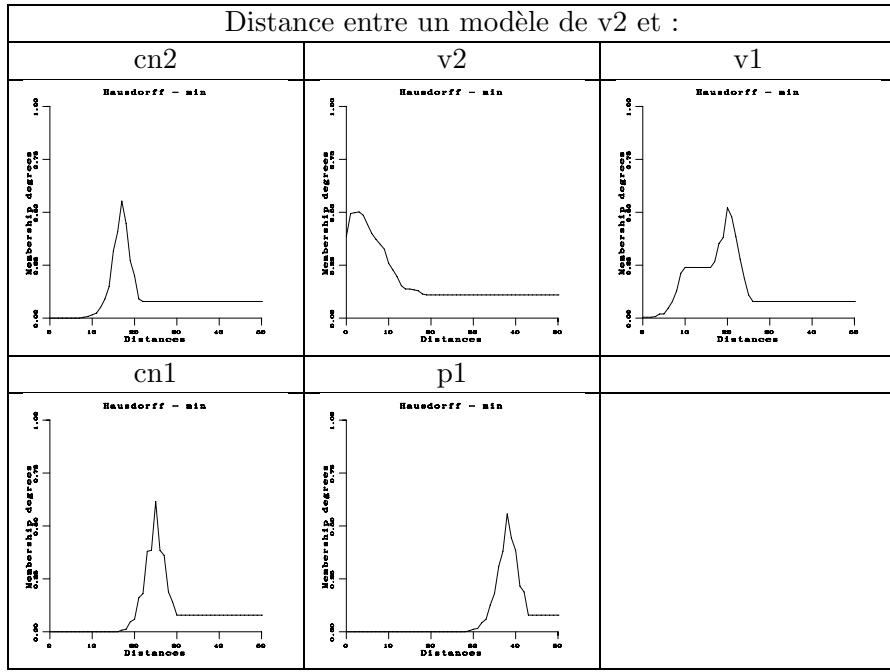
$$\begin{aligned} d_H(X, Y) = n \Leftrightarrow X \subseteq D^n(Y) \text{ et } Y \subseteq D^n(X) \\ \text{et } (X \not\subseteq D^{n-1}(Y) \text{ ou } Y \not\subseteq D^{n-1}(X)). \end{aligned} \quad (5.73)$$

La traduction de ces équations donne le nombre flou suivant :

$$\delta_H(\mu, \mu')(0) = t[\inf_{x \in S} T[\mu(x), c(\mu'(x))], \inf_{x \in S} T[\mu'(x), c(\mu(x))]], \quad (5.74)$$

$$\begin{aligned} \delta_H(\mu, \mu')(n) = t[\inf_{x \in S} T[D_\nu^n(\mu)(x), c(\mu'(x))], \inf_{x \in S} T[D_\nu^n(\mu')(x), c(\mu(x))], \\ T(\sup_{x \in S} t[\mu(x), c(D_\nu^{n-1}(\mu')(x))], \sup_{x \in S} t[\mu'(x), c(D_\nu^{n-1}(\mu)(x))])]. \end{aligned} \quad (5.75)$$

Ces définitions n'ont pas exactement les mêmes propriétés que leurs équivalents binaires. Toutes les distances sont positives, dans le sens où leur support est inclus dans \mathbb{R}^+ . Par construction elles sont symétriques en μ et μ' . La propriété de séparabilité n'est pas toujours satisfaite. Cependant, si μ est normalisé, on a pour la distance minimum $\delta_N(\mu, \mu)(0) = 1$ et $\delta_N(\mu, \mu)(n) = 0$ pour $n > 1$. Pour la distance de Hausdorff, $\delta_H(\mu, \mu')(0) = 1$ implique $\mu = \mu'$ pour la t-conorme T de Lukasiewicz ($T(a, b) = \min(1, a + b)$), alors que cela implique μ et μ' binaires et égaux pour $T = \max$. L'inégalité triangulaire n'est en général pas satisfaite.



TAB. 5.2 – Distances de Hausdorff entre ensembles flous selon l’approche morphologique.

Quelques exemples de densités de distances sont illustrés dans la table 5.2 (nombres flous obtenus par l’approche morphologique pour la distance de Hausdorff et la t-norme min). Les objets flous de départ sont quelques structures du cerveau (figure 5.1) qui sont comparées à un modèle de v2. Les valeurs non nulles obtenues pour v2 montrent les différences entre la structure dans l’image et le modèle. Les autres résultats correspondent bien au fait que v2 est proche de cn2 et v1, assez loin de cn1 et très loin de p1.

Selon une **approche de théorie des graphes**, une fonction de similarité entre graphes flous peut induire une distance entre ensembles flous, si ceux-ci sont représentés à un haut niveau par une structure de graphe (le moyen le plus simple pour cela serait de considérer la grille discrète sur laquelle est définie l’image comme un graphe, mais cette représentation serait assez peu structurée). Une telle distance porte typiquement sur la comparaison à la fois des attributs des nœuds et sur ceux des arcs [226].

Discussion. Pour des applications de traitement d’images, la première classe de méthodes (comparant seulement les fonctions d’appartenance) est adaptée à des applications où les deux ensembles flous à comparer représentent la même structure ou une structure issue d’une image et un modèle. Les applications envisageables sont du domaine de la reconnaissance de formes à partir de modèles ou à partir de recherche d’analogies entre cas.

Au contraire, les définitions de la deuxième classe, qui combinent l’information de distance spatiale à la comparaison des fonctions d’appartenance, sont beaucoup plus riches et permettent une analyse des structures et des objets dans les images plus générale, plus complète et plus proche des problèmes d’image. Elles sont donc particulièrement prometteuses dès que la disposition topologique et spatiale des structures d’intérêt est importante (c’est le cas la plupart du temps, par exemple en segmentation, classification, interprétation de scènes, etc.). Les distances de ce type peuvent fournir, comme application directe, des

mesures de conflit entre ensembles flous qui soient bien adaptées à l'information spatiale des images.

5.5.4 Distance géodésique dans un ensemble flou

Dans cette partie, nous nous intéressons à la distance entre deux points d'un ensemble flou μ , dans une optique de généralisation de la distance géodésique binaire, largement utilisée en traitement d'images, en particulier en morphologie mathématique. L'idée fondamentale est de combiner $d_E(x, y)$ à un degré mesurant à quel point un chemin entre x et y reste dans l'ensemble flou μ . La connexité floue c_μ [282, 283] (voir partie 5.4.1) est donc bien adaptée à ce propos. Dans le cas idéal, nous voudrions que la distance géodésique floue, pour conduire à des comportements et une interprétation satisfaisants, vérifie les propriétés de positivité, symétrie, séparabilité, l'inégalité triangulaire, que d_μ dépende du plus court chemin entre x et y qui « sorte » de μ « le moins possible », que d_μ tende vers l'infini s'il n'est pas possible d'aller de x à y sans passer par un point de degré d'appartenance à μ égal à 0, que d_μ soit décroissante par rapport à $\mu(x)$, $\mu(y)$ et c_μ , que d_μ soit égale à la distance géodésique classique si μ est binaire. Plusieurs définitions ont été proposées dans [29, 32], avec des propriétés différentes. Celle qui a les meilleures propriétés est la suivante :

$$d_\mu(x, y) = \frac{l(L_2^*(x, y))}{c_\mu(x, y)}, \quad (5.76)$$

où $L_2^*(x, y)$ est le plus court chemin entre x et y sur lequel c_μ soit atteint (ce chemin s'interprète comme une géodésique descendant le moins possible dans les degrés d'appartenance), et $l(L_2^*(x, y))$ désigne sa longueur. Toutes les propriétés requises sauf l'inégalité triangulaire sont satisfaites par cette formule. À partir de cette définition, il est possible de construire une véritable distance, satisfaisant l'égalité triangulaire, sans perdre les autres propriétés, mais qui n'est pas très opératoire en général.

5.6 Position relative directionnelle entre deux objets

Contrairement aux relations spatiales étudiées jusqu'à présent, les relations de position relative directionnelle (« à droite de », « au-dessus », etc.) sont des notions qui sont intrinsèquement vagues et mal définies, même si les objets sont binaires. Ainsi il ne s'agit plus simplement d'étendre des définitions bien formalisées au cas d'ensembles flous mais de formaliser de telles notions imprécises. Plusieurs travaux dans ce sens peuvent être trouvés dans la littérature.

5.6.1 Les principales approches floues

Une des premières approches floues a été proposée dans [197]. Elle repose sur la projection des ensembles flous sur un axe défini par la direction à laquelle on s'intéresse (par exemple l'axe x pour la relation « à gauche de »). La projection d'un objet de référence bidimensionnel R fournit un ensemble flou à une dimension donnant le degré avec lequel chaque point satisfait la relation par rapport à R . L'évaluation de la relation d'un objet A par rapport au premier est obtenue par la somme normalisée de ces degrés ponctuels, pondérés par l'appartenance de chaque point à la projection de A . Ainsi, si $A^f(x)$ et $R^f(x)$ désignent les projections

normalisées des ensembles flous A et R sur l'axe x , le degré avec lequel A est à gauche de R est défini par :

$$\mu_{gauche}^R(A) = \mu_\pi^R(A) = \frac{\int_{-\infty}^{+\infty} A^f(x) \int_x^{+\infty} R^f(y) dy dx}{\int_{-\infty}^{+\infty} A^f(y) dy \int_{-\infty}^{+\infty} R^f(y) dy}. \quad (5.77)$$

La plupart des méthodes s'appuient sur une définition linguistique de quelques relations de base. En 2D, une valeur linguistique est représentée par une fonction d'appartenance sur l'intervalle $[-\pi, +\pi]$, et dépend de l'angle θ entre le segment joignant deux points et l'axe x [246, 191]. Par exemple, la valeur « à droite de » (correspondant à un angle 0) peut être définie par :

$$\mu_0(\theta) = \begin{cases} \cos^2(\theta) & si \theta \in [-\frac{\pi}{2}, +\frac{\pi}{2}] \\ 0 & sinon \end{cases} \quad (5.78)$$

Les relations μ_α pour d'autres directions α sont déduites par simple translation sur $[-\pi, +\pi]$ de cette fonction. L'évaluation d'une relation entre deux objets se fait par comparaison entre ces valeurs linguistiques et un calcul d'angles entre les points des deux objets. Ce calcul peut prendre différentes formes. Dans [202, 191], seuls les centres de gravité c_R et c_A des deux objets sont considérés (si les objets sont flous, on calcule les centres de gravité pondérés par les fonctions d'appartenance), et la relation dans une direction α est évaluée par :

$$\mu_\alpha^R(A) = \mu_\alpha(\theta(c_R, c_A)). \quad (5.79)$$

D'autres approches évitent de réduire les objets à un point et considèrent l'histogramme d'angles normalisé, calculé pour toutes les paires de points des deux objets. L'évaluation d'une relation revient alors à comparer cet histogramme à l'ensemble flou de la direction d'intérêt. Cette comparaison peut se faire en calculant l'ensemble flous de compatibilité entre ces deux ensembles flous [245, 246], par une méthode d'agrégation par moyennage [202, 191], ou par pattern matching flou [33]. Une extension de l'histogramme d'angles, l'histogramme de forces, a été proposée dans [235], dans laquelle la contribution de chaque paire de points à l'histogramme est pondérée par une fonction de la distance entre ces points, permettant ainsi de donner par exemple plus d'importance aux points proches. La valeur de l'histogramme pondéré vaut alors pour chaque angle θ :

$$H^R(A)(\theta) = \sum_{a \in R, b \in A \mid \theta(a,b)=\theta} \varphi(||\vec{ab}||), \quad (5.80)$$

où φ est une fonction décroissante, par exemple $\varphi(x) = \frac{1}{x^r}$ (pour $r = 0$, on retrouve l'histogramme d'angles).

Parallèlement à ces approches qui passent dans des espaces de représentation d'objets de dimension plus faible (0 pour les centres de gravité, 1 pour les projections et les histogrammes d'angles en 2D, 2 pour les histogrammes d'angles en 3D), une approche complètement différente a été proposée dans [33, 41, 42], où le raisonnement est effectué complètement dans l'espace de l'image \mathcal{S} , par une méthode morphologique. L'idée est de définir un ensemble flou dans l'espace représentant pour chaque point le degré avec lequel la relation est satisfaite par rapport à l'objet de référence R . Cet ensemble peut être calculé par une dilatation floue (voir paragraphe 5.3) par un élément structurant radial ν , dont la fonction d'appartenance en un point décroît en fonction de l'écart entre les coordonnées angulaires de ce point et celles définissant la direction d'intérêt \vec{u}_α , par exemple :

$$\forall P \in \mathcal{S}, \nu(P) = \max[0, 1 - \frac{2}{\pi} \arccos \frac{\vec{OP} \cdot \vec{u}_\alpha}{\|\vec{OP}\|}]. \quad (5.81)$$

Cet ensemble flou (spatial) correspond à une fonction de la visibilité de chaque point de l'espace depuis l'objet dans la direction la plus proche de celle de la relation. Il peut être calculé une fois pour toutes, pour une direction donnée, et permet d'évaluer ensuite très rapidement le degré avec lequel n'importe quel objet A satisfait cette relation par rapport à R . Cette évaluation peut être effectuée par une méthode de pattern matching flou, donnant une évaluation sous forme d'intervalle $[N, \Pi]$, ou par une moyenne M :

$$N_\alpha^R(A) = \inf_{x \in S} T[\mu_\alpha(R)(x), 1 - \mu_A(x)], \quad \Pi_\alpha^R(A) = \sup_{x \in S} t[\mu_\alpha(R)(x), \mu_A(x)],$$

$$M_\alpha^R(A) = \frac{1}{|A|} \sum_{x \in S} t[\mu_A(x), \mu_\alpha(R)(x)], \quad (5.82)$$

où t est une t-norme et T une t-conorme.

Nous ne détaillons pas ici les propriétés de ces différentes approches. Une différence notable, qui peut avoir des conséquences algorithmiques importantes suivant les applications, est que les approches à partir d'histogrammes d'angles nécessitent un calcul d'histogramme par paire d'objets qui permet d'évaluer toutes les relations souhaitées, alors que la méthode morphologique nécessite un calcul de dilatation floue par direction, et permet l'évaluation de la relation de plusieurs objets à l'objet de référence.

5.6.2 Un exemple simple

Nous illustrons quelques unes de ces définitions sur les exemples de la figure 5.13, dans lesquels les objets sont très simples mais satisfont clairement plusieurs relations à des degrés divers.

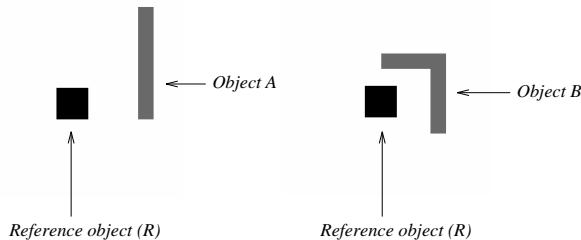


FIG. 5.13 – Deux exemples simples illustrant la difficulté de donner une définition en tout ou rien de la position relative directionnelle entre deux objets.

La figure 5.14 montre les histogrammes d'angles pondérés pour le premier exemple. L'objet A est surtout à droite de R (angles proches de 0), mais satisfait également d'autres relations, et certaines parties sont presques au-dessus de R . Lorsque r augmente dans la fonction $\varphi(x) = \frac{1}{x}$, la partie de A à droite de R prédomine car elle est la plus proche de R .

Les tables 5.3 et 5.4 montrent les résultats pour le premier exemple. Toutes les méthodes sont concordantes pour dire que A est surtout à droite de R , ce degré augmentant avec r . Les valeurs absolues peuvent être un peu différentes selon les approches, mais l'ordre entre les valeurs et le comportement général restent les mêmes.

Les tables 5.5 et 5.6 montrent les résultats obtenus sur le deuxième exemple. Deux relations principales (parmi les quatre étudiées) sont satisfaites : à droite et au-dessus. La

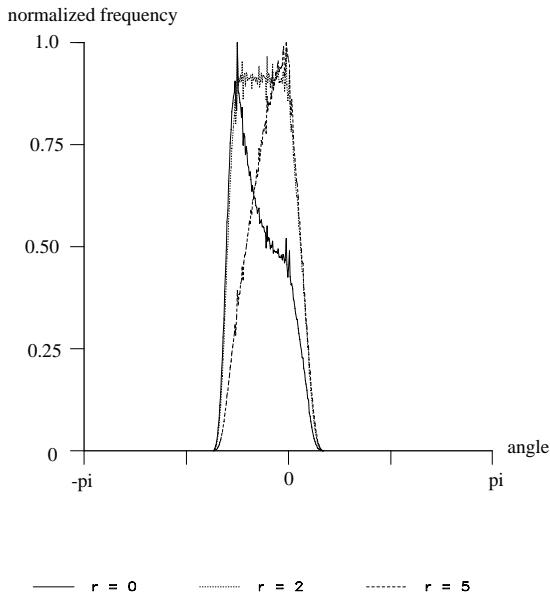


FIG. 5.14 – Histogrammes de forces pour l'exemple de gauche de la figure 5.13. Train plein : $r = 0$, pointillés : $r = 2$, tiretés : $r = 5$.

Objet A par rapport à l'objet R							
Relation	Centre	Agrégation			Compatibilité		
		$r = 0$	$r = 2$	$r = 5$	$r = 0$	$r = 2$	$r = 5$
Gauche	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Droite	0,76	0,73	0,79	0,86	0,62	0,67	0,75
En-dessous	0,00	0,00	0,01	0,01	0,05	0,06	0,06
Au-dessus	0,24	0,27	0,20	0,13	0,38	0,33	0,25

TAB. 5.3 – Position relative de A par rapport à R (figure 5.13), avec les méthodes du centre d'inertie, d'agrégation et de compatibilité. Les histogrammes sont calculés pour $r = 0$, $r = 2$ et $r = 5$.

méthode du centre d'inertie ne rend pas compte de la relation au-dessus, ce qui montre les limites de cette approche trop réductrice. Comme la partie de B qui est au-dessus de R est plus proche de R que la partie qui est à droite, les valeurs de la relation à droite décroissent quand r augmente alors que celles de la relation au-dessus augmentent. L'approche morphologique met bien en évidence l'ambiguïté des relations pour ces objets. Des parties de B satisfont complètement la relation au-dessus alors que d'autres parties ne la satisfont pas du tout.

Objet <i>A</i> par rapport à l'objet <i>R</i>				
Relation	FPM			Approche morphologique
	$r = 0$	$r = 2$	$r = 5$	
Gauche	[0,00, 0,00]	[0,00, 0,00]	[0,00, 0,00]	[0,00, 0,00] $M = 0,00$
Droite	[0,37, 0,68]	[0,39, 0,98]	[0,58, 1,00]	[0,50, 1,00] $M = 0,81$
En-dessous	[0,00, 0,10]	[0,00, 0,12]	[0,00, 0,13]	[0,00, 0,35] $M = 0,05$
Au-dessus	[0,32, 0,63]	[0,02, 0,61]	[0,00, 0,42]	[0,00, 0,73] $M = 0,44$

TAB. 5.4 – Position relative de *A* par rapport à *R* (figure 5.13), avec la méthode de fuzzy pattern matching (FPM) entre les relations et les histogrammes, pour $r = 0$, $r = 2$ and $r = 5$, et avec la méthode morphologique (les intervalles $[N, \Pi]$ et la valeur moyenne sont donnés).

Objet <i>B</i> par rapport à l'objet <i>R</i>						
Relation	Centre	Agrégation			Compatibilité	
		$r = 0$	$r = 2$	$r = 5$	$r = 0$	$r = 2$
Gauche	0,00	0,00	0,00	0,01	0,05	0,06
Droite	0,83	0,63	0,54	0,33	0,55	0,49
En-dessous	0,00	0,03	0,02	0,01	0,17	0,16
Au-dessus	0,17	0,34	0,43	0,66	0,45	0,51

TAB. 5.5 – Position relative de *B* par rapport à *R* (figure 5.13), avec les méthodes du centre d'inertie, d'agrégation et de compatibilité.

Objet <i>B</i> par rapport à l'objet <i>R</i>				
Relation	FPM			Approche morphologique
	$r = 0$	$r = 2$	$r = 5$	
Gauche	[0,00, 0,06]	[0,00, 0,10]	[0,00, 0,11]	[0,00, 0,44] $M = 0,02$
Droite	[0,34, 0,81]	[0,10, 0,75]	[0,00, 0,39]	[0,29, 1,00] $M = 0,81$
En-dessous	[0,00, 0,28]	[0,00, 0,26]	[0,00, 0,11]	[0,00, 0,60] $M = 0,11$
Au-dessus	[0,19, 0,66]	[0,25, 0,91]	[0,61, 1,00]	[0,00, 1,00] $M = 0,52$

TAB. 5.6 – Position relative de *B* par rapport à *R* (figure 5.13), avec la méthode de fuzzy pattern matching (FPM) entre les relations et les histogrammes, et avec la méthode morphologique (les intervalles $[N, \Pi]$ et la valeur moyenne sont donnés).

5.7 Classification floue

Nous présentons dans cette partie les principales méthodes de classification. Outre leur utilisation directe pour des problèmes de classification, elles sont souvent utilisées pour estimer des fonctions d'appartenance dans des espaces de caractéristiques.

Le principe de la classification floue est d'affecter un élément à classer non pas à une classe comme dans les approches classiques, mais à toutes les classes avec un certain degré. Cette approche permet de prendre en compte facilement les transitions douces entre classes, l'ambiguïté, l'appartenance partielle à plusieurs classes. La classification nette est un cas particulier dans lequel chaque élément a un degré d'appartenance non nul (et égal à 1) à exactement une classe, et un degré d'appartenance nul à toutes les autres.

5.7.1 C-moyennes floues

La méthode de classification la plus populaire est appelée C-moyennes floues [23] et généralise la méthode classique des k-moyennes [137, 136].

Soit X un ensemble d'éléments à classer (définis dans un certain espace de caractéristiques) et $P = \{C_1, C_2, \dots, C_C\}$ l'ensemble de classes (le nombre de classes C est fixé). Une partition floue est définie par :

1. $\forall x_i \in X, \forall C_j \in P, \mu_{ij} \in [0, 1]$, où μ_{ij} désigne le degré d'appartenance de l'élément x_i à la classe C_j ;
2. $\forall x_i \in X, \sum_{j=1}^C \mu_{ij} = 1$;
3. $\forall j \leq C, 0 < \sum_{i=1}^N \mu_{ij} < N$, où N est le cardinal de X .

Comme pour les k-moyennes, le principe des C-moyennes floues est de mettre à jour de manière itérative les centres des classes m_j (moyennes) et les degrés d'appartenance aux classes. Ces itérations visent à optimiser un critère d'erreur quadratique pondéré qui s'écrit :

$$J_m = \sum_{j=1}^C \sum_i^N \mu_{ij}^m \|x_i - m_j\|^2, \quad (5.83)$$

où la norme est définie dans l'espace des caractéristiques, et m est un paramètre dans $]1, +\infty[$ appelé facteur de flou. Cette optimisation est faite sous la contrainte d'obtenir une partition floue.

A chaque itération, les degrés d'appartenance qui minimisent le critère sont :

$$\mu_{ij} = \frac{1}{\sum_{j=1}^C \left[\frac{\|x_i - m_i\|}{\|x_i - m_j\|} \right]^{\frac{2}{m-1}}}, \quad (5.84)$$

et les centres de classes sont donnés par :

$$m_j = \frac{\sum_i \mu_{ij}^m x_i}{\sum_i \mu_{ij}^m}. \quad (5.85)$$

La matrice de dispersion de la classe C_j est définie par :

$$M_j = \sum_{i=1}^N \mu_{ij}^m (x_i - m_j)(x_i - m_j)^t, \quad (5.86)$$

et la dispersion intra-classes par :

$$M = \sum_{j=1}^C M_j. \quad (5.87)$$

Il est facile de vérifier que la fonctionnelle J_m est la trace de la matrice de dispersion M .

L'algorithme des C-moyennes floues s'écrit alors :

- initialiser les centres de classe,
- itérer la séquence suivante :

1. calculer les degrés d'appartenance de chaque point à chaque classe selon l'équation 5.84),

2. calculer les nouveaux centres des classes selon l'équation 5.85.

- arrêter les itérations lorsque la partition ne change plus ou lorsque la décroissance de la fonctionnelle devient suffisamment faible.

Cet algorithme converge vers un minimum local de la fonctionnelle [23], ce qui nécessite d'avoir une initialisation pas trop éloignée de la solution.

Considérons un exemple où l'espace de caractéristiques est de dimension 1 et où l'on cherche deux classes. La figure 5.15 montre l'allure des fonctions d'appartenance. Ces fonctions deviennent de plus en plus proches de fonctions binaires lorsque m tend vers 1 (figure 5.16).

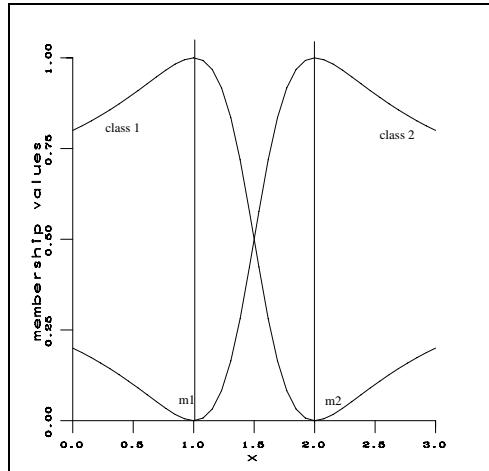


FIG. 5.15 – Allure des fonctions d'appartenance obtenues par les C-moyennes floues pour deux classes de centres $x = 1$ et $x = 2$ et pour $m = 2$.

Ces figures illustrent un des principaux inconvénients de cette approche : si la forme des fonctions d'appartenance est intuitivement satisfaisante à proximité des centres de classes, en revanche elle ne l'est plus quand on s'en éloigne. A cause de la contrainte de normalisation des degrés d'appartenance, la fonction s'annule au niveau du centre de l'autre classe puis recommence à croître. Elle n'est donc pas décroissante en fonction de la distance au centre de la classe. Cet effet est amplifié dans les cas où le nombre de classes est plus grand. Un exemple avec trois classes est illustré figure 5.17.

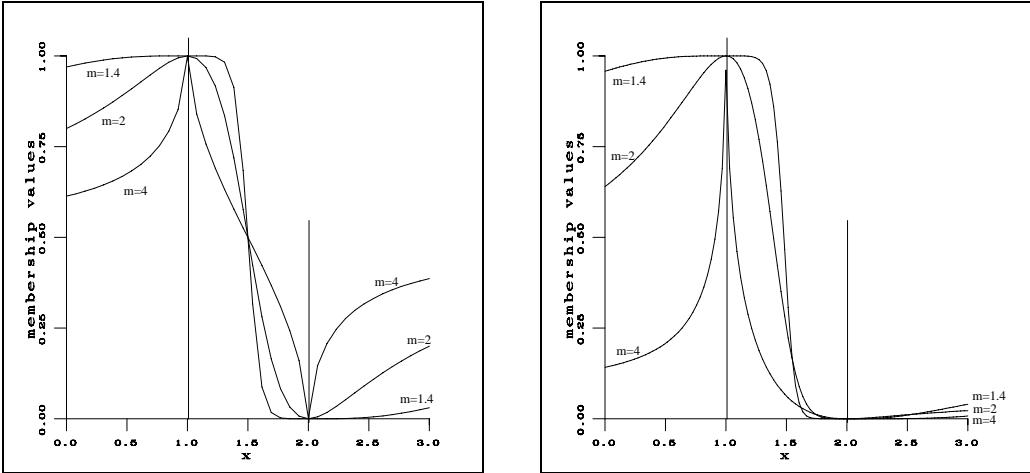


FIG. 5.16 – Fonctions d’appartenance obtenues pour deux classes de centres $x = 1$ et $x = 2$, pour $m = 1, 4$, $m = 2$ et $m = 4$. A gauche : fonction d’appartenance à la classe C_1 . A droite : le terme μ_{i1}^m de la fonctionnelle J_m .

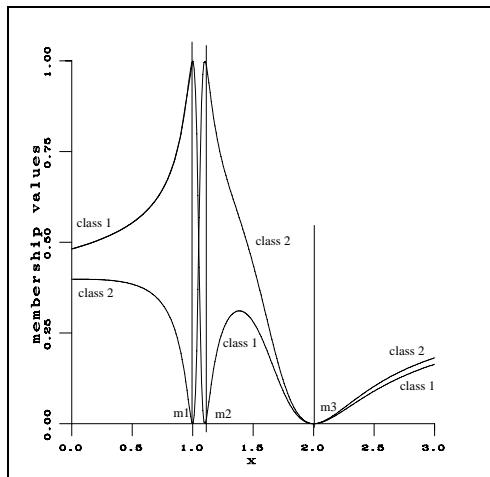


FIG. 5.17 – Exemple à trois classes de centres $x = 1$, $x = 1,1$ et $x = 2$.

5.7.2 C-moyennes possibilistes

Les C-moyennes possibilistes évitent ces inconvénients en rajoutant un terme dans la fonctionnelle et en abandonnant la contrainte de normalisation [201] :

$$J = \sum_{j=1}^C \sum_{i=1}^N \mu_{ij}^m \|x_i - m_j\|^2 + \sum_{j=1}^C \eta_j \sum_{i=1}^N (1 - \mu_{ij})^m \|x_i - m_j\|^2. \quad (5.88)$$

Les degrés d’appartenance deviennent alors :

$$\mu_{ij} = \frac{1}{1 + \frac{\|x_i - m_j\|^2}{\eta_j^{\frac{1}{m-1}}}}. \quad (5.89)$$

Le paramètre η_j contrôle la décroissance de la fonction d’appartenance à la classe C_j en fixant la distance au centre de la classe pour laquelle le degré d’appartenance vaut $1/2$. Le

paramètre m contrôle le flou comme dans la méthode précédente. Cette fois les fonctions ont une forme qui correspond mieux à l'intuition puisqu'elles sont décroissantes en fonction de la distance à la classe (figure 5.18).

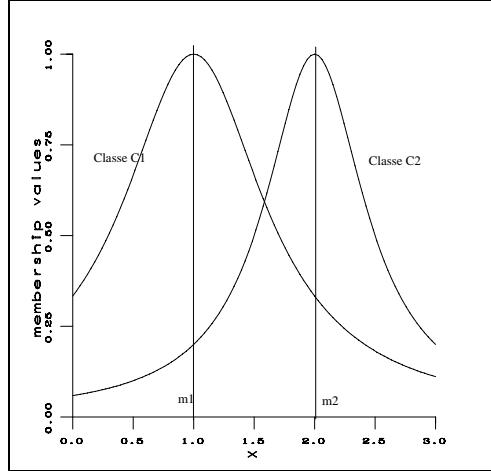


FIG. 5.18 – Fonctions d'appartenance obtenues avec les C-moyennes possibilistiques pour deux classes de centres $x = 1$ et $x = 2$, avec $m = 1$, $\eta_1 = 0,5$ et $\eta_2 = 0,25$.

Cette approche permet aisément de définir un rejet en ambiguïté (si les valeurs d'appartenance à deux classes sont trop proches) et un rejet en distance (si les degrés d'appartenance à toutes les classes sont trop faibles).

La détermination de η_j est un problème difficile. Dans [201] deux solutions sont proposées pour déterminer ces paramètres à partir des données :

1. l'une utilise la dispersion :

$$\eta_j = \frac{\sum_{i=1}^N \mu_{ij} \|x_i - m_j\|^2}{\sum_{i=1}^N \mu_{ij}}$$

2. l'autre utilise les échantillons de la classe dont la fonction d'appartenance est plus grande qu'une valeur donnée α :

$$\eta_j = \frac{\sum_{i \in \{\mu_{ij}\}_\alpha} \|x_i - m_j\|^2}{|\{\mu_{ij}\}_\alpha|}$$

où $\{\mu_{ij}\}_\alpha$ désigne l' α -coupe de μ .

5.7.3 k-plus proches voisins flous

La méthode des k-plus proches voisins (kppv) [92] a également été étendue au flou [134].

Dans le cas binaire, l'appartenance (binaire) d'un point x à une classe i peut s'écrire de la manière suivante :

$$\mu_i(x) = U\left[\sum_{j=1}^k \mu_i(x_j)\right] \quad (5.90)$$

avec

$$U(t) = \begin{cases} 1 & \text{si } t > \frac{k}{2} \\ 0 & \text{sinon} \end{cases}$$

Il suffit alors de remplacer la fonction d'Heaviside par une fonction douce pour étendre la méthode au cas flou :

$$\mu_i(x) = \frac{1}{1 + \exp[\frac{1}{b_i}(\frac{k}{2} - t_i(x))]} \quad (5.91)$$

avec

$$t_i(x) = \sum_{j=1}^k \mu_i(x_j)$$

ou encore, en tenant compte des distances aux voisins :

$$t_i(x) = \sum_{j=1}^k \mu_i(x_j) \exp[-\lambda(\frac{d(x, x_j)}{d_m^i})^2]$$

Le paramètre b_i permet de contrôler le flou, et peut être défini par exemple en fonction de l'entropie floue :

$$b_i = H(i) = \frac{1}{K} \sum_x [\mu_i(x) \log(\mu_i(x)) + (1 - \mu_i(x)) \log(1 - \mu_i(x))].$$

5.8 Opérations locales d'amélioration ou de détection

Nous regroupons dans cette partie les techniques locales de filtrage au sens le plus général, visant à améliorer le contraste de l'image, à supprimer le bruit, à extraire les contours, etc. Notons que ces buts sont tous bien différents et souvent contradictoires. Cependant, les principes des techniques sont similaires, et peuvent être regroupés en deux classes : les techniques d'optimisation fonctionnelle d'une part, et les techniques à base de règles d'autre part. Ces aspects étant déjà largement développés dans d'autres ouvrages (par exemple [24, 207, 332, 11]), nous n'en donnons ici que les grandes lignes.

5.8.1 Approches fonctionnelles

Ces techniques consistent à minimiser ou maximiser une fonctionnelle, qui peut être interprétée comme une représentation analytique d'un objectif. Par exemple, rehausser le contraste d'une image selon cette technique revient à diminuer la quantité de flou de l'image. Cela peut être effectué par une simple modification des fonctions d'appartenance (par exemple par des opérateurs d'intensification), par la minimisation d'un indice de flou tel que l'entropie, ou même par la détermination d'un seuil optimal (par exemple au sens de la minimisation d'un indice de flou) qui pousse le rehaussement jusqu'à la binarisation de la fonction d'appartenance [261, 262]. D'autres méthodes consistent simplement à modifier les filtres classiques (médian par exemple) en incorporant des fonctions de pondération floues [208].

5.8.2 Techniques à base de règles

Les techniques à base de règles reposent sur des modèles idéaux (de filtres, de contours, etc.). Ces cas idéaux étant rarement rencontrés, des variations et des écarts par rapport à ces modèles sont permis grâce à des représentations floues de ces modèles, sous forme de règles floues.

Par exemple, un opérateur de lissage peut être exprimé par [287, 288] :

SI	un pixel est <i>plus foncé</i> que ses voisins
ALORS	<i>augmenter</i> son niveau de gris
SINON SI	le pixel est <i>plus clair</i> que ses voisins
ALORS	<i>diminuer</i> son niveau de gris
SINON	le laisser inchangé

Dans cette représentation, les termes en italiques sont définis par des ensembles flous ou des opérations floues. Typiquement les caractéristiques de niveaux de gris sont définies par des variables linguistiques dont la sémantique est constituée d'ensembles flous sur l'intervalle de niveaux de gris de l'image. Les actions sont des fonctions floues appliquées aux niveaux de gris et aux pixels. La mise en œuvre de ces règles n'est pas spécifique au traitement d'images et est exactement la même qu'en logique floue classique (voir chapitres 1 et 4).

Les règles sont parfois simplement une représentation différente de méthodes fonctionnelles. Par exemple, si l'on représente un détecteur de contours par :

SI	un pixel <i>appartient au contour</i>
ALORS	<i>augmenter beaucoup</i> son niveau de gris
SINON	<i>diminuer beaucoup</i> son niveau de gris

Il est clair que le résultat final est une image où le niveau de gris de chaque pixel est une fonction croissante du gradient de l'image en ce pixel, ce qui est équivalent à une représentation fonctionnelle.

Des règles plus complexes peuvent être trouvées, par exemple dans [206, 269], où un détecteur de contours est exprimé par une série de règles impliquant le gradient, la symétrie et la raideur du contour. Des systèmes de règles floues ont également été proposés pour relier des morceaux de contours, en fonction de critères de proximité et d'alignement.

Un des avantages de cette approche est que la représentation est proche de l'intuition, ce qui permet en particulier de construire assez facilement des opérateurs adaptatifs. Par exemple, des règles peuvent permettre de décider en fonction du contexte local quel type d'opérateur appliquer.

Chapitre 6

Définitions générales de la fusion d'informations

La fusion est devenue un aspect important de traitement de l'information dans plusieurs domaines très différents, dans lesquels les informations à fusionner, les objectifs, les méthodes, et donc la terminologie, peuvent varier beaucoup, même si les analogies sont également nombreuses. Ce chapitre vise à préciser le contexte de la fusion dans le domaine du traitement et de l'interprétation des images, à préciser les concepts et à dégager des définitions. Les spécificités de la fusion en image seront précisées dans le chapitre suivant.

6.1 Choix d'une définition

Nous considérons ici toujours un sens large du terme « information ». En particulier, il couvre à la fois des données (par exemple des mesures, des images, des signaux, etc.) et des connaissances (sur les données, sur le domaine, sur des contraintes, etc.) qui peuvent être génériques ou spécifiques.

La définition de la fusion d'informations que nous utiliserons dans toute la suite est la suivante :

Définition 1 *La fusion d'informations consiste à combiner des informations issues de plusieurs sources afin d'améliorer la prise de décision.*

Cette définition, issue en grande partie des discussions menées au sein du groupe de travail sur la fusion d'informations du GdR-PRC ISIS¹, est suffisamment générale pour englober la diversité des problèmes de fusion que l'on rencontre en traitement des images. Son intérêt est qu'elle est focalisée sur les étapes de combinaison et de décision, ces deux opérations pouvant prendre des formes différentes suivant les problèmes et les applications.

Pour chaque type de problème et d'application, cette définition pourra être plus spécifique, en répondant à un certain nombre de questions : quel est le but de la fusion ? comment s'exprime la décision ? quelles sont les informations à fusionner ? quelles sont leurs origines ? quelles sont leurs caractéristiques (incertitude, relations entre les informations, génériques

¹<http://www-isis.enst.fr>

ou factuelles, statiques ou dynamiques, etc.) ? quelle méthodologie choisir ? comment évaluer et valider la méthode et les résultats ? quelles sont les difficultés principales, les limites ? etc.

Situons cette définition par rapport à celles proposées par quelques autres groupes de travail qui ont structuré le domaine de la fusion d'informations.

La définition 1 est un peu plus spécifique que celle proposée par le groupe de travail européen FUSION [51], qui a travaillé sur la fusion dans différents domaines de 1996 à 1999². La définition générale retenue dans ce projet est : regrouper des informations issues de plusieurs sources et exploiter l'information regroupée afin de répondre à des questions, prendre des décisions, etc. Dans cette définition, qui est également focalisée sur la combinaison et les buts, les buts s'arrêtent souvent avant l'étape de décision, et ne sont pas restreints à l'amélioration de l'information globale. Ils incluent par exemple l'obtention d'un point de vue global, par exemple dans les problèmes de fusion d'opinions ou de préférences de personnes, qui est un des thèmes abordés dans ce projet, mais qui sort du cadre de ce cours. Ici l'amélioration de la connaissance fait référence au monde tel qu'il est, et non au monde tel qu'on voudrait qu'il soit comme c'est le cas en fusion de préférences par exemple.

Un des premiers efforts notables d'éclaircissement du domaine a été effectué par le groupe de travail sur la fusion de données du « Joint Directors of Laboratories » (JDL), du département de la défense US. Ce groupe a été créé en 1986 et s'est attaché à préciser et codifier la terminologie de la fusion de données dans une sorte de dictionnaire (« Data Fusion Lexicon ») [1]. La méthode proposée est dédiée essentiellement à des applications de défense (telles que la poursuite de cibles, la reconnaissance et l'identification automatiques de cibles) et est focalisée sur les fonctionnalités, en identifiant des processus, des fonctions et des techniques [171]. Elle insiste sur la description d'une hiérarchie d'étapes de traitement dans un système. La définition que nous utilisons ici s'oppose à celle du JDL et choisit un autre point de vue, en s'attachant plus à la description des méthodes de combinaison et de décision qu'à celle d'un système. Elle est mieux adaptée à la variété des situations qui peuvent être rencontrées en traitement d'images. Elle est en ce sens plus générale.

Un autre groupe de travail européen de l'EARSel (Association of Remote Sensing Laboratories) a étendu la définition du JDL au domaine plus général de l'imagerie satellitaire [345] : la fusion de données constitue un cadre formel dans lequel s'expriment les données provenant de sources diverses ; elle vise à l'obtention d'informations de plus grande qualité ; la définition exacte de « plus grande qualité » dépendra de l'application. Cette définition englobe la plupart des définitions proposées par plusieurs auteurs en imagerie satellitaire, qui sont regroupées dans [345]. La définition 1 va plus loin que celle-ci, jusqu'à la décision.

La fusion peut être comprise dans un sens plus ou moins large. D'autres concepts tels que l'estimation, la révision, l'association de données, la fouille de données sont parfois considérés comme des problèmes de fusion dans un sens large du terme. Précisons ces concepts :

Fusion et estimation L'estimation vise à combiner plusieurs valeurs d'un paramètre ou d'une distribution pour en déduire une valeur plausible. On retrouve donc les étapes de combinaison et de décision qui sont les deux ingrédients principaux de la définition 1. D'un autre côté, les méthodes de fusion numériques nécessitent souvent une étape préliminaire d'estimation des distributions à combiner (voir paragraphe 6.4 ci-dessous), et l'estimation apparaît alors comme une des étapes d'un processus de fusion.

²Ce chapitre a beaucoup bénéficié des discussions au sein de ce groupe de travail, et nous tenons ici à en remercier tous les participants.

Fusion et révision ou mise à jour La révision ou la mise à jour consiste à compléter ou modifier une information en fonction de nouveaux apports d'informations. Elle peut être considérée comme un des domaines de la fusion. Parfois, la fusion est considérée dans un sens plus restreint, où la combinaison est symétrique. La révision, elle, ne l'est pas, et fait une distinction entre l'information a priori et les nouvelles informations. Ici, nous considérons entre autres des processus dynamiques (en particulier en robotique), et il nous semble donc important d'inclure la révision et la mise à jour dans la fusion (par exemple pour des applications à la compréhension de l'environnement par un robot). La révision concerne l'ajout d'une nouvelle information permettant de modifier ou préciser l'information disponible auparavant sur le phénomène observé, alors que la mise à jour concerne une modification du phénomène conduisant à modifier l'information sur celui-ci (typiquement dans un processus temporel).

Fusion et association L'association de données est l'action qui permet de trouver parmi différents signaux, issus de deux sources ou plus, ceux qui sont émis par le même objet (source ou cible). Selon Bar-Shalom et Fortman [18], l'association de données est l'étape la plus difficile en poursuite de cibles multiples. Elle consiste à détecter et associer des mesures bruitées, dont les origines sont incertaines à cause de plusieurs facteurs, tels que les fausses alarmes aléatoires dans la détection, le fouillis, les cibles interférentes, les pièges et autres contre-mesures. Les modèles principaux utilisés dans ce domaine sont soit déterministes (à partir de tests d'hypothèses classiques), soit probabilistes (essentiellement bayésiens) [18, 280, 211]. La méthode la plus courante [18] repose sur le filtre de Kalman sous hypothèse gaussienne. Plus récemment, d'autres méthodes d'estimation ont été proposées, comme le modèle d'estimation interactive multiple (« Interactive Multiple Model estimator », IMM), qui peut s'adapter à différents types de mouvement, réduire le bruit, tout en préservant une bonne précision des estimations des états [353]. Ainsi les problèmes traités par l'association de données sont assez différents de ceux couverts par la définition 1.

Fusion et fouille de données La fouille de données (« Data Mining ») consiste à extraire des parties intéressantes de l'information et des données, qui peuvent être par exemple des données spéciales (au sens d'une propriété particulière) ou rares. Elle peut être opposée à la fusion visant à expliquer, où des tendances générales sont recherchées, ou à la fusion visant à généraliser et à induire des connaissances plus génériques à partir de données. Ici nous écarterons la fouille de données des problèmes de fusion.

6.2 Caractéristiques générales des données

Dans cette partie, nous décrivons succinctement les caractéristiques générales des informations à fusionner, qui doivent souvent être prises en compte dans un processus de fusion. Des exemples plus détaillés et plus spécifiques seront donnés pour chaque domaine dans les chapitres suivants.

Une première caractéristique concerne le type d'information à fusionner. Il peut s'agir d'observations directes, de résultats de traitements sur ces observations, de connaissances plus génériques, exprimées sous forme de règles par exemple, ou d'avis d'experts. Ces informations peuvent être exprimées sous forme numérique ou sous forme symbolique (voir paragraphe 6.3). Une attention particulière doit être portée à l'échelle utilisée pour représenter les informations. Cette échelle n'a pas forcément de signification absolue, mais le minimum

à garantir est que les informations puissent être comparées selon cette échelle, c'est-à-dire que les échelles induisent un ordre sur les populations. Cela se traduit par des propriétés de commensurabilité, voire de normalisation.

Le niveau de l'information qui va être fusionnée est également un aspect très important. On distingue généralement le bas niveau (typiquement les mesures originales), d'un niveau plus élevé nécessitant des étapes préliminaires de traitement, d'extraction de primitives et de structuration de l'information. Suivant le niveau auquel on se place, les contraintes ne sont pas les mêmes, ni les difficultés. Cela sera illustré par exemple dans le cas de la fusion d'images dans le chapitre 7.

D'autres distinctions sur les types de données sont également intéressantes à souligner, car elles donnent lieu à des modélisations et à des types de traitements différents. La distinction entre données fréquentes et données rares en fait partie. Les informations peuvent également être factuelles ou génériques. Les connaissances génériques peuvent être par exemple un modèle du phénomène observé, des règles générales, des contraintes d'intégrité. Les informations factuelles sont plus directement reliées aux observations. Souvent, ces deux types d'information ont des spécificités différentes. Les informations génériques sont en général moins spécifiques (et font office de « défaut ») que les informations factuelles qui sont directement relatives au phénomène particulier observé. Le défaut est considéré si l'information spécifique n'est pas présente ou pas assez fiable. Dans le cas contraire, et si les informations sont contradictoires, on préférera les informations les plus spécifiques. Enfin, les informations peuvent être statiques ou dynamiques, et là encore leur modélisation et leur prise en compte diffèrent dans les deux cas.

Les informations manipulées dans un processus de fusion sont d'une part les informations à fusionner, et d'autre part des informations supplémentaires qui servent à guider ou aider la combinaison. Il peut s'agir d'informations sur les informations à combiner telles que des informations sur les sources, sur leur dépendance, sur leur fiabilité, des préférences sur les informations à combiner, etc. Il peut s'agir également d'informations contextuelles, sur le domaine. Ces informations supplémentaires ne sont pas forcément exprimées dans le même formalisme que les informations à combiner (en général elles ne le sont pas), mais elles peuvent intervenir dans le choix de la modélisation des informations à fusionner.

Une des caractéristiques importantes de l'information en fusion est son imperfection. Celle-ci est toujours présente (sinon la fusion ne serait pas nécessaire). Elle peut prendre diverses formes, qui sont brièvement décrites ci-dessous. Notons que ces notions ne trouvent pas toujours de définition consensuelle dans la littérature. Nous en proposons ici des définitions qui sont assez intuitives et conviennent bien au problème de la fusion, mais qui ne sont sûrement pas universelles. Les subtilités qui peuvent y être apportées sont volontairement omises ici, car elles seront reprises et illustrées dans les chapitres suivants pour chacun des domaines de fusion abordés dans ce livre.

Incertitude L'incertitude est relative à la vérité d'une information, et caractérise son degré de conformité à la réalité [126]. Elle fait référence à la nature de l'objet ou du fait concerné, à sa qualité, à son essence ou à son occurrence.

Imprécision L'imprécision concerne le contenu de l'information et mesure donc un défaut quantitatif de connaissance, sur une mesure [126]. Elle concerne le manque d'exactitude en quantité, en taille, en durée, le manque de définition d'une proposition qui est ouverte à diverses interprétations ou qui a des frontières vagues et mal définies. Cette

notion est souvent abusivement confondue avec celle d'incertitude, car les deux types d'imperfection sont souvent présents simultanément, et l'un peut induire l'autre. Il est important de les distinguer car ils sont souvent antagonistes, même si ces deux termes peuvent être inclus dans une acception plus large de l'incertitude. Inversement, des typologies en un plus grand nombre de classes ont été proposées [195].

Incomplétude L'incomplétude caractérise l'absence d'information apportée par la source sur certains aspects du problème. L'incomplétude des informations issues de chaque source est la raison principale qui motive la fusion. L'information fournie par chaque source est en général partielle, elle ne fournit qu'une vision du monde ou du phénomène observé, en n'en mettant en évidence que certaines caractéristiques.

Ambiguïté L'ambiguïté exprime la capacité d'une information de conduire à deux interprétations. Elle peut provenir des imperfections précédentes, par exemple de l'imprécision d'une mesure qui ne permet pas de différencier deux situations, ou de l'incomplétude qui induit des confusions possibles entre des objets et des situations qui ne peuvent être séparées selon les caractéristiques mises en évidence par la source. Un des objectifs de la fusion est de lever les ambiguïtés d'une source grâce aux informations apportées par les autres sources ou par les connaissances supplémentaires.

Conflit Le conflit caractérise deux ou plusieurs informations conduisant à des interprétations contradictoires et donc incompatibles. Les situations conflictuelles sont fréquentes dans les problèmes de fusion, et posent toujours des problèmes difficiles à résoudre. Tout d'abord, la détection des conflits n'est pas forcément facile. Ils peuvent facilement être confondus avec d'autres types d'imperfection, ou même avec la complémentarité des sources. Ensuite, leur identification et leur typologie est une question qui se pose souvent, mais de manière différente suivant leur domaine. Enfin, leur résolution peut prendre différentes formes. Elle peut reposer sur l'élimination de sources non fiables, sur la prise en compte d'informations supplémentaires, etc. Dans certains cas, il peut être préférable de retarder la combinaison et d'attendre d'autres informations susceptibles de lever les conflits, ou même de ne pas fusionner du tout.

D'autres caractéristiques de l'information sont plus positives, et sont exploitées pour limiter les imperfections :

Redondance La redondance est la qualité de sources qui apportent plusieurs fois la même information. La redondance entre les sources est souvent observée, dans la mesure où les sources donnent des informations sur le même phénomène. Idéalement, la redondance est exploitée pour réduire les incertitudes et les imprécisions.

Complémentarité La complémentarité est la propriété des sources qui apportent des informations sur des grandeurs différentes. Elle vient du fait qu'elles ne donnent en général pas d'informations sur les mêmes caractéristiques du phénomène observé. Elle est exploitée directement dans le processus de fusion pour avoir une information globale plus complète et pour lever les ambiguïtés.

Les outils permettant de modéliser les différents types d'informations, et de mesurer les imperfections de l'information ainsi que la redondance et la complémentarité seront décrits dans le chapitre 8.

6.3 Numérique / symbolique

De nombreuses discussions ont eu lieu dans la communauté de la fusion sur la dualité entre fusion numérique et fusion symbolique. L'objectif dans cette partie n'est pas de retracer en détails ces discussions mais plutôt de présenter les différents niveaux auxquels cette question peut être posée. Par une judicieuse description de ces niveaux, il est souvent possible de réduire ces débats. Les trois niveaux que nous distinguons ici sont ceux du type de données, du type de traitement appliqué aux données et du rôle des représentations. Ils sont détaillés dans les paragraphes suivants.

6.3.1 Données et informations

Par information numérique, nous entendons une information qui est directement donnée sous forme de nombre. Ces nombres peuvent représenter des mesures physiques, des niveaux de gris dans une image, une intensité de signal, la distance donnée par un télémètre, ou encore la réponse à un opérateur numérique de traitement. Ils peuvent être soit directement lus dans les données à fusionner, soit attachés au domaine ou à la connaissance contextuelle.

Par information symbolique, nous entendons toute information qui est donnée sous forme de symboles, de propositions, de règles, etc. De telles informations peuvent être attachées aux données à fusionner, ou à la connaissance du domaine (par exemple, des propositions sur les propriétés du domaine concerné, une information structurelle, des règles générales sur le phénomène observé, etc.).

La classification des informations et des données en numériques et symboliques ne peut pas toujours être faite de manière binaire. En effet, les informations peuvent être également hybrides, et des nombres peuvent représenter des codages d'informations de nature non numérique. C'est typiquement le cas de l'évaluation d'une donnée ou d'un traitement, de la quantification de l'imprécision ou de l'incertitude. Dans de tels cas, les valeurs absolues des nombres sont souvent peu importantes, et c'est surtout leur place dans une échelle, ou leur ordre si plusieurs grandeurs sont évaluées, qui joue un rôle prépondérant. Le terme « hybride » désigne alors des nombres utilisés comme des symboles pour représenter une information, mais avec une quantification, qui permet ensuite de les manipuler de manière numérique. Ces nombres peuvent être utilisés pour des informations symboliques aussi bien que numériques.

6.3.2 Traitements

En ce qui concerne le traitement de l'information, on appelle traitement numérique tout calcul sur des nombres. En fusion d'informations, cela couvre toutes les approches qui combinent des nombres par un calcul formel. Il est important de remarquer que ce type de traitement ne fait pas forcément d'hypothèse sur le type d'information qui est représentée par des nombres. A l'origine, les informations peuvent être de nature aussi bien numérique que symbolique.

Les traitements symboliques incluent le calcul formel sur des propositions (par exemple des approches de type logique ou les grammaires, sur lesquelles on peut trouver plus de détails dans [51]), prenant éventuellement en compte des connaissances numériques. Les approches

structurelles, telles que des approches à base de graphes, qui sont largement utilisées en reconnaissance des formes structurelles (en particulier pour la fusion), peuvent être mises dans cette classe.

Nous appelons traitement hybride les méthodes où la connaissance a priori est utilisée d'une manière symbolique pour contrôler des traitements numériques, par exemple en déclarant des règles propositionnelles qui suggèrent, permettent ou au contraire interdisent certaines opérations numériques. Typiquement, une proposition qui définit dans quels cas deux sources sont indépendantes peut être utilisée pour choisir la manière dont des probabilités sont combinées.

6.3.3 Représentations

Ainsi que le montrent les deux paragraphes précédents, les représentations et leurs types peuvent jouer des rôles très différents. Des représentations numériques peuvent être utilisées pour des données intrinsèquement numériques mais aussi pour l'évaluation et la quantification de données symboliques. Les représentations numériques en fusion d'informations sont utilisées de manière importante pour quantifier l'imprécision, l'incertitude ou la fiabilité de l'information (cette information peut être de nature numérique aussi bien que symbolique), et donc pour représenter plutôt de l'information sur les données à combiner que ces données elles-mêmes. Ces représentations seront plus détaillées dans les chapitres sur les méthodes numériques de fusion. Les représentations numériques sont aussi souvent utilisées pour des degrés de croyance attachés à des connaissances numériques ou symboliques, et pour des degrés de cohérence ou d'incohérence (ou de conflit) entre les informations (le cas le plus typique est peut-être celui de la fusion de bases de données ou de réglementations). Notons que le même formalisme numérique peut être utilisé pour représenter différents types de données ou de connaissances [35] : l'exemple le plus évident est celui de l'utilisation des probabilités pour représenter des données aussi différentes que des fréquences ou des degrés de croyance subjectifs [93].

Les représentations symboliques peuvent être utilisées dans des systèmes logiques, ou des systèmes à base de règles, mais aussi en tant que connaissance a priori ou connaissance contextuelle ou générique pour guider un traitement numérique, en tant que support structural par exemple en fusion d'images, et bien sûr en tant que sémantique attachée aux objets manipulés.

Dans de nombreux exemples, une forte dualité peut être observée entre les rôles des représentations numériques et symboliques, et peut être exploitée dans le cas de fusion de sources hétérogènes.

6.4 Systèmes de fusion et types d'architecture

En général, la fusion n'est pas une tâche simple. Elle peut se décomposer de manière schématique en plusieurs tâches. Nous les décrivons succinctement ici, car elles serviront de guide à la description des outils théoriques dans le chapitre 8. Considérons un problème général de fusion pour lequel on dispose de m sources S_1, S_2, \dots, S_m , et pour lequel le but est de prendre une décision dans un ensemble de n décisions possibles d_1, d_2, \dots, d_n . Les principales étapes à résoudre pour construire le processus de fusion sont les suivantes :

1. Modélisation : cette étape comporte le choix d'un formalisme, et des expressions des informations à fusionner dans ce formalisme. Cette modélisation peut être guidée par les informations supplémentaires (sur les informations et sur le contexte ou le domaine). Supposons pour fixer les idées que chaque source S_j fournit une information représentée par M_i^j sur la décision d_i . La forme de M_i^j dépend bien sûr du formalisme choisi. Elle peut être par exemple une distribution dans un formalisme numérique, ou une formule dans un formalisme logique.
2. Estimation : la plupart des modélisations nécessitent une phase d'estimation (par exemple toutes les méthodes utilisant des distributions). Là encore les informations supplémentaires peuvent intervenir.
3. Combinaison : cette étape concerne le choix d'un opérateur, compatible avec le formalisme de modélisation retenu, et guidé par les informations supplémentaires.
4. Décision : c'est l'étape ultime de la fusion, qui permet de passer des informations fournies par les sources au choix d'une décision d_i .

Nous ne donnons pas plus de détails sur ces étapes ici, car cela nécessiterait d'entrer dans les formalismes et les aspects techniques. Cela fait l'objet du chapitre 8.

La manière dont ces étapes sont agencées définit le système de fusion et son architecture. Dans le cas idéal la décision est prise à partir de tous les M_i^j , pour toutes les sources et toutes les décisions. C'est ce qu'on appelle la fusion globale. Dans ce modèle global, aucune information n'est négligée. La complexité de ce modèle et de sa mise en œuvre a conduit à développer des systèmes simplifiés, mais aux performances plus limitées [55].

Un deuxième modèle consiste ainsi à prendre d'abord des décisions locales au niveau de chaque source séparément. Dans ce cas, une décision $d(j)$ est prise à partir de toutes les informations issues de la source S_j seulement. Cette décision est dite décentralisée. Puis dans une deuxième étape, ces décisions locales sont fusionnées en une décision globale. Ce modèle s'impose lorsque les sources ne sont pas disponibles simultanément. Il permet des réponses rapides, grâce à des procédures spécifiques à chaque source, et peut être très facilement adapté à l'introduction de sources supplémentaires. Ce type de modèle s'appuie avec profit sur des techniques de contrôle adaptatif, et utilise souvent des architectures distribuées. Il est aussi appelé fusion de décisions [328, 96]. Son principal inconvénient provient de ce qu'il tient mal compte des relations entre capteurs et des corrélations ou dépendances possibles entre sources. De plus, ce modèle conduit très facilement à des décisions locales contradictoires ($d(j) \neq d(k)$ pour $j \neq k$) et la résolution de ces conflits implique des arbitrages de niveau supérieur, difficiles à optimiser, puisque l'information de départ n'est plus disponible. Des modèles de ce type sont souvent mis en œuvre dans des applications en temps réel par exemple dans le domaine militaire.

Un troisième modèle, « orthogonal » au précédent, consiste à combiner par une opération F tous les M_i^j relatifs à la même décision d_i , pour obtenir une forme fusionnée $M_i = F(M_i^1, M_i^2, \dots, M_i^m)$. Puis une décision est prise sur le résultat de cette combinaison. Ici, aucune décision intermédiaire n'est prise, et l'information est manipulée dans le formalisme choisi jusqu'à la dernière étape, diminuant ainsi les contradictions et les conflits. Ce modèle, comme le modèle global, est un modèle centralisé qui nécessite de disposer simultanément de toutes les sources. Plus simple que le modèle global, il est moins souple que le modèle distribué, rendant plus difficile l'ajout éventuel de sources d'informations.

Enfin, un modèle intermédiaire, hybride, consiste à choisir de manière adaptative les informations nécessaires pour un problème donné en fonction des spécificités des sources. Ce type de modèle copie souvent l'expert humain et fait intervenir des connaissances symboliques sur les sources et sur les objets. Il est donc très utilisé dans les systèmes à base de règles. Des architectures de type multi-agents sont bien adaptées à ce modèle.

6.5 Fusion en traitement d'images et fusion dans d'autres domaines

La fusion en image a des spécificités qui doivent être prises en compte à toutes les étapes de la construction d'un processus de fusion. Ces spécificités nécessitent également de modifier et de complexifier certains outils théoriques, souvent issus d'autres domaines. C'est typiquement le cas de l'information spatiale en fusion d'images ou en robotique. Ces spécificités seront détaillées pour le cas de la fusion en image dans le chapitre suivant.

La quantité des données à traiter et leur hétérogénéité sont souvent plus importantes que dans d'autres domaines (problèmes de combinaison d'opinions d'experts par exemple). Cela induit un niveau de complexité supplémentaire, à prendre en compte non seulement dans la modélisation, mais aussi dans les algorithmes.

Les données sont surtout objectives (données par des capteurs), ce qui les oppose aux données subjectives telles qu'elles peuvent être données par des individus. Elles contiennent cependant une certaine part de subjectivité (par exemple dans le choix des capteurs ou des sources d'informations, ou encore des paramètres d'acquisition). La subjectivité peut également intervenir dans l'expression des buts. Les données objectives sont en général dégradées, soit à cause de l'imperfection des systèmes d'acquisition, soit à cause des traitements qui leur sont appliqués.

En fait, une des difficultés principales vient du fait que les types de connaissances manipulées sont très hétérogènes. Il ne s'agit pas que de mesures et d'observations (qui elles-mêmes peuvent être hétérogènes), mais aussi de cas généraux, d'exemples typiques, de modèles génériques, etc.

Les différences principales par rapport à d'autres domaines d'application de la fusion d'informations viennent tout d'abord du fait que la question essentielle qui est posée (donc le but de la fusion) n'est pas la même. En traitement d'images, il s'agit essentiellement, selon la définition 1, d'améliorer la connaissance du monde (tel qu'il est). Cela suppose la notion d'existence d'une vérité, même si l'on n'en a qu'une vision partielle et déformée, ou difficile d'accès. Cela diffère de la fusion de préférences (comment on veut que le monde soit), de la fusion de réglementations (comment le monde devrait être), de problèmes de vote, où typiquement la vérité n'existe pas, etc. [51].

Chapitre 7

Fusion en traitement des images

Ce chapitre décrit les spécificités de la fusion en traitement d'images. Les définitions générales proposées dans le chapitre 6 sont reprises et discutées dans ce cadre particulier. Nous insistons sur le caractère propre des images et de leur représentation lors d'un problème de fusion, et nous insistons sur ce qui distingue la fusion en traitement des images de la plupart des autres domaines d'applications de la fusion.

7.1 Objectifs de la fusion en traitement des images

Bien sûr les images sont apparues très tôt comme des sources d'information importantes de systèmes existant de fusion de l'information et des systèmes de fusion de données ont utilisé des images. Prenons par exemple une application de suivi global de situations écologiques. On y fait appel à des images de télédétection pour donner des information météorologiques. La donnée issue de l'image peut alors être intégrée à un modèle physique par l'intermédiaire d'une estimation, en chaque pixel, du couvert nuageux. On peut introduire dans une équation de bilan thermodynamique la teneur en vapeur d'eau ainsi estimée en chaque point de l'image.

Ce n'est pourtant pas ce type d'application qui a donné naissance au domaine original de la **fusion d'images**. On doit plutôt rechercher dans la pratique des experts en interprétation des images, dans des milieux très différents de la société, les modèles dont les traiteurs d'images ont cherché à s'inspirer. Donnons deux exemples, mais nous pourrions les multiplier et les diversifier aisément.

Un radiologue en milieu hospitalier prend ses décisions à partir de nombreux clichés qui donnent des points de vue différents sur une même structure anatomique. Ces clichés sont souvent étalés sur un négatoscope et le médecin construit son opinion à partir d'un examen alterné et attentif de plusieurs vues. À partir d'une hypothèse, issue de l'un des clichés, il confirme son interprétation sur d'autres, la précise par recouplements et addition de vues différentes, ou au contraire l'écarte par des informations contradictoires issues d'une de ces vues. Cette situation est très classique en imagerie médicale, domaine où les techniques d'acquisition se diversifient : rayons X, résonance magnétique, imagerie nucléaire, imagerie ultrasonore, chacune pouvant se décliner selon des modalités diverses en fonction des protocoles d'acquisition. Elle est confortée également par les efforts déployés dans toutes les structures hospitalières pour regrouper les sources d'images dans un même service ou faire

converger ces images sur une même console au lieu où s'élabore le diagnostic. Ces efforts ont conduit à l'introduction progressive des systèmes d'archivage et de consultation intégrés à l'hôpital (les PACS des anglo-saxons¹).

Pour second exemple, nous choisirons un expert de télédétection, chargé d'interpréter une scène complexe. Il dispose d'un nombre important d'images, issues de capteurs variés, par exemple des images du domaine visible dans diverses gammes de longueur d'onde, ou des images infra-rouge ou encore des images de radar. Chaque source l'informe sur un aspect particulier de la scène et lui permet donc d'élaborer des scénarios. Là encore, l'expertise se fait par une confrontation des diverses représentations, leur combinaison permettant tantôt de conforter le jugement, tantôt de l'infirmier. Cette expertise est le résultat d'un long apprentissage et devient de plus en plus complexe lorsque les sources d'images se diversifient et leur nombre s'accroît. Pourtant, autant pour les applications satellitaires où se complètent de nombreux capteurs, que pour les applications aéroportées où l'on tire profit de sources très variées (cartes, cadastres, plans d'occupation des sols, relevés géologiques ou agronomiques, modèles altimétriques, etc.), l'usage s'engage résolument vers une complexité croissante des sources d'images.

Dans ce cadre, la fusion d'images apparaît donc comme une tâche singulière et différente de la fusion de données. En effet, il n'est pas clair qu'il soit possible de concevoir un cadre fonctionnel dans lequel chaque information prendrait place comme c'était le cas lorsque l'image satellitaire permettait d'introduire la mesure de « couvert nuageux » dans un schéma d'ensemble mettant en jeu par exemple l'évapo-transpiration du couvert végétal et les conditions météo. Dans le cadre le plus général du traitement d'images, la fusion des images intervient pour aider à la prise de décision dans un schéma complexe, et mal formalisé la plupart du temps, où les diverses images apportent une part de « vérité » qui contribue, en collaboration et en opposition avec les autres sources, à une interprétation d'ensemble. Nous souhaitons donc, en développant des méthodes automatiques de fusion d'images, préparer et raccourcir l'étape d'élaboration de l'expertise humaine et, éventuellement, dans les situations de profusion des sources d'images, gérer la multiplicité sans sacrifier les apports potentiels d'une combinatoire complexe, au-delà même de ce que les experts humains savent aujourd'hui en tirer.

La tâche de fusion en traitement des images est donc très étroitement liée à la prise de décision. Au contraire, elle a peu à voir avec l'étape qu'on lui associe souvent de recalage géométrique des images, mais cette étape est généralement considérée aujourd'hui comme incontournable et beaucoup de travaux en fusion d'image se contentent de l'effectuer puis de laisser à un opérateur humain le soin de mettre en œuvre la prise de décision.

Rappelons que l'objectif du recalage est de superposer exactement les pixels correspondant à un même objet observé dans les diverses images. Cette étape peut être facilitée s'il existe un référentiel absolu reconnu pour décrire la scène, c'est le cas par exemple dans les applications de cartographie ou de géographie qui exploitent les référentiels géocodés, ainsi que pour certaines applications médicales pour lesquelles des repères anatomiques conventionnels ont été établis. Les techniques de recalage sont nombreuses et utilisent des principes variés : corrélation, programmation dynamique, flot optique, déformation élastique, ... (on pourra consulter [223] et [227] pour une synthèse des méthodes en imagerie aérienne et satellitaire et en imagerie médicale respectivement).

¹Picture Archiving and Communication Systems (systèmes d'archivage et d'échange d'images)

Pour quels objectifs est on amené à faire de la fusion d'image ? Tout d'abord pour améliorer les trois tâches principales de la reconnaissance des formes, détecter, reconnaître et identifier :

Détection : il s'agit dans ce cas de valider l'hypothèse de présence ou d'absence de l'objet cherché (présence d'un véhicule sur une route, ou d'une sténose dans un vaisseau sanguin). Cet objectif de détection est parfois combiné à celui de suivi des objets détectés dans une séquence de vues.

Reconnaissance ou classification : un objet détecté est alors associé à l'une des familles d'objets connus ou attendus en fonction de ses critères photométriques, géométriques ou morphologiques. Cette opération peut être conduite sur des objets à des niveaux très variés, depuis le pixel jusqu'à des ensembles complexes de composantes de l'image.

Identification : un objet détecté et reconnu est identifié lorsqu'il est associé à un prototype unique de sa classe. Ainsi, après avoir détecté un véhicule en imagerie infrarouge, la reconnaissance permettra d'en déduire son type : camion, moto ou voiture, et l'identification conclura au camion du laitier, objet classiquement surveillé dans ce type d'image ...

Mais d'autres applications que celles de la reconnaissance des formes exigent aussi la mise en œuvre de méthodes de fusion. Ces opérations peuvent intervenir au cours du processus de reconnaissance, mais à un stade plus préliminaire et ne débouchent pas nécessairement sur une prise de décision.

Segmentation : il s'agit là d'un objectif plus focalisé que la classification, en cherchant à extraire le plus précisément possible des objets déterminés. Il peut s'agir simplement d'utiliser la complémentarité des sources d'information pour mieux identifier les limites des composantes homogènes de l'image. Par exemple, en fusionnant le tracé précis fourni par une image aérienne, et la structure topologique du réseau et sa géométrie approximative données par une carte, on peut fournir une excellente description d'un réseau routier.

Reconstruction : la multiplicité des points de vue est favorable à la reconstruction tridimensionnelle des scènes observées, et, si cette reconstruction peut prendre des formes classiques (comme en stéréovision, ou en tomographie), dans d'autres situations où les conditions d'acquisition sont moins bien contrôlées, il n'est possible de reconstruire qu'une information tridimensionnelle approchée qui combine de façon plus ou moins empirique les divers aspects disponibles.

Détection de changements : ce type de décision concerne typiquement les images acquises à des dates différentes, qu'il s'agisse d'une carte et d'une image, ou d'images multi-dates pour le suivi des cultures ou d'une pathologie. Il peut s'agir également de séquences d'images multi-sources (de cadence plus élevée que les images multi-dates).

Mise à jour de la connaissance sur un phénomène ou une scène : ici, contrairement au cas précédent, la décision consiste à utiliser les informations provenant de différentes sources (éventuellement multi-dates) pour modifier ou compléter une connaissance précédente, par exemple compléter un réseau routier avec les nouveaux ronds-points pour mettre à jour une carte.

Parmi ces différents problèmes de décision, certains sont proches de la combinaison d'experts. En effet, chaque image peut être considérée comme un expert donnant son opinion selon ses compétences. Mais en général l'information dans les problèmes de fusion d'experts est plus éparses qu'en image. L'apprentissage est donc plus difficile car les données sont moins

nombreuses, mais en revanche l'utilisateur a souvent moins de contraintes sur les coûts algorithmiques des méthodes. En image, la quantité des données à fusionner est à la fois un avantage pour l'apprentissage et un inconvénient pour les charges de calcul.

Si l'on compare la problématique de fusion d'image aux problèmes de fusion de données par agrégation et optimisation multi-critères, on constate qu'une différence essentielle réside en ce que dans ces derniers problèmes on cherche une solution satisfaisant au mieux un ensemble de contraintes plus ou moins souples, tandis qu'en traitement des images, chaque source procure (de manière plus ou moins explicite) un degré de satisfaction (de l'appartenance à une classe par exemple, qui peut alors être considérée comme un critère) et la décision consiste plutôt à choisir la meilleure (meilleure classe par exemple).

7.2 Les situations de fusion

Selon les applications, les problèmes de fusion peuvent se produire dans des situations différentes, dans lesquelles les types d'informations ne sont pas les mêmes. Les principales situations de fusion en traitement d'images sont les suivantes.

Plusieurs images du même capteur : il s'agit par exemple de plusieurs canaux du même satellite, d'images multi-échos en IRM, ou encore de séquences d'images pour des scènes en mouvement. Les données sont alors relativement homogènes car elles correspondent à des mesures physiques similaires.

Plusieurs images de capteurs différents : c'est le cas le plus fréquent, où les principes physiques différents des capteurs permettent d'avoir des points de vue complémentaires sur la scène. Il peut s'agir d'images ERS et SPOT, d'images IRM et ultrasonores, etc. L'hétérogénéité est alors beaucoup plus importante, les différents capteurs ne parlant pas du même aspect du phénomène. Les images donnent chacune une vision partielle et ne sont pas informatives sur les caractéristiques auxquelles elles ne sont pas dédiées (par exemple une IRM anatomique ne donne pas d'information fonctionnelle, et une image TEP a une résolution trop limitée pour que l'anatomie y soit vue de manière suffisamment précise).

Plusieurs informations extraites d'une même image : il s'agit de situations dans lesquelles on extrait divers types d'informations d'une image à l'aide de plusieurs détecteurs, opérateurs, classificateurs, etc., qui s'appuient sur des caractéristiques différentes des données, cherchent à extraire des objets différents, rendant ainsi les informations à fusionner souvent très hétérogènes. Les informations extraites peuvent concerter le même objet (fusion de détecteurs de contours par exemple) ou des objets différents et on cherche alors une interprétation globale de la scène et une cohérence entre les objets. Elles peuvent être de niveaux différents (très locales ou plus structurelles quand on s'intéresse aux relations spatiales entre objets).

Images et autre source d'information : par autre source d'information, on entend par exemple un modèle, qui peut être soit particulier comme une carte, soit générique comme un atlas anatomique, des bases de connaissances, des règles, des informations issues d'experts, etc. Les informations sont à nouveau de types très différents, à la fois dans leur nature, et dans leur représentation initiale (images quand il s'agit d'une carte ou d'un atlas numérisé, mais aussi descriptions linguistiques, bases de données, etc.).

7.3 Caractéristiques des données en fusion d'images

Les spécificités de la fusion en traitement d'images ne permettent que difficilement d'exploiter les progrès réalisés dans d'autres domaines de la fusion d'informations. Une des raisons est la complexité des données et des connaissances, qui rendent impossible la recherche d'un système global pour regrouper en une relation unique toutes les composantes de l'image.

La complexité vient d'une part du volume d'informations à traiter (par exemple, une seule image IRM du cerveau occupe un volume de 8 à 16 mega-octets). Ces gros volumes de données, sur lesquels des apprentissages statistiques sont souvent possibles, sont une des raisons qui expliquent la large utilisation de méthodes probabilistes et statistiques en fusion d'images.

La complexité vient d'autre part de la forte hétérogénéité des informations à combiner, qu'il s'agisse d'images issues de capteurs différents, d'images et de modèles, ou de caractéristiques différentes extraites d'une ou plusieurs images. Cette hétérogénéité est présente à la fois dans la nature des informations et dans leur représentation. Les données peuvent être fréquentes, par exemple lorsqu'il s'agit des cas typiques rencontrés dans une application donnée, pour lesquels il est possible d'avoir des données statistiques par exemple. Elles peuvent aussi être rares (par exemple des images pathologiques) et dans ce cas il est beaucoup plus difficile de les modéliser de manière statistique. La combinaison des deux types de données est fréquemment rencontrée en traitement d'images. De plus, elles peuvent être soit factuelles (typiquement, une photographie d'une scène à un instant donné) ou génériques (modèle, règles, connaissances générales sur l'application). Combiner des informations de spécificités différentes pose souvent des problèmes de conflit à résoudre. En traitement d'images, ce n'est pas une tâche facile, car l'information factuelle n'est pas toujours suffisamment sûre et précise pour qu'on puisse systématiquement lui donner la priorité sur des informations moins spécifiques et plus génériques qui peuvent admettre des exceptions.

La combinaison des informations est souvent guidée ou contrainte par des informations supplémentaires, sur l'information à combiner, sur le contexte et le domaine d'application). C'est aussi une source de forte hétérogénéité. Parmi les exemples d'informations supplémentaires sur l'information on peut citer la fiabilité d'une source, soit globale, soit conditionnelle aux objets observés. Ce cas est très souvent rencontré en classification d'images multi-sources, où une image peut être fiable pour une classe mais pas pour une autre. Donnons maintenant quelques exemples d'informations supplémentaires sur le domaine et le contexte :

- les fleuves sont sombres dans le canal XS3 de SPOT (information liant le type d'acquisition à une observation),
- le LCR est sombre dans les images IRM en T1,
- les routes se croisent pour former des carrefours (contrainte d'intégrité).

Ces informations génériques sont utilisées pour guider le processus de fusion. Le dernier exemple est un cas typique de règle avec exception. La règle donne le cas le plus général, mais elle n'est plus vraie dans le cas d'impasses par exemple.

La fusion active est un des moyens pour réduire la complexité en choisissant à chaque instant les meilleures informations à fusionner. Ce choix peut être effectué en fonction d'un résultat partiel de fusion obtenu dans une étape antérieure, de mesures d'informations, d'informations externes susceptibles de guider la fusion, de l'identification d'ambiguités à lever, etc.

Les informations fusionnées en traitement d'images sont fortement entachées d'imperfections (incertitude, imprécision, incomplétude, ambiguïté, conflit etc., selon la distinction proposée dans le chapitre 6). Ces imperfections trouvent leur origine à différents niveaux, depuis les phénomènes observés jusqu'aux traitements. Par exemple, la transition douce entre tissus sains et pathologiques est une imprécision due au phénomène physiologique. De la même façon, des caractéristiques similaires entre deux tissus différents se retrouvent sur les images mesurant ce type de caractéristique et se traduisent par une incertitude sur l'appartenance d'un point particulier à un tissu ou à un autre, cette incertitude étant due à la fois au phénomène et au capteur. La délocalisation de l'information spatiale, due au regroupement en un même pixel de l'information contenue dans tout un volume, est due au capteur et à sa résolution, et constitue une imprécision sur la localisation de l'information sur l'image (effet de volume partiel). Les phénomènes de Gibbs au niveau des transitions nettes, qui apparaissent en IRM ou en imagerie radar par exemple, sont des sources d'imprécisions dues aux algorithmes de reconstruction numérique des images. La représentation d'informations (symboliques) sous forme schématique (par des cartes ou des atlas) est source à la fois d'imprécision et d'incertitude. Imprécisions et incertitudes sont ensuite renforcées dans les primitives extraites des images et sur lesquelles s'appuie la fusion. L'exemple le plus familier est celui de la détection de contours par des filtrages gaussiens à différentes échelles : en augmentant l'écart type de la gaussienne, on gagne en certitude sur la présence de contours mais on perd en précision sur leur localisation. Cet antagonisme entre précision et certitude a bien été identifié comme un trait caractéristique de la démarche en reconnaissance des formes [308]. De cet antagonisme naissent souvent des contradictions en fusion d'images puisque l'on dispose de plusieurs mesures sur un même événement : si ces données sont précises, alors elles sont probablement incertaines, et elles risquent donc d'être en contradiction les unes avec les autres ; si l'on renforce leur certitude, cela se fait souvent au prix de plus d'imprécision, rendant les données peu informatives si cette imprécision devient trop importante. Il conviendra donc qu'un système de décision en fusion gère explicitement incertitude et imprécision pour éviter les incohérences.

L'imprécision n'est pas une caractéristique propre aux données, mais elle peut également être attachée aux objectifs et aux buts, en particulier lorsqu'ils sont exprimés sous forme linguistique vague.

Enfin, le caractère spatial de l'information, spécifique au traitement d'images, mérite une attention particulière. Son introduction dans les méthodes de fusion, souvent inspirées d'autres domaines dans lequel ne figure pas ce caractère spatial, n'est pas immédiate mais pourtant nécessaire afin de garantir la cohérence spatiale des résultats. L'imprécision est également présente à ce niveau. À bas niveau, il s'agit de problèmes de recalage, ou de volume partiel par exemple. À plus haut niveau, il s'agit par exemple de relations entre objets qui peuvent être intrinsèquement vagues et mal définies (relations du type « à gauche de » par exemple).

Comme dans les autres applications de la fusion, la redondance et la complémentarité entre les images à fusionner sont des atouts pour réduire les imperfections telles que l'incertitude et l'imprécision, pour lever des ambiguïtés, pour compléter l'information, pour résoudre des conflits. Voici quelques exemples de complémentarité en fusion d'images :

- sur l'information elle-même : parties cachées qui peuvent être différentes dans des images de profondeur ou des images aériennes,
- sur le type d'information : information anatomique ou information fonctionnelle pour le même sujet dans des modalités d'imagerie différentes,

- qualité de l'information : deux images de même type mais avec des paramètres d'acquisition différents peuvent donner des informations de qualités différentes pour différentes structures.

La redondance provient de manière plus évidente du fait que ce sont des images de la même scène que l'on cherche à combiner. Pour certains problèmes de fusion, comme les études de groupe en imagerie fonctionnelle, la redondance (quelles sont les zones activées chez tous les sujets) et la complémentarité (où sont les différences) sont elles-mêmes sujet d'étude.

Le conflit est un problème très délicat, comme dans d'autres applications de la fusion d'informations, comme cela a été discuté dans le chapitre 6. En image, les exemples de conflit qui ne sont qu'apparents et facilement confondus avec la complémentarité se trouvent dans les cas où une image n'est pas capable de différencier deux classes alors qu'une autre les distingue. Les imprécisions et incertitudes sont également sources de conflit. Par exemple la mauvaise localisation d'un contour peut provoquer un conflit entre plusieurs détecteurs de contours. Les conflits dus à des spécificités différentes des informations à combiner sont fréquents dans les applications de fusion d'images et de modèles. Par exemple la reconnaissance de structures cérébrales par fusion d'images IRM et de données issues d'un atlas anatomique doit faire face à la variabilité entre les individus, souvent non représentée dans l'atlas, ou encore à la présence possible de pathologies dans des images de patient, non présentes dans le modèle générique. Des problèmes similaires se rencontrent dans les problèmes de fusion d'images aériennes et satellites et de cartes numériques. Ici les conflits peuvent provenir du tracé imprécis de la carte, de modifications dans la scène non portées sur une carte plus ancienne, etc.

7.4 Contraintes

Les contraintes spécifiques au traitement d'images à prendre en compte dans un processus de fusion sont de plusieurs ordres.

Du point de vue de l'architecture du système de fusion, il est rare que les systèmes décentralisés soient imposés. Le cas le plus fréquent est celui de la fusion « off-line » où l'on dispose simultanément de toutes les informations. Des systèmes centralisés peuvent alors être utilisés.

Les contraintes de temps réel sont assez peu fréquentes, sauf dans les applications de surveillance ou multi-média, dans lesquelles elles sont amenées à prendre une place grandissante.

En revanche, les contraintes de cohérence spatiale sont très fortes et constituent un point important des recherches en fusion d'images. De plus en plus de travaux sont donc consacrés à la prise en compte de l'information spatiale, soit au niveau local via le contexte spatial, soit à un niveau plus structurel via des relations spatiales entre les structures ou les objets de la scène.

Le fait que les données soient volumineuses peut imposer des contraintes de temps de calcul. Ainsi les manipulations au niveau des pixels sont limitées à des opérations simples. Des opérations plus sophistiquées nécessitent souvent l'introduction d'informations de plus haut niveau, et une représentation plus structurelle de l'information.

La complexité et la quantité des données à manipuler impose souvent que soit effectué

un choix dans les informations et les connaissances à fusionner. Ce choix est bien sûr guidé principalement par des critères de pertinence par rapport aux objectifs de décision, mais aussi par des critères portant sur la facilité d'accès et de représentation des informations et des connaissances, et sur leur qualité.

Enfin, un problème crucial est celui de l'évaluation. Dans les problèmes de fusion d'images, la « vérité » ou la « bonne » solution existe en général plus ou moins, mais elle est le plus souvent difficile d'accès. Ainsi l'évaluation et la validation d'une méthode de fusion ne peut se faire que sur des simulations, des acquisitions sur fantômes, ou par comparaison avec une décision manuelle. Cette situation s'oppose à celle rencontrée dans les problèmes de vote ou de choix social, dans lesquels la vérité n'existe pas, mais on cherche à trouver une « meilleure » solution au sens d'un compromis, ou de critères d'équité et d'éthique.

7.5 Aspects numériques et symboliques en fusion d'images

Si le côté numérique de l'information manipulée en traitement d'images est évident, le côté symbolique mérite qu'on s'y arrête un peu plus.

Les informations symboliques peuvent être attachées aux données à fusionner (par exemple une information graphique dans une carte ou un atlas anatomique, des attributs calculés sur les données ou sur des objets extraits préalablement des données), ou attachés à la connaissance du domaine. Typiquement, les informations sur le domaine sont souvent représentées par des règles, des représentations structurelles telles que des graphes, souvent utilisés en reconnaissance des formes dans les images, des contraintes à prendre en compte dans les algorithmes. Une information structurelle peut par exemple spécifier qu'un réseau routier peut être représenté par un graphe en utilisant des routes et des carrefours, ou exprimer sous forme propositionnelle des règles générales sur la scène telles que « les ventricules cérébraux sont toujours à l'intérieur de la matière blanche », etc. Les informations structurelles peuvent également être représentées de manière iconique, donc proche des données images, par exemple dans le cas de cartes numérisées ou d'atlas anatomiques. On combine alors la représentation géométrique des structures et la nature ou la sémantique de ces structures. Mais les informations sur le domaine peuvent également purement numériques quand il s'agit par exemple de caractéristiques de l'acquisition comme les longueurs d'ondes utilisées en imagerie satellitaire, ou les temps d'acquisition en imagerie médicale. Des représentations hybrides, dans lesquelles des nombres sont utilisés comme des symboles avec une quantification, sont utilisées en fusion d'images, pour quantifier la qualité d'un détecteur, pour évaluer une donnée symbolique ou la confiance dans une mesure, la fiabilité d'une image pour certaines classes ou structures, etc. Du point de vue des traitements, cet aspect hybride se manifeste par exemple lorsqu'une proposition établissant que la reconnaissance d'une structure ne dépend que du contexte local conduit à une modélisation appropriée dans un cadre markovien. Les représentations symboliques sont ainsi utilisées en tant que connaissance a priori ou connaissance contextuelle ou générique pour guider un traitement numérique. Elles servent aussi de support structurel par exemple en fusion d'images et de cartes [248].

Chapitre 8

Méthodes numériques de fusion d'informations

Dans ce chapitre, nous présentons les trois approches numériques principales de la fusion d'informations : les approches probabilistes et bayésiennes sont les plus répandues et sont décrites dans la partie 8.1 ; la théorie des fonctions de croyances est présentée dans la partie 8.2 ; puis l'utilisation pour la fusion de la théorie des ensembles flous et des possibilités introduite dans le chapitre 4 est décrite dans la partie 8.3. Puisqu'une des caractéristiques principales de la fusion d'images est l'importance de l'information spatiale, nous examinerons dans la section 8.4 comme elle peut être introduite dans ces méthodes de fusion.

8.1 Fusion probabiliste et bayésienne

Les approches probabilistes traitent essentiellement de l'incertitude de l'information. Elles reposent sur des théories mathématiques solides et bien maîtrisées en traitement du signal et des images, telles que la théorie bayésienne de la décision, la théorie de l'estimation, les mesures entropiques, etc., ce qui en fait des outils privilégiés pour la fusion.

Les informations et leurs imperfections (essentiellement celles dont la nature s'exprime en termes d'incertitude) sont modélisées par des distributions de probabilités ou de mesures statistiques. Nous verrons dans la section 8.1.1 comment ce formalisme permet de mesurer l'information. Puis nous aborderons les différentes étapes du processus de fusion : la modélisation et l'estimation dans la section 8.1.2, la combinaison bayésienne dans la section 8.1.3, puis la combinaison bayésienne vue comme un problème d'estimation dans la section 8.1.4. Les règles les plus classiques de décision sont présentées dans les sections 8.1.5 et 8.1.6. Les sections suivantes présentent des exemples d'applications et d'autres outils théoriques y sont détaillés, dans les domaines de la classification multisources en traitement d'images à la section 8.1.7.

8.1.1 Mesures d'information

Lorsque l'on dispose d'un ensemble de l sources d'information I_j , une première tâche consiste souvent à le transformer en un sous-ensemble plus réduit, donc de traitement plus

simple, sans perdre d'information.

L'approche de l'analyse en composantes principales, qui projette chaque source d'information sur les vecteurs propres de la matrice de covariance, est souvent employée et permet ainsi d'obtenir l nouvelles sources décorrélées, classées par ordre décroissant d'énergie. Une troncature aux l' ($l > l'$) premières sources permet souvent de conserver l'essentiel de l'énergie de l'ensemble d'origine.

Mais en pratique cette méthode montre très vite ses limites en traitement d'images par exemple, car elle ne permet de prendre en compte ni les dépendances complexes entre images, ni les variations spatiales de dépendance.

Pour exprimer l'apport d'information dû à l'ajout d'une nouvelle source d'information I_{l+1} à un ensemble déjà connu $\{I_1, \dots, I_l\}$, on préfère l'approche proposée par Shannon et reposant sur les notions d'information et d'entropie [203, 224]. A partir de la probabilité jointe des l premières sources $p(I_1, \dots, I_l)$ (estimée le plus souvent par dénombrement, par exemple à partir de l'histogramme multidimensionnel des niveaux de gris pour des images), on définit l'entropie (ou information moyenne par pixel dans le cas des images) des l premières sources par :

$$H(I_1, \dots, I_l) = - \sum p(I_1, \dots, I_l) \log p(I_1, \dots, I_l), \quad (8.1)$$

et l'entropie apportée par la $(l + 1)^{\text{ème}}$ source s'exprime, soit en fonction des entropies, soit en fonction des probabilités par :

$$\begin{aligned} H(I_{l+1} | I_1, \dots, I_l) &= H(I_1, \dots, I_{l+1}) - H(I_1, \dots, I_l) \\ &= - \sum p(I_1, \dots, I_{l+1}) \log p(I_{l+1} | I_1, \dots, I_l). \end{aligned} \quad (8.2)$$

Pour deux sources, on définit ainsi la redondance¹ entre elles par :

$$R(I_1, I_2) = H(I_1) + H(I_2) - H(I_1, I_2), \quad (8.3)$$

et la complémentarité de la source I_2 par rapport à la source I_1 , c'est-à-dire la quantité moyenne d'information qu'il faut ajouter à I_2 pour retrouver I_1 :

$$C(I_1 | I_2) = H(I_1 | I_2), \quad (8.4)$$

ce qui conduit à la relation :

$$H(I_1) = R(I_1, I_2) + C(I_1 | I_2). \quad (8.5)$$

Des approches analogues peuvent être envisagées dans un cadre non probabiliste, en s'appuyant par exemple sur l'entropie floue [219]. Le formalisme est pour l'instant moins développé dans cette direction.

En fusion, on utilisera typiquement des sources fortement redondantes pour confirmer une décision incertaine et des images complémentaires pour élargir le champ des décisions. Des sources complémentaires peuvent conduire soit à des décisions conflictuelles soit à des décisions consensuelles.

En traitement d'images, la notion d'entropie a été élargie pour caractériser non seulement la dispersion des mesures sur l'espace de mesure, mais également la cohérence spatiale des

¹Cette redondance ne peut malheureusement pas être étendue à plus de deux sources sans perdre potentiellement sa propriété de positivité.

mesures en prenant en compte des probabilités d'occurrence de configurations particulières de pixels, soit dans le cadre de classifications [225, 224], soit dans celui de champs markoviens [336, 344].

Les notions d'entropie globale ne sont pas toujours bien adaptées aux problèmes de fusion, et des notions d'entropie conditionnelle aux classes à reconnaître par exemple sont souvent préférables : elles permettent une analyse plus fine de l'information qu'apporte chaque source pour chaque classe et sont donc mieux adaptées aux problèmes pour lesquels une source est meilleure pour certaines classes et moins bonne pour d'autres. Bien que la définition formelle de tels concepts ne pose pas de difficulté particulière, ils ne sont pas encore beaucoup utilisés en fusion et c'est vraisemblablement un point qui mériterait d'être approfondi.

8.1.2 Modélisation et estimation

La théorie la plus exploitée dans la littérature est de loin la théorie des probabilités, associée à la théorie bayésienne de la décision [135]. L'information y est modélisée par une probabilité conditionnelle, par exemple, la probabilité pour qu'un pixel appartienne à une classe particulière, étant donné les images disponibles. Ainsi, la mesure introduite dans la section 6.4 s'écrit-elle sous la forme :

$$M_i^j(x) = p(x \in C_i | I_j). \quad (8.6)$$

Cette probabilité est calculée à partir de caractéristiques $f_j(x)$ de l'information extraites des sources. Par exemple, en image, il peut s'agir dans les cas les plus simples du niveau de gris du pixel considéré ou d'informations plus complexes nécessitant des traitements préliminaires. L'équation 8.6 ne dépend alors plus de toute la source I_j mais s'écrit sous la forme simplifiée :

$$M_i^j(x) = p(x \in C_i | f_j(x)). \quad (8.7)$$

En traitement des images, en l'absence de modélisations fonctionnelles fortes des phénomènes observés, les probabilités $p(f_j(x) | x \in C_i)$, ou plus généralement $p(I_j | x \in C_i)$ (qui représente la probabilité, conditionnelle à la classe C_i , de l'information fournie par la source I_j), sont apprises par dénombrement sur des zones de test (ou par apprentissage sur ces zones des paramètres d'une loi donnée) et on en déduit les probabilités des équations 8.6 et 8.7 par application de la règle de Bayes.

8.1.3 Combinaison dans un cadre bayésien

Dans le modèle bayésien, la fusion peut être effectuée de manière équivalente à deux niveaux :

- soit au niveau de la modélisation, et l'on calcule alors des probabilités de la forme :

$$p(x \in C_i | I_1, \dots, I_l), \quad (8.8)$$

à l'aide de la règle de Bayes :

$$p(x \in C_i | I_1, \dots, I_l) = \frac{p(I_1, \dots, I_l | x \in C_i) p(x \in C_i)}{p(I_1, \dots, I_l)}, \quad (8.9)$$

où les différents termes sont estimés par apprentissage ;

- soit par la règle de Bayes elle-même, où l'information issue d'un capteur vient mettre à jour l'information sur x estimée d'après les capteurs précédents (c'est la seule forme utilisable si les informations sont disponibles successivement et non simultanément) :

$$p(x \in C_i | I_1, \dots, I_l) = \frac{p(I_1 | x \in C_i) p(I_2 | x \in C_i, I_1) \dots p(I_l | x \in C_i, I_1, \dots, I_{l-1}) p(x \in C_i)}{p(I_1) p(I_2 | I_1) \dots p(I_l | I_1, \dots, I_{l-1})}.$$

Très souvent, étant données la complexité de l'apprentissage à partir de plusieurs capteurs et la difficulté d'obtenir des statistiques suffisantes, ces équations sont simplifiées sous l'hypothèse d'indépendance. Là encore, des critères ont été proposés pour vérifier la validité de ces hypothèses. Les formules précédentes deviennent alors :

$$p(x \in C_i | I_1, \dots, I_l) = \frac{\prod_{j=1}^l p(I_j | x \in C_i) p(x \in C_i)}{p(I_1, \dots, I_l)}. \quad (8.10)$$

Cette équation fait apparaître clairement le type de combinaison des informations, sous la forme d'un produit, donc une fusion conjonctive. Il est notable que la probabilité *a priori* joue exactement le même rôle dans la combinaison que chacune des sources auxquelles elle est combinée également par un produit.

8.1.4 Combinaison vue comme un problème d'estimation

Une autre manière de voir la fusion probabiliste consiste à considérer que chaque source donne une probabilité (d'appartenance à une classe) par exemple, et que la fusion consiste à combiner ces probabilités pour trouver la probabilité globale d'appartenance à la classe. Cette vision revient à considérer la fusion comme un problème d'estimation, et permet d'utiliser des opérateurs de combinaison différents du produit. En particulier les méthodes de moyenne ou moyenne pondérée, de médiane, de consensus sont souvent employées [88, 89, 145]. Des estimateurs robustes peuvent également être employés, afin de limiter ou supprimer l'influence des valeurs aberrantes (« outliers »). Enfin, des méthodes issues de la théorie des variables régionalisées [233], telles que le krigeage ou le krigeage universel, pourraient également être utilisées dans ce cadre.

8.1.5 Décision

La dernière étape concerne la décision, par exemple le choix de la classe à laquelle appartient un point. Cette décision binaire peut être assortie d'une mesure de la qualité de cette décision, pouvant éventuellement conduire à la rejeter. La règle la plus utilisée pour la décision probabiliste et bayésienne est le maximum *a posteriori* :

$$x \in C_i \text{ si } p(x \in C_i | I_1, \dots, I_l) = \max \{p(x \in C_k | I_1, \dots, I_l), 1 \leq k \leq n\},$$

mais de très nombreux autres critères ont été développés par les probabilistes et les statisticiens, pour qu'ils s'adaptent au mieux aux besoins de l'utilisateur et au contexte de sa décision : maximum de vraisemblance, maximum d'entropie, marginale maximale, espérance maximale, risque minimal, etc. Cependant, la grande variété de ces critères laisse l'utilisateur à nouveau démunie devant la justification d'un choix et l'éloigne de l'objectivité recherchée initialement par ces méthodes.

8.1.6 Autres approches en détection

Le domaine de la détection par fusion multicapteurs a été largement étudié et a donné lieu à plusieurs approches. On distingue la détection centralisée dans laquelle les mesures données par les différents capteurs sont considérées comme un vecteur sur lequel est prise la décision, et la détection décentralisée dans laquelle chaque capteur donne une réponse binaire (détection ou non), ces réponses étant ensuite combinées par un opérateur de fusion.

Dans le premier cas [341], les règles de décision utilisées reposent souvent sur le risque moyen, le maximum de gain, le minimum de risque, tous issus de l'approche bayésienne, mais aussi sur des critères tels que celui de Neyman-Pearson, consistant à maximiser la probabilité de détection pour une probabilité de fausse alarme donnée. Cela suppose implicitement que la fausse alarme est considérée comme l'erreur la plus grave, ce qui n'est pas toujours le cas suivant les applications (par exemple dans le cas du déminage humanitaire, c'est la non-détection qui est l'erreur la plus grave).

Dans le deuxième cas, si l'on dispose de l capteurs, chacun donnant une réponse binaire, l'opérateur de fusion est considéré comme une fonction logique de ces l réponses (qui constituent l'entrée de l'opérateur). L'ensemble des opérateurs possibles est de cardinal très élevé (2^{2^l}), ce qui rend les méthodes reposant sur une énumération exhaustive des possibilités impraticables dès que l dépasse quelques unités. Cependant, le nombre de possibilités est fortement réduit par la contrainte de monotonie que doit satisfaire l'opérateur [329]. Parmi les méthodes les plus intéressantes dans ce domaine, citons les méthodes entropiques pour l'optimisation de l'opérateur de fusion [111]. On aboutit ainsi à une règle de fusion optimale qui s'exprime comme une somme pondérée des décisions locales, que l'on compare à un seuil fonction des probabilités de fausse alarme et de détection des différents détecteurs locaux, des probabilités *a priori* et des coûts. Les méthodes entropiques peuvent ainsi être utilisées à plusieurs niveaux : pour choisir les capteurs les plus pertinents, pour optimiser les détecteurs locaux (au niveau de chaque capteur) et pour optimiser l'opérateur de fusion.

8.1.7 Un exemple de fusion bayésienne en imagerie satellitaire

Nous illustrons dans cette section la fusion bayésienne par un exemple simple de classification multisource en imagerie satellitaire, dans lequel la fusion est effectuée au niveau des pixels, à partir de l'information de niveaux de gris. Cet exemple a été développé dans [77]. La figure 8.1 montre un exemple des six images à fusionner. Il s'agit d'images SPOT en mode spectral multibande XS dans le vert (XS1), le rouge (XS2) et le proche infrarouge (XS3), avec un pas d'échantillonnage de 20 mètres, recalées dans un référentiel commun (ce qui permet la fusion au niveau du pixel).

Les classes considérées sont les villes ou zones urbaines (classe C_1), les fleuves (classe C_2) et une classe C_3 comportant toutes les autres structures (essentiellement des zones de végétation).

La caractéristique principale des villes sur ces images étant leur texture, l'ensemble des trois images initiales est complété par trois images de texture obtenues par un algorithme d'estimation des paramètres d'un champ markovien gaussien [109]. Ces images de texture sont également montrées figure 8.1.

L'apprentissage des probabilités conditionnelles est effectué à partir de l'histogramme

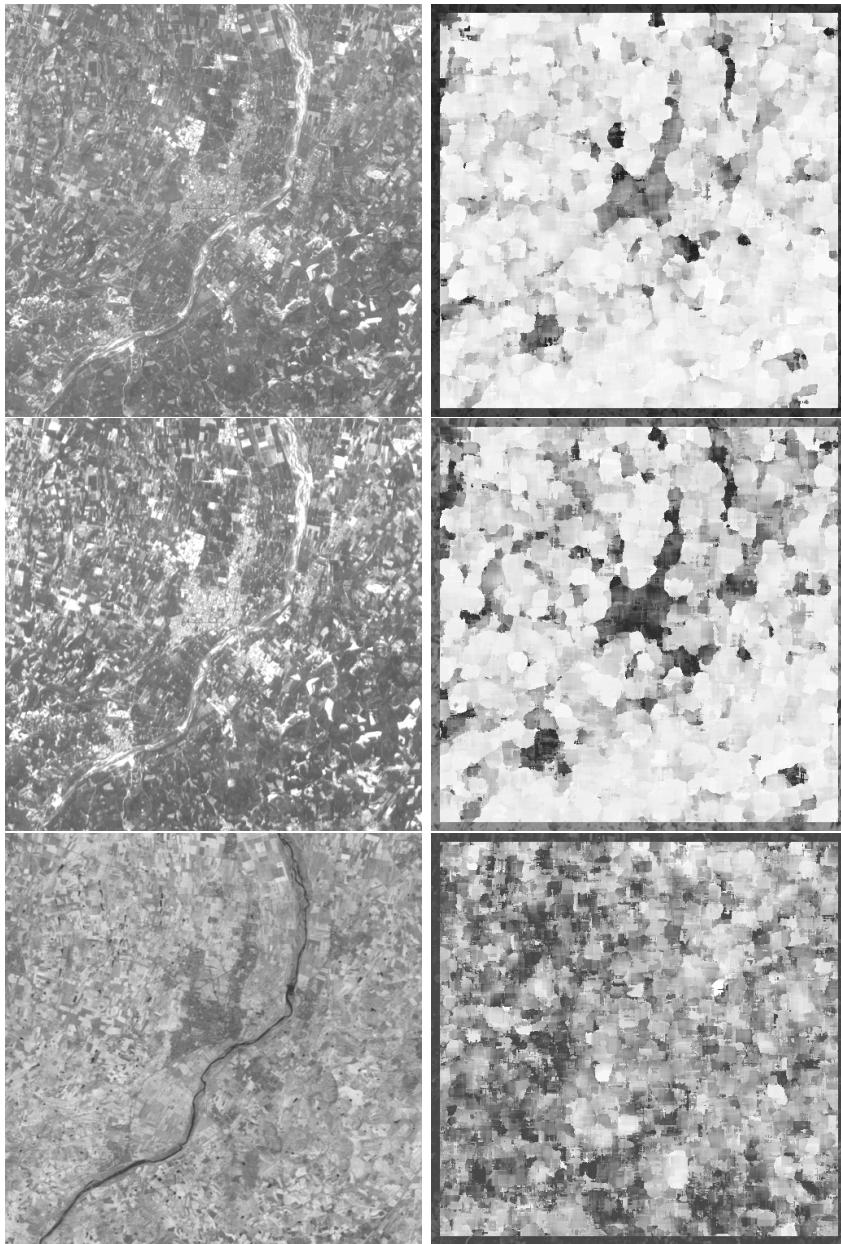


FIG. 8.1 – Images du satellite SPOT en mode spectral multibande XS de taille 512×512 pixels et de niveau de prétraitement 1-B en visée verticale (région de Vignola près de Modène en Italie) © Spot Image. De haut en bas, le canal vert (XS1), le rouge (XS2) et le proche infrarouge (XS3). A gauche : images SPOT originales. A droite : Images de paramètres de température d'un champ markovien gaussien sur les trois canaux de SPOT (images de texture).

des niveaux de gris. Ces estimations peuvent être lissées par des fenêtres de Parzen par exemple. La figure 8.2 illustre les résultats de l'apprentissage sur une des images. En réalité, pour éviter des hypothèses d'indépendance non justifiées, c'est la probabilité jointe des trois canaux XS conditionnellement aux classes qui est estimée et de même pour les trois images de texture.

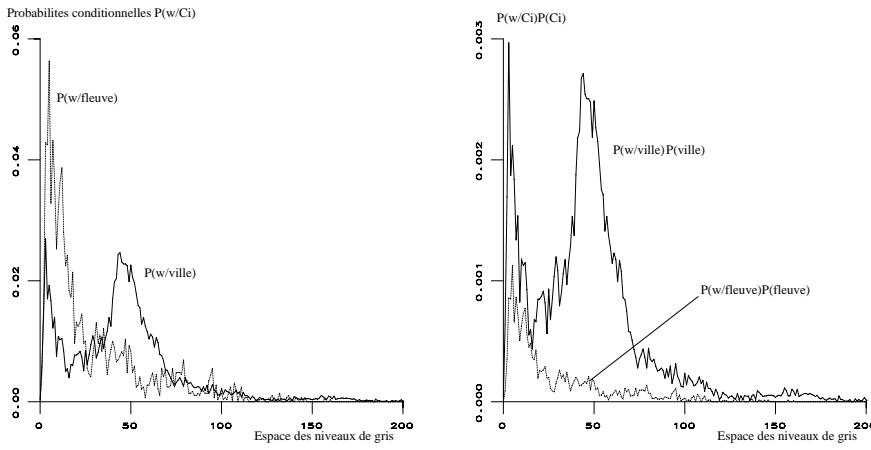


FIG. 8.2 – A gauche : probabilités conditionnelles des villes et des fleuves de l'image de texture sur XS2. A droite : probabilités conditionnelles aux deux classes multipliées par leur probabilité *a priori* $P(\text{ville}) = 11\%$ et $P(\text{fleuve}) = 2\%$.

Une des difficultés des méthodes bayésiennes est l'estimation des probabilités *a priori*. Si celles-ci sont fixées selon la proportion des classes présentes dans les images, cela conduit à diminuer fortement la probabilité des classes faiblement représentées (voir figure 8.2 à droite) et rend leur détection très difficile. Le choix effectué ici consiste à choisir ces estimations lorsqu'il y a peu de conflit entre les classes, et à prendre des probabilités *a priori* uniformes lorsque les propriétés statistiques des classes révèlent un fort conflit.

La figure 8.3 montre le résultat de la fusion, pour un critère de maximum *a priori*. Le fleuve est superposé en blanc à l'image originale, ainsi que les contours des zones urbaines. Le reste correspond à la classe C_3 .

8.1.8 Discussion

La large diffusion des méthodes probabilistes, en particulier des méthodes bayésiennes, est due, plus qu'à la justification qu'en donne Cox [93] (voir chapitre 3), à la connaissance acquise au cours d'expériences nombreuses permettant de guider les phases de modélisation et d'apprentissage.

L'avantage essentiel des méthodes probabilistes vient de ce qu'elles reposent sur une base mathématique solide et ont été l'objet de nombreux travaux. Elles proposent donc un éventail d'outils très riche permettant aussi bien la modélisation (par exemple par des familles de lois paramétriques aux propriétés bien étudiées) que l'apprentissage des modèles (pour des lois paramétriques ou non paramétriques) (voir par exemple [77, 209, 220]). Elles proposent également des règles d'usage soit théoriques (bornes, valeurs asymptotiques) soit heuristiques (tests d'hypothèses, critères de validité, tables de confiance). Enfin, la modélisation probabiliste soutenue par l'interprétation fréquentiste largement répandue dans le monde de la physique et du traitement du signal, est actuellement un concept universellement partagé qui sert naturellement de base de comparaison aux autres modélisations.

Un autre avantage des approches probabilistes et statistiques, cette fois du point de vue de la combinaison, est encore qu'elles reposent sur de solides bases mathématiques,

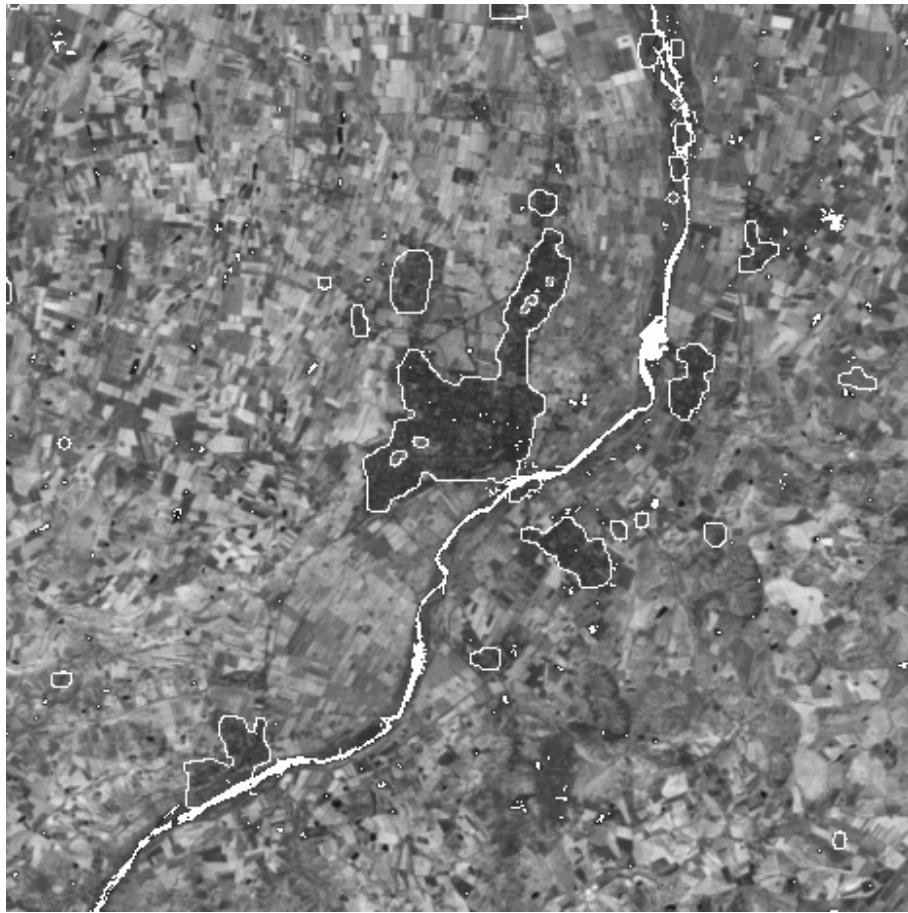


FIG. 8.3 – Images de détection par l’inférence bayésienne des classes ville, fleuve et C_3 sur la scène de Vignola par fusion des 6 images du satellite SPOT

et peuvent être utilisées pour la mise à jour de réseaux complexes de connaissances [265, 266]. Elles permettent d’introduire des informations qui s’expriment facilement sous forme de probabilités, telles que le contexte spatial dans le cadre des champs de Markov (voir partie 8.4) ou la qualité des informations exprimée comme la probabilité pour qu’une mesure soit fiable [168].

Cependant, et malgré leurs bases mathématiques solides, ces méthodes sont également l’objet de critiques et souffrent de plusieurs inconvénients. Nous les regroupons dans cette partie, en soulignant toutefois que certains d’entre eux sont contestés par les inconditionnels des probabilités.

Tout d’abord, si elles représentent bien l’incertain qui entache l’information, elles ne permettent pas aisément de représenter son imprécision, et elles conduisent souvent à confondre ces deux notions. Ensuite, elles nécessitent que, lors de l’apprentissage, des contraintes très strictes soient vérifiées par les mesures (imposées par les axiomes de base des probabilités) et par l’ensemble de classes considéré (exhaustivité). Ces contraintes peuvent rendre l’apprentissage très délicat (comment caractériser des zones qui ne soient pas du blé en imagerie aérienne² ?), ou, si le problème à traiter est complexe, conduire pratiquement à des in-

²Ce problème est un exemple du problème plus général que l’on rencontre en classification et en recon-

cohérences car l'utilisateur ne peut alors prendre en compte tout le réseau des dépendances probabilistes (cas des boucles logiques [265]). L'apprentissage de lois de probabilités nécessite, outre les hypothèses, un nombre de données important. Typiquement, l'apprentissage non paramétrique d'une loi multidimensionnelle dans des images ou des zones de taille limitée n'est souvent pas pertinent, et l'on se tourne alors vers des modèles paramétriques, qui à leur tour nécessitent des hypothèses sur la forme des lois.

L'estimation des probabilités *a priori* est souvent délicate et primordiale dans les cas où l'on a peu d'informations (distributions très plates des probabilités conditionnelles). Si, dans le cas du traitement d'images, les probabilités conditionnelles peuvent être souvent bien estimées par apprentissage à partir de fréquences d'occurrence, ce n'est en général pas le cas des probabilités *a priori*. Leur évaluation sort du cadre des probabilités fréquentistes et fait souvent appel à des concepts plus subjectifs. De plus, la combinaison bayésienne est contrainte, comme pour la modélisation, par les axiomes des probabilités, et son utilisation en pratique nécessite souvent des hypothèses simplificatrices (comme l'indépendance) rarement vérifiées. La théorie probabiliste et bayésienne combine les informations de façon conjonctive, par des produits de probabilités conditionnelles, ce qui conduit souvent en pratique à un effondrement des probabilités des événements qui sont déduits d'une longue chaîne de déduction.

La contrainte d'additivité peut être trop forte pour certains problèmes. Prenons l'exemple que donne Smets [312], dans le domaine du diagnostic médical. Si un symptôme s est toujours présent quand un patient a une pathologie A , et que l'on observe ce symptôme s , alors la probabilité pour que le patient ait A augmente. La contrainte d'additivité impose alors que la probabilité pour que le patient n'ait pas A diminue, alors qu'il n'y a pas de raison qui justifie cela (paradoxe de Hempel), si le symptôme s peut aussi être observé dans le cas d'autres pathologies³.

L'application des méthodes bayésiennes nécessite souvent beaucoup de connaissances sur le problème, et leur utilisation dans de bonnes conditions suppose qu'une nouvelle réflexion soit menée pour chaque problème à traiter⁴. Par exemple, le diagnostic « bayésien » peut être formalisé de la manière suivante :

$$p(A_i|O) = \frac{p(O|A_i)p(A_i)}{\sum_j p(O|A_j)p(A_j)} = \frac{p(O|A_i)p(A_i)}{p(O|A_i^C)p(A_i^C) + p(O|A_i)p(A_i)}, \quad (8.11)$$

naissance de formes : en général le complémentaire d'une classe n'est pas une classe.

³La contrainte d'additivité et la règle de Bayes (équation 3.8) impliquent que si $p(s|A) = 1$, alors $p(A|s) = \frac{p(A)}{p(s)}$ et donc $p(A|s) \geq p(A)$. Au contraire, $p(A^C|s) = 1 - p(A|s)$, donc $p(A^C|s) \leq p(A^C)$. Si l'argument de Smets réfutant cette inégalité peut être criticable, il peut également être interprété de la manière suivante : le modèle des probabilités additives peut être trop simpliste dans ce cas. En particulier, si parler de la probabilité d'une pathologie A peut avoir un sens, il est moins certain que parler de celle de $[A^C|s]$ en ait un. En effet, A^C ne correspond pas à *une* pathologie, mais est un ensemble infini, mal connu, imprécis, et il est difficile d'affirmer que $[A^C|s]$ est une proposition bien définie, binaire, représentant correctement la réalité. Ainsi, tout modèle permettant de déduire $p(A^C|s)$ peut être facilement contestable. Nous espérons ne pas trahir la pensée de Smets par cette interprétation.

⁴Cette situation s'oppose à celle des approches fréquentistes, en particulier à la théorie de Fisher, qui est essentiellement automatique : *Faced with a new situation, the working statistician can apply maximum likelihood in an automatic fashion, with little chance (in experienced hands) of going far wrong and considerable chance of providing a nearly optimal inference. In short, he does not have to think a lot about the specific situation in order to get on towards its solution* [138]. C'est une théorie d'archétypes, qui, en permettant de séparer les problèmes, permet d'obtenir des solutions raisonnables là où l'approche bayésienne, où tout doit être traité « d'un seul coup », serait trop complexe. Ce caractère automatique est certainement une des raisons qui font la popularité de la théorie de Fisher, malgré les inconvénients des approches fréquentistes.

où $p(A_i|O)$ désigne la probabilité pour que le patient ait la pathologie A_i , étant donné un ensemble d'observations O (examens cliniques, images, etc.), $p(O|A_i)$ désigne la probabilité conditionnelle des observations étant donnée la pathologie, et $p(A_i)$ est la probabilité *a priori* de A_i . La décision est prise à partir des $p(A_i|O)$. L'utilisation de cette formule nécessite soit de connaître l'ensemble de toutes les pathologies, soit d'avoir des statistiques reliant les observations à A_i^C (« non pathologie A_i »). Les deux solutions semblent irréalistes. De plus, toutes les distributions de probabilités intervenant dans la formule doivent pouvoir être estimées. Le problème est alors la limite des tests statistiques reliant les symptômes ou les observations aux pathologies et la difficulté d'avoir une estimation des probabilités *a priori*. Ces limites sont bien sûr plus générales et ne sont pas spécifiques à cet exemple particulier.

La modélisation probabiliste permet de raisonner sur des singletons seulement, qui représentent les différentes hypothèses, sous la contrainte du monde clos. Nous avons vu dans l'exemple précédent du diagnostic médical que cette hypothèse ne correspond pas à la réalité. De plus, les singletons ne permettent pas de représenter des situations complexes. Prenons le cas d'images affectées par l'effet de volume partiel (situation courante en imagerie médicale). Les modèles classiques dans la littérature pour représenter ce phénomène consistent à affecter à un point des probabilités d'appartenance aux tissus dont il est composé, ces probabilités étant proportionnelles à la quantité de chaque tissu dans le volume représenté par le point. Or, cela ne correspond à aucune réalité. En effet, ce type de modèle probabiliste signifie que l'on a une incertitude sur la classe à laquelle appartient le point (on sait que le point peut appartenir à plusieurs classes mais on ne sait pas laquelle), alors qu'il s'agit en fait d'une appartenance à plusieurs classes à la fois. Cela nous semble être un exemple typique où les modèles probabilistes utilisés ne modélisent pas correctement le phénomène observé.

Les ensembles flous ou la théorie des fonctions de croyance (ou de Dempster-Shafer) permettent des modélisations plus proches de la réalité pour certains problèmes et d'interprétation moins contestable. Par exemple, les ensembles flous permettent de rendre exactement compte du phénomène d'appartenance partielle qui est effectivement observé. Il ne suffit pas ici de relâcher la contrainte d'additivité des probabilités pour résoudre le problème : il s'agit bien de modéliser un type de phénomène tout à fait différent. Nous y reviendrons dans les deux chapitres suivants.

Une autre limite provient de la difficulté d'introduire dans le système de raisonnement des connaissances qui ne se traduisent pas simplement par des probabilités.

Dans le même ordre d'idées, il est difficile de modéliser l'absence de connaissances, des connaissances imprécises (au contraire des connaissances incertaines qui sont naturellement représentées par des probabilités), ou encore l'ignorance que l'on peut avoir sur un phénomène. Le principe de raison insuffisante ne suffit pas à prendre en compte l'ignorance et peut conduire à des contradictions suivant la manière dont on l'exprime.

Le même type de problème se pose avec le principe du maximum d'entropie. L'exemple bien connu de Shafer sur la probabilité pour qu'il y ait de la vie sur la planète Sirius l'illustre bien [302]⁵. Ces inconvénients, qui ne sont pas mieux résolus avec une version subjective

⁵Nous ne présentons pas ici l'exemple original de Shafer, qui peut prêter à discussion, mais un exemple donné par Dubois qui s'en approche. L'ignorance sur l'existence de la vie sur Sirius s'exprime classiquement en probabilités par $p(\text{vie}) = p(\text{non vie}) = 0,5$. Si l'on pose le problème autrement, en supposant qu'il peut y avoir trois possibilités, vie végétale, vie animale ou pas de vie, l'expression de l'ignorance conduira à affecter une probabilité de $1/3$ à chacune des trois hypothèses. On obtient alors $p(\text{vie}) = 2/3$. De manière analogue, on peut obtenir autant de valeurs différentes que de manières d'exprimer le problème. On peut trouver des

des probabilités, se manifestent dès qu'il s'agit de modéliser un raisonnement humain où les décisions sont prises à partir de données à la fois imprécises et incertaines, partielles, pas complètement fiables, conflictuelles, où les contraintes et les objectifs ne sont pas toujours très précis.

8.2 Fusion dans la théorie des fonctions de croyance

La théorie des fonctions de croyance (ou théorie de Dempster-Shafer) date des années 1970 mais son utilisation en fusion de signaux et d'images est relativement récente. Pourtant les premières applications sont prometteuses, et nous montrons dans ce chapitre quelles sont les caractéristiques de cette théorie qui justifient que l'on s'y intéresse, aussi bien du point de vue de la représentation des connaissances et de leurs imperfections (imprécision, incertitude, ambiguïté, ignorance, conflit) que de leur combinaison.

Bien que cette théorie soit inspirée des notions de probabilités supérieures et inférieures et donc souvent vue d'un point de vue probabiliste, elle peut être interprétée de manière plus générale d'un point de vue subjectif, comme un modèle formel quantitatif de degrés de confiance [314]. Un des atouts essentiels de cette théorie est qu'elle manipule des sous-ensembles plutôt que des singletons, ce qui lui confère une grande souplesse de modélisation pour de multiples situations rencontrées en fusion de signaux et d'images. Elle fournit également des représentations à la fois de l'incertitude et de l'imprécision, ainsi que de l'ignorance. Pour cela, plusieurs fonctions sont utilisées pour modéliser l'information et la manipuler, au lieu des seules probabilités utilisées dans le chapitre précédent. Cette théorie permet de mesurer des conflits entre les sources et de les interpréter en termes de fiabilité des sources, de monde ouvert ou de contradictions d'observations. Bien que plusieurs modes de combinaison soient possibles, la combinaison de type conjonctive est la plus utilisée dans les domaines qui nous intéressent ici et c'est surtout celle-ci que nous détaillerons. Cela rapporte l'essentiel du travail de l'utilisateur sur l'étape de modélisation et de représentation des informations et des connaissances disponibles.

Par la suite, nous illustrerons très souvent nos propos avec l'exemple de la classification multisource.

8.2.1 Modélisation

La théorie des fonctions de croyance permet, de manière analogue à la théorie des possibilités, comme nous le verrons dans le chapitre suivant, de représenter à la fois l'imprécision et l'incertitude à l'aide de fonctions de masse m , de plausibilité Pls et de croyance Bel [169, 302, 314]. Les fonctions de masse sont définies sur tous les sous-ensembles de l'espace D , appelé ici espace de discernement, (contenant par exemple les classes auxquelles on s'intéresse) et pas simplement sur les singletons comme les probabilités qui ne mesurent que la probabilité d'appartenance à une classe donnée.

Posons $D = \{C_1, C_2, \dots, C_n\}$ où chaque C_i désigne une hypothèse en faveur de laquelle une décision peut être prise (typiquement une classe dans un problème de classification multisource). Une fonction de masse est définie comme une fonction de 2^D (ensemble des

exemples analogues en traitement des images, en particulier dans les problèmes de classification.

parties de D) dans $[0, 1]$. En général on impose $m(\emptyset) = 0$, et une normalisation de la forme :

$$\sum_{A \subseteq D} m(A) = 1, \quad (8.12)$$

qui garantit une sorte de commensurabilité entre plusieurs jeux de masse.

La contrainte $m(\emptyset) = 0$ correspond à une hypothèse de monde clos, dans lequel toutes les solutions possibles sont effectivement représentées dans D (ce qui suppose que l'on est capable de les énumérer). Si l'on relâche cette contrainte et que l'on accepte d'avoir une masse strictement positive sur \emptyset , cela correspond alors à une hypothèse de monde ouvert, dans lequel des solutions hors de D sont envisageables.

Un élément focal est un sous-ensemble A de D tel que $m(A) > 0$. La réunion des éléments focaux est appelée noyau.

Une fonction de croyance Bel est une fonction totalement croissante définie de 2^D dans $[0, 1]$:

$$\begin{aligned} \forall A_1 \in 2^D, \dots A_k \in 2^D, \\ Bel(\cup_{i=1 \dots k} A_i) \geq \sum_{I \subseteq \{1 \dots k\}, I \neq \emptyset} (-1)^{|I|+1} Bel(\cap_{i \in I} A_i), \end{aligned} \quad (8.13)$$

où $|I|$ désigne le cardinal de I , et telle que $Bel(\emptyset) = 0$, $Bel(D) = 1$.

Etant donnée une fonction de masse m , la fonction Bel définie par :

$$\forall A \in 2^D, Bel(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B) \quad (8.14)$$

est une fonction de croyance. Inversement, à partir d'une fonction de croyance définie comme une fonction totalement croissante (inégalité 8.13) telle que $Bel(\emptyset) = 0$, $Bel(D) = 1$, on peut définir une fonction de masse par :

$$\forall A \in 2^D, m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} Bel(B). \quad (8.15)$$

Cette fonction de masse vérifie alors l'équation 8.14.

La fonction de croyance mesure la confiance totale que l'on a dans un sous-ensemble A . L'ensemble vide est exclu de la somme car il interviendrait sinon à la fois dans l'évaluation de A et dans l'évaluation de A^C ($\emptyset \subset A$ et $\emptyset \subset A^C$).

Ainsi, avoir une masse nulle sur un sous-ensemble A ne signifie pas que cet ensemble est impossible, mais simplement que l'on n'est pas capable d'affecter un degré précisément à A . En effet, on peut avoir des masses non nulles sur des sous-ensembles de A , conduisant donc à $Bel(A) \neq 0$. Cette remarque est très importante pour la modélisation, puisqu'elle permet de ne pas affecter de degrés si l'on n'est pas capable de le faire (ainsi, on ne force pas d'information là où l'on n'en dispose pas).

Dans l'hypothèse de monde ouvert, on a :

$$Bel(D) = 1 - m(\emptyset). \quad (8.16)$$

Une fonction de plausibilité Pls est également une fonction de 2^D dans $[0, 1]$ définie par :

$$\forall A \in 2^D, Pls(A) = \sum_{B \cap A \neq \emptyset} m(B) = 1 - Bel(A^C). \quad (8.17)$$

Plus généralement, pour tenir compte de la possibilité d'être en monde ouvert, on a :

$$Pls(A) = \sum_{B \cap A \neq \emptyset} m(B) = Bel(D) - Bel(A^C). \quad (8.18)$$

La plausibilité mesure la confiance maximum que l'on peut avoir dans A . Cette fonction a une interprétation naturelle dans le modèle des croyances transférables [314] où l'on considère que l'apport d'information peut permettre de transférer des croyances sur des sous-ensembles plus précis. La plausibilité représente alors la croyance maximale que l'on pourrait potentiellement affecter à un sous-ensemble A si l'on apprend par exemple que la solution se trouve dans A (toute la confiance mise dans un sous-ensemble B intersectant A est alors transférée sur A afin de mettre à 0 la confiance sur A^C).

On a les propriétés suivantes :

$$\forall A \in 2^D, Pls(A) \geq Bel(A), \quad (8.19)$$

$$\forall A \in 2^D, Bel(A) + Bel(A^C) \leq 1, \quad (8.20)$$

$$\forall A \in 2^D, Pls(A) + Pls(A^C) \geq 1, \quad (8.21)$$

$$\forall A \in 2^D, Bel(A) + Bel(A^C) = 1 \Leftrightarrow Bel(A) = Pls(A). \quad (8.22)$$

L'intervalle $[Bel(A), Pls(A)]$ est appelé intervalle de confiance et sa longueur est une mesure de l'ignorance que l'on a sur un événement A et son complémentaire.

Si l'on affecte des masses uniquement aux hypothèses simples ($m(A) = 0$ pour $|A| > 1$), alors les trois fonctions m , Bel et Pls sont égales et sont une probabilité. On parle alors de fonctions de masse bayésiennes. Dans les cas plus complexes, ce n'est pas le cas et il n'y a pas d'équivalence directe avec des probabilités. Des fonctions analogues aux fonctions de crédibilité et de plausibilité pourraient être obtenues par exemple à partir de probabilités conditionnelles à des comportements pessimistes et optimistes respectivement, mais leur formalisation serait beaucoup plus délicate que ce que propose la théorie des fonctions de croyance.

Parmi les fonctions de masse particulières, on distingue les fonctions à support simple, pour lesquelles toute la masse porte sur un sous-ensemble non vide A et sur l'ensemble de discernement D :

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

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

Dans le cas où s vaut 0, alors toute la masse porte sur D . Cette fonction représente l'ignorance totale, dans le sens où aucun sous-ensemble ne peut être distingué.

La possibilité d'affecter des masses aux hypothèses composées, et donc de travailler sur 2^D plutôt que sur D constitue un des avantages de cette théorie. Elle permet en effet une modélisation très souple et très riche, en particulier de l'ambiguïté ou de l'hésitation entre classes. Citons quelques exemples de situations dans lesquelles la fusion par la théorie des fonctions de croyance peut être employée :

- dans les cas limites (que l'on peut considérer comme idéaux) où l'on connaît toute l'information sur le problème à traiter ;
- lorsqu'une source donne des informations seulement sur certaines classes : par exemple certaines images TEP⁶ donnent des informations sur les limites du cerveau mais pas de la tête ;
- lorsqu'une source n'est pas capable de différencier deux hypothèses : la théorie des fonctions de croyance permet alors de considérer la disjonction de ces deux classes, sans introduire d'information arbitraire forçant leur séparation ;
- lorsque l'on veut modéliser les effets de volume partiel, donc typiquement représenter l'appartenance d'un pixel ou un voxel à plusieurs classes ;
- lorsque l'on veut représenter la fiabilité globale d'une source : cela peut être réalisé en affectant une masse non nulle à D ;
- dans les cas où la fiabilité d'une source dépend des classes (par exemple des images cérébrales fonctionnelles ne donnent pas d'informations très fiables sur l'anatomie, alors qu'au contraire les images IRM sont très fiables pour les classes anatomiques) ;
- dans les cas où l'on veut introduire des informations *a priori* : même si ces informations ne sont pas représentables facilement par des probabilités, elles peuvent être introduites si elles induisent un moyen de choisir les éléments focaux (en particulier les disjonctions d'hypothèses), de définir ou de modifier les fonctions de masse.

8.2.2 Estimation des fonctions de masse

L'estimation des fonctions de masse est un problème difficile, qui n'a pas de solution universelle. La difficulté est augmentée ici si l'on veut affecter des masses aux hypothèses composées [153, 218]. En traitement d'images par exemple, on peut les construire à trois niveaux : au niveau le plus haut (souvent abstrait et symbolique), la représentation de l'information est utilisée d'une manière similaire à ce qui est fait en intelligence artificielle et les masses sont affectées à des propositions, et souvent données par des experts [15, 165, 256]. Ce type d'information n'est le plus souvent pas directement dérivé de mesures dans les données, et les méthodes correspondantes ne sont donc pas spécifiques du traitement des images. A un niveau intermédiaire, les masses sont calculées à partir d'attributs et peuvent s'appuyer par exemple sur des modèles géométriques en image [7, 79, 340, 95]. Ce niveau est bien adapté à des problèmes de reconnaissance des formes à partir de modèles mais il est difficile de l'utiliser pour des problèmes de fusion sur des structures complexes sans modèle. A bas niveau (le pixel en traitement d'images), beaucoup de méthodes sont envisageables, et la plupart s'appuient sur des méthodes statistiques de reconnaissance des formes.

La manière la plus simple que l'on puisse imaginer consiste à calculer les masses sur les singletons dans une source (image par exemple) I_j par :

$$m_j(\{C_i\})(x) = M_i^j(x), \quad (8.23)$$

où $M_i^j(x)$ est estimée le plus souvent comme une probabilité. Les masses sur tous les autres sous-ensembles de D sont alors nulles. Il est clair que ce modèle est très réducteur et n'exploite pas les caractéristiques intéressantes de la théorie des fonctions de croyance. Beaucoup d'approches s'appuient toutefois sur un tel modèle initial, puis répartissent les masses sur

⁶Images de tomographie par émission de positons, utilisées en particulier en imagerie cérébrale fonctionnelle.

l'ensemble des hypothèses composées ou n'utilisent que certaines hypothèses composées, dans une démarche simplificatrice et souvent très heuristique [340, 209, 277, 364]. Mais d'autres approches peuvent également être envisagées. Dans la suite, nous présentons quelques modèles de la littérature.

Modification de modèles probabilistes

Le modèle le plus simple et le plus souvent utilisé consiste à employer la technique d'affaiblissement [302]. Les nouvelles masses m' sont calculées à partir des masses initiales m de la manière suivante (l'indice j représentant la source d'information ainsi que l'élément x sur lequel on raisonne sont omis ici) :

$$m'(\{C_i\}) = \alpha m(\{C_i\}), \quad (8.24)$$

$$m'(D) = 1 - \alpha + \alpha m(D), \quad (8.25)$$

où $\alpha \in [0, 1]$ est le coefficient d'affaiblissement. Dans le cas où les masses initiales sont apprises sur les singletons seulement, par exemple à partir de probabilités, alors $m(D) = 0$ et $m'(D) = 1 - \alpha$. Cette technique est souvent utilisée pour affaiblir une source en fonction de sa fiabilité, et permet d'affecter une masse à D qui sera faible si la source est fiable et importante si la source ne l'est pas. Dans les cas extrêmes, la valeur $\alpha = 0$ est utilisée pour une source qui n'est pas fiable du tout, et toute la masse est alors affectée à D , ce qui représente l'ignorance totale. La valeur $\alpha = 1$ est utilisée pour une source fiable dans laquelle toute la masse est affectée aux singletons et où il n'y a aucune ambiguïté entre classes.

Ce type de modèle est très simple. L'apprentissage des masses sur les singletons peut bénéficier des techniques classiques d'apprentissage statistique. Cependant les disjonctions d'hypothèses ne sont pas modélisées, ce qui limite beaucoup la portée de ce modèle.

Deux modèles d'inspiration probabiliste ont été proposés par A. Appriou [10], et prennent en compte d'autres disjonctions que D . Ces modèles supposent une estimation initiale de probabilités conditionnelles $p(f(x)|C_i)$ (où $f(x)$ désigne les caractéristiques de x extraites de la source et sur lesquelles s'appuie la fusion), notées plus simplement $p(x|C_i)$. La fonction de masse associée à une source est calculée par combinaison de fonctions de masse associées à chaque singleton, définies dans le premier modèle par :

$$m^i(\{C_i\})(x) = \frac{\alpha_i R p(x|C_i)}{1 + R p(x|C_i)}, \quad (8.26)$$

$$m^i(D \setminus \{C_i\})(x) = \frac{\alpha_i}{1 + R p(x|C_i)}, \quad (8.27)$$

$$m^i(D)(x) = 1 - \alpha_i, \quad (8.28)$$

où α_i est un coefficient d'affaiblissement lié à la classe C_i , qui permet de prendre en compte la fiabilité de la source pour cette classe en particulier (et non plus globale comme dans le modèle précédent), et R est un coefficient de pondération des probabilités. Si $R = 0$, seule la fiabilité de la source est prise en compte, sinon les données sont également prises en compte.

Dans le deuxième modèle, les masses associées à chaque singleton sont définies par :

$$m^i(\{C_i\})(x) = 0, \quad (8.29)$$

$$m^i(D \setminus \{C_i\})(x) = \alpha_i(1 - Rp(x|C_i)), \quad (8.30)$$

$$m^i(D)(x) = 1 - \alpha_i + \alpha_i Rp(x|C_i). \quad (8.31)$$

Ce modèle correspond au cas où $p(x|C_i)$ nous donne une information essentiellement sur ce qui n'est pas C_i .

La masse associée à la source est ensuite calculée comme $\oplus_i m^i$, où \oplus est la somme orthogonale de Dempster (voir section 8.2.3). Ce modèle est bien adapté dans les cas où l'on apprend facilement une classe contre toutes les autres, ce qui est fréquent en reconnaissance des formes dans les images, ou dans les cas où chaque classe est déterminée à partir d'un détecteur adapté (par exemple un détecteur de routes dans une image aérienne permet de définir la probabilité d'appartenance à la route par rapport à celle de toutes les autres classes, mais n'est pas capable de distinguer ces autres classes).

Dans [118], les disjonctions sont définies en fonction d'un critère de significativité des probabilités conditionnelles. Si une seule probabilité $p(x|C_i)$ est significative (ce qui nécessite de définir des seuils), alors un modèle simple de masse portant sur les singltons est utilisé. Si plusieurs probabilités sont significatives, les disjonctions des hypothèses correspondantes sont également prises en compte. Par exemple, si trois valeurs sont significatives et telles que $p(x|C_i) > p(x|C_j) > p(x|C_k)$, la fonction de masse est construite par :

$$m(\{C_i\})(x) = p(x|C_i) - p(x|C_j), \quad (8.32)$$

$$m(\{C_i \cup C_j\})(x) = p(x|C_j) - p(x|C_k), \quad (8.33)$$

$$m(\{C_i \cup C_j \cup C_k\})(x) = p(x|C_k), \quad (8.34)$$

puis les masses sont normalisées. Si aucune probabilité n'est significative, la masse porte entièrement sur D .

Modification de modèles de distances

Une approche de type reconnaissance des formes est proposée dans [107]. Si chaque classe C_i est représentée par un prototype (ou un centre) x_i , une fonction de masse associée à chaque classe peut-être définie dans laquelle C_i et D sont les seuls éléments focaux :

$$m^i(\{C_i\})(x) = \alpha e^{-\gamma d^2(x, x_i)}, \quad (8.35)$$

$$m^i(D)(x) = 1 - \alpha e^{-\gamma d^2(x, x_i)}. \quad (8.36)$$

Les paramètres α et γ permettent de jouer sur la quantité d'ignorance et la forme des fonctions de masse. La distance $d^2(x, x_i)$ permet d'affecter une masse d'autant plus importante que $x \ll \text{ressemble} \gg$ à la classe C_i . Les m^i sont ensuite combinées selon la règle de Dempster (voir section 8.2.3) pour avoir une masse prenant en compte l'information sur toutes les classes.

Cette approche peut également être appliquée aux k plus proches voisins. La distance est alors la distance de x à l'un de ses voisins, et la masse est affectée selon le modèle précédent à la classe à laquelle appartient ce voisin et à D . Les fonctions calculées pour chacun des voisins de x sont ensuite combinées par la règle de Dempster.

***A priori* sur les éléments focaux composés (disjonctions)**

Dans de nombreuses applications, il est possible de disposer d'informations *a priori* qui permettent de déterminer de manière supervisée quels sont les éléments focaux à prendre en compte. Ces méthodes ont été utilisées par exemple dans [37, 244, 243, 334]. Dans [37], des images du cerveau sont combinées pour détecter des pathologies (voir section 8.2.7). Les fonctions de masse sont estimées automatiquement à partir des niveaux de gris [50] et les classes non distinguées dans certaines images par leur niveaux de gris sont regroupées en disjonctions. Dans [334], les résultats de détecteurs de plusieurs structures sont fusionnés pour interpréter une image radar. Ce sont alors les capacités des détecteurs à différencier ou non différentes classes de structures qui permettent de définir les éléments focaux et les disjonctions de classes à prendre en compte. Dans [244, 243], des attributs extraits d'images de différents capteurs sont combinés pour différencier des mines d'objets inoffensifs, dans un programme de déminage humanitaire. Les mesures à combiner peuvent être caractéristiques d'une classe ou de l'espace de discernement complet. Par exemple, la profondeur des objets permet d'affecter une masse aux objets inoffensifs si elle est importante, mais ne permet pas de distinguer les types d'objets si elle est faible et la masse est alors affectée à D .

Ce type d'approche est très efficace si l'on dispose de telles informations, mais elle reste supervisée, et donc applicable à des problèmes où le cardinal de D reste raisonnable.

Apprentissage des éléments focaux composés

Les méthodes d'apprentissage des éléments focaux s'appuient souvent sur des classifications préalables effectuées dans chaque source séparément. Typiquement à partir de matrices de confusion il est possible d'identifier les classes confondues selon une source, dont la réunion constituera un élément focal de la fonction de masse attachée à cette source.

De manière complètement non supervisée, les intersections entre les classes détectées dans une source et celles détectées dans une autre source peuvent définir les singletons de l'espace de discernement, les classes détectées dans chaque source devenant alors des disjonctions [231].

Des mesures de dissonance et consonance sont proposées dans [240]. L'idée consiste à modifier une fonction de masse initiale portant uniquement sur les singletons en affaiblissant les masses sur les singletons en fonction du degré de consonance de ceux-ci, et en créant des masses sur des disjonctions de deux classes en fonction du degré de dissonance entre ces classes. Cette méthode a été appliquée à la fusion de plusieurs classifieurs. La consonance d'une classe est calculée d'après le nombre d'éléments affectés à cette classe par tous les classifieurs, et la dissonance d'après le nombre d'éléments classés différemment.

Dans le cas où les éléments sont caractérisés par une mesure dans un espace à une dimension (typiquement représentés par un histogramme), les masses sur les hypothèses composées peuvent être définies dans les zones de recouvrement ou d'ambiguïté entre deux classes voisines. Une autre méthode, s'inspirant des méthodes de seuillage hiérarchique, est proposée dans [279] où chaque pic de l'histogramme correspond à un singleton. Puis l'histogramme est progressivement seuillé à des hauteurs décroissantes, et des disjonctions sont créées lorsque des maxima se regroupent. Cette méthode est à rapprocher des arbres de composantes utilisés par exemple en morphologie mathématique avec le concept de topologie des coupes [116], ainsi que des intervalles de confiance et de leurs liens avec les distributions de possibilité

[133] (voir chapitre 5).

Introduction de disjonctions par morphologie mathématique

Sans se restreindre à des espaces de représentation de dimension 1, la méthode proposée dans [38] permet de calculer des masses sur des disjonctions par érosions et dilations de masses définies dans un premier temps sur des singletons. Les propriétés de ces opérations morphologiques (en particulier de dualité) permettent de les interpréter comme des croyances et plausibilités, dont les masses sont ensuite déduites.

8.2.3 Combinaison conjonctive

Règle de Dempster

Soit m_j ($j = 1 \dots l$) la fonction de masse définie pour la source j . La combinaison conjonctive des fonctions de masse est effectuée selon la règle orthogonale de Dempster [302, 314], définie $\forall A \subseteq D$ par :

$$(m_1 \oplus m_2 \oplus \dots \oplus m_l)(A) = \sum_{B_1 \cap \dots \cap B_l = A} m_1(B_1)m_2(B_2)\dots m_l(B_l). \quad (8.37)$$

Des justifications axiomatiques de cette règle peuvent être trouvées dans [314]. Les différences entre ces axiomes et ceux de Cox [93] (qui permettent de justifier les règles des probabilités) expliquent les origines des différences entre les deux théories [30]. Ces aspects sont reportés en annexe 8.2.6.

Conflit et normalisation

Dans l'équation 8.37, non normalisée, la masse affectée par la combinaison à l'ensemble vide est en général non nulle. Elle s'interprète souvent comme le conflit entre les sources. Notons que cette mesure de conflit n'est pas une mesure absolue mais dépend de la modélisation effectuée (en particulier de la répartition des masses sur les différents sous-ensembles de D). Le conflit peut avoir deux sources essentielles : soit les sources ne sont pas fiables, soit elles donnent des informations sur des phénomènes différents. Dans le premier cas, il est acceptable de combiner les sources et une solution pour prendre en compte le conflit est d'affaiblir les sources en fonction de leur fiabilité. Nous y reviendrons un peu plus loin. Dans le deuxième cas, la combinaison n'a pas de sens. Des méthodes de regroupement des sources selon les phénomènes qu'elles observent ont été proposées, visant à combiner les sources à l'intérieur de chaque groupe uniquement. Ces groupes sont calculés de sorte à minimiser le conflit dans chaque groupe [243, 297].

Dans une hypothèse de monde ouvert, une masse non nulle sur l'ensemble vide peut également représenter une solution non prévue dans D . Sous l'hypothèse du monde fermé, où tout ce qui est possible est représenté dans D , cette interprétation n'est pas acceptable,

ce qui conduit à normaliser le résultat de la combinaison sous la forme⁷ :

$$(m_1 \oplus \dots \oplus m_l)(A) = \frac{\sum_{B_1 \cap \dots \cap B_l = A} m_1(B_1) \dots m_l(B_l)}{1 - \sum_{B_1 \cap \dots \cap B_l = \emptyset} m_1(B_1) \dots m_l(B_l)}, \quad (8.38)$$

$$\text{et } (m_1 \oplus m_2 \oplus \dots \oplus m_l)(\emptyset) = 0$$

si le dénominateur de l'équation 8.38 est non nul, c'est-à-dire si :

$$k = \sum_{B_1 \cap \dots \cap B_l = \emptyset} m_1(B_1) m_2(B_2) \dots m_l(B_l) < 1. \quad (8.39)$$

Cette quantité (qui mesure le conflit entre les sources) est donc directement prise en compte dans la combinaison sous forme de facteur de normalisation. Elle représente la masse qui serait affectée à l'ensemble vide si l'on n'avait pas cette normalisation (équation 8.37). Il est important de prendre en compte cette valeur pour juger de la qualité de la combinaison : celle-ci peut ne pas avoir grand sens en cas de fort conflit et conduire à des décisions criticables.

Prenons un exemple simple où $D = \{C_1, C_2, C_3\}$ et deux fonctions de masse n'ayant que les singletons comme éléments focaux et les valeurs suivantes :

	C_1	C_2	C_3
m_1	0,9	0,0	0,1
m_2	0,0	0,9	0,1

Leurs fusions non normalisée et normalisée donnent :

	C_1	C_2	C_3	\emptyset
$m_1 \oplus m_2$ non normalisée	0,0	0,0	0,01	0,99
$m_1 \oplus m_2$ normalisée	0,0	0,0	1,0	0,0

Toute la masse est alors focalisée sur C_3 qui est la seule classe où les deux sources sont d'accord, mais pour dire que ce n'est qu'une solution très faiblement plausible. La normalisation a masqué le conflit. La forme non normalisée est le plus souvent préférable en cas de conflit. Ici, elle permet d'affecter la masse essentiellement à l'ensemble vide, et l'origine du conflit peut venir de l'hypothèse de monde ouvert, d'une faible fiabilité d'au moins une des deux sources, ou du fait qu'une source voit un objet dans la classe C_1 alors que la seconde source voit un autre objet dans la classe C_2 .

D'autres méthodes que la normalisation ont été proposées pour éliminer la masse sur l'ensemble vide. Par exemple, cette masse est affectée à D dans [349], donc transformée en ignorance. Dans [127], une méthode plus fine est proposée : par exemple, si les éléments focaux A_1 et A_2 de deux sources sont en conflit ($A_1 \cap A_2 = \emptyset$), alors le produit $m_1(A_1)m_2(A_2)$ est affecté à $m(A_1 \cup A_2)$. Cela suppose qu'au moins une des deux sources est fiable mais que l'on ne sait pas laquelle, et la forme disjonctive du résultat est l'attitude la plus prudente.

⁷Cette forme normalisée est la règle de Dempster au sens strict [302], la forme non normalisée ayant été proposée plus tard [314] mais semblant aujourd'hui préférable dans la plupart des applications.

Propriétés

Examinons maintenant les propriétés de la règle de combinaison. Elle est commutative et associative. La fonction de masse définie par :

$$m_0(D) = 1 \text{ et } \forall A \subseteq D, A \neq D, m_0(A) = 0 \quad (8.40)$$

est élément neutre pour la combinaison. Cette masse représente une source complètement non informative, qui ne distingue aucun élément de D . Qu'elle ne joue aucun rôle dans la combinaison correspond donc bien à l'intuition. La définition de cette fonction de masse remplace le principe d'indifférence utilisé en probabilités (équirépartition des probabilités sur tous les éléments), et représente mieux l'absence d'information.

La loi \oplus n'est pas idempotente. Considérons à nouveau l'exemple précédent, mais avec cette fois les fonctions de masse suivantes :

	C_1	C_2	C_3
m_1	0,7	0,2	0,1
m_2	0,7	0,2	0,1

Leurs fusions non normalisée et normalisée donnent :

	C_1	C_2	C_3	\emptyset
$m_1 \oplus m_2$ non normalisée	0,49	0,04	0,01	0,46
$m_1 \oplus m_2$ normalisée	0,91	0,07	0,02	0,0

Cet exemple illustre la non-idempotence de la règle de combinaison. Les valeurs les plus fortes sont renforcées et les plus faibles diminuées. Il est important également de noter que le conflit entre deux fonctions de masse identiques est non nul, et qu'il est d'autant plus fort que la masse est répartie sur les singltons.

A l'origine, cette règle de combinaison était réputée applicable seulement sous l'hypothèse d'indépendance des sources. Il a été montré [275, 276] que la règle est encore applicable sans cette hypothèse, en s'appuyant sur l'analogie avec les ensembles fermés aléatoires. De manière moins technique et plus philosophique, l'indépendance dans le cadre des fonctions de croyance ne doit pas être comprise au sens statistique, mais dans un sens plus « cognitif » [316]. On parle alors d'indépendance cognitive. Imaginons par exemple des experts dont on veut combiner les opinions. Ils ne sont vraisemblablement pas indépendants statistiquement (s'ils sont experts du même domaine), mais on peut attendre d'eux qu'ils le soient cognitivement, c'est-à-dire que chacun se forge une opinion sans consulter les autres. C'est à ce type d'indépendance que s'applique la règle de Dempster, ce qui se traduit par la non-idempotence de la règle, conduisant à un renforcement des fonctions de masse identiques. Sous hypothèse de dépendance, on souhaiterait au contraire avoir une règle idempotente. Nous reviendrons sur ces considérations dans la théorie des ensembles flous.

Lorsque les fonctions m , Bel , et Pls sont des probabilités (c'est-à-dire lorsque les seuls éléments focaux sont des singltons), la loi de combinaison de Dempster est cohérente avec les lois classiques des probabilités. Cela fait donc apparaître les probabilités comme la limite de la théorie des croyances lorsqu'il n'y a pas d'ambiguïté ni d'imprécision et que seule l'incertitude des données doit être prise en compte.

La règle de Dempster a un comportement conjonctif, puisqu'elle donne des éléments focaux qui sont les intersections des éléments focaux des fonctions de masse initiales. Elle renforce donc la focalisation et diminue la longueur des intervalles de confiance [*Bel*, *Pls*].

En pratique, le calcul de la combinaison s'effectue en établissant la table d'intersection des éléments focaux. Par exemple, si m_1 porte sur $C_1 \cup C_2$ (typiquement dans le cas d'une source qui n'est pas capable de différencier ces deux classes) et C_3 , et m_2 sur C_1 et $C_2 \cup C_3$, les éléments focaux de $m_1 \oplus m_2$ sont donnés par la table d'intersection suivante :

	$C_1 \cup C_2$	C_3
C_1	C_1	\emptyset
$C_2 \cup C_3$	C_2	C_3

Les éléments focaux ne sont alors plus que les singletons et l'ensemble vide. Cet exemple illustre comment la combinaison conjonctive réduit l'imprécision et résout (ou diminue en général) l'ambiguïté de chaque source.

En particulier, les intervalles [*Bel*, *Pls*] sont réduits après la combinaison, puisque les fonctions de masse sont plus focalisées que les masses initiales (elles portent sur des ensembles plus petits).

La masse sur l'ensemble vide (mesurant le conflit) augmente au contraire lors de la fusion.

Supposons maintenant que la modélisation inclue une part d'ignorance et qu'une masse non nulle soit affectée à D dans les deux sources. La table d'intersection donne alors :

	$C_1 \cup C_2$	C_3	D
C_1	C_1	\emptyset	C_1
$C_2 \cup C_3$	C_2	C_3	$C_2 \cup C_3$
D	$C_1 \cup C_2$	C_3	D

Cette fois, l'ambiguïté n'est que partiellement réduite et il reste une masse non nulle sur les éléments imprécis (disjonctions de classes). Le conflit en revanche diminue, ce qui est une propriété générale de la combinaison de masses affaiblies en renforçant D .

Affaiblissement

L'affaiblissement permet de modéliser la fiabilité des sources à l'aide d'un coefficient α ($\alpha \in [0, 1]$) utilisé pour augmenter la masse sur D . L'idée est de renforcer l'ignorance d'autant plus que la source est moins fiable. Une fonction de masse m est alors modifiée en fonction de masse m' selon les formules suivantes :

$$m'(A) = \alpha m(A) \quad \forall A, A \neq D,$$

$$m'(D) = 1 - \alpha(1 - m(D)) = 1 - \alpha + \alpha m(D).$$

L'effet de l'affaiblissement est d'augmenter les intervalles [*Bel*, *Pls*] et, lors de la combinaison, de réduire le conflit.

Conditionnement

Prenons maintenant le cas particulier d'une source qui donne une information certaine sur un sous-ensemble B de D . Cette information se modélise de la manière suivante :

$$m_B(B) = 1 \text{ et } \forall A \subseteq D, A \neq B, m_B(A) = 0. \quad (8.41)$$

Toutes les sources doivent alors être « conditionnées » par m_B , afin de prendre en compte le fait que la vérité ne peut être que dans B . Le conditionnement se fait simplement en combinant une fonction de masse m avec m_B :

$$\forall A \subseteq D, m \oplus m_B(A) = \sum_{A=B \cap C} m(C), \quad (8.42)$$

qui s'écrit également :

$$\forall A \subseteq D, A \not\subseteq B, m \oplus m_B(A) = 0, \quad (8.43)$$

$$\forall A \subseteq D, A \subseteq B, m \oplus m_B(A) = \sum_{X \subseteq B^C} m(A \cup X). \quad (8.44)$$

Le conditionnement correspond au modèle de croyances transférables [314] : la connaissance de B conduit à transférer toute la masse sur les sous-ensembles inclus dans B . Ainsi, la croyance initialement affectée à un sous-ensemble $A = A_1 \cup A_2$ (avec $A_1 \subseteq B$ et $A_2 \subseteq B^C$) représentait le fait que la vérité pouvait être n'importe où dans A . La connaissance de B permet maintenant de préciser l'information et de réduire A à A_1 . En quelque sorte, la croyance diffuse dans A est maintenant concentrée dans la seule partie qui est incluse dans B .

Le conditionnement effectué par la règle conjonctive est l'équivalent, dans le cadre des fonctions de croyance, des probabilités conditionnelles, qui correspondent également à une conjonction. On a en effet :

$$P(X|B) = \frac{P(X \cap B)}{P(B)}.$$

Fonctions de masse séparables

Considérons maintenant des fonctions de masse à support simple. Si m_1 et m_2 sont des fonctions à support simple de même support A , avec des poids s_1 et s_2 , alors la combinaison donne une fonction de même support avec un poids $s_1 + s_2 - s_1 s_2$. De telles fonctions ne sont jamais conflictuelles.

Si les deux fonctions ont des supports différents A_1 et A_2 , alors la combinaison donne :

$$\begin{aligned} (m_1 \oplus m_2)(A_1 \cap A_2) &= s_1 s_2 \\ (m_1 \oplus m_2)(A_1) &= s_1(1 - s_2) \\ (m_1 \oplus m_2)(A_2) &= s_2(1 - s_1) \\ (m_1 \oplus m_2)(D) &= (1 - s_1)(1 - s_2) \\ (m_1 \oplus m_2)(B) &= 0 \quad \forall B, B \neq A_1, B \neq A_2, B \neq A_1 \cap A_2, B \neq D. \end{aligned}$$

En particulier, si $A_1 \cap A_2 = \emptyset$, alors les deux fonctions sont conflictuelles, une masse non nulle étant affectée à l'ensemble vide. Lorsque $s_1 s_2 \neq 1$, la fonction résultante est appelée fonction de masse séparable.

Complexité

Dans le cas général, ainsi que le montre la formule 8.37, la combinaison est de complexité exponentielle. En pratique, il est rare que tous les sous-ensembles de D soient à prendre en compte, et la complexité reste souvent plus raisonnable. Une complexité linéaire est obtenue si les masses sont modélisées selon la structure de Barnett [19], c'est-à-dire où les éléments focaux de chaque source sont uniquement les singletons et les compléments des singletons (fonctions séparables). Cette structure est adaptée à des problèmes de reconnaissance des formes dans lesquels chaque source est un détecteur qui permet de distinguer une classe contre toutes les autres. Mais elle n'est pas générale et ne s'applique pas aux sources qui nécessitent des éléments focaux qui soient des disjonctions quelconques.

8.2.4 Autres modes de combinaison

D'autres modes de combinaison, tels que des modes disjonctifs ou de compromis, sont possibles, en remplaçant l'intersection dans la formule 8.37 par une autre opération ensembliste. Par exemple, une fusion disjonctive est obtenue en prenant la réunion [316] :

$$(m_1 \oplus_{\cup} \dots \oplus_{\cup} m_l)(A) = \sum_{B_1 \cup \dots \cup B_l = A} m_1(B_1) \dots m_l(B_l). \quad (8.45)$$

Notons que cette combinaison ne peut pas faire apparaître de conflit. Elle élargit les éléments focaux et fournit donc une information moins précise que chacune des sources. Ce mode de fusion peut être intéressant quand on ne sait pas modéliser *a priori* les fiabilités des sources, leurs ambiguïtés et imprécisions. Par exemple, si une source est focalisée sur A et une autre sur B avec $A \cap B = \emptyset$, une manière de ne pas lever le conflit est de conclure que la vérité est dans $A \cup B$, ce que permet la fusion disjonctive.

Toutefois, dans la plupart des applications en fusion d'images, on cherche à obtenir une fonction de masse combinée plus focalisée que les fonctions de masse initiales. Ainsi, on préfère la fusion conjonctive, ce qui implique de prendre en compte les imprécisions, fiabilités, ambiguïtés de chaque source à l'étape de modélisation. Elle constitue alors l'étape la plus cruciale et qui requiert le plus d'attention.

8.2.5 Décision

Une fois calculées les fonctions de masse combinées, les fonctions de croyance et de plausibilité sont déduites par les équations 8.14 et 8.17. La dernière étape est celle de la décision, donc du choix d'un sous-ensemble de D maximisant un certain critère. Dans la suite, m , Bel et Pls désignent les fonctions de masse, de croyance et de plausibilité obtenues après combinaison.

Dans la théorie des fonctions de croyances, plusieurs règles de décision sont possibles et sont le plus souvent appliquées au choix d'un singleton C_i .

Le maximum de plausibilité :

$$x \in C_i \text{ si } Pls(C_i)(x) = \max \{Pls(C_k)(x), 1 \leq k \leq n\}, \quad (8.46)$$

cette règle étant optimale au sens de critères d'inspiration probabiliste pour des fonctions de masse dérivées de probabilités [9].

Le maximum de crédibilité :

$$x \in C_i \text{ si } Bel(C_i)(x) = \max \{Bel(C_k)(x), 1 \leq k \leq n\}, \quad (8.47)$$

qui est équivalent au critère du maximum de plausibilité dans le cas où le résultat de la combinaison ne porte que sur les singletons.

Le maximum de crédibilité sans recouvrement des intervalles de confiance (sans risque d'erreur) :

$$x \in C_i \text{ si } Bel(C_i)(x) \geq \max \{Pls(C_k)(x), 1 \leq k \leq n, k \neq i\}, \quad (8.48)$$

cette dernière condition étant particulièrement stricte et pouvant ne conduire à aucune décision.

Le maximum de crédibilité avec rejet [231] :

$$\begin{aligned} x \in C_i \text{ si } & Bel(C_i)(x) = \max \{Bel(C_k)(x), 1 \leq k \leq n\} \\ & \text{et} \quad Bel(C_i)(x) \geq Bel(C_i^C)(x), \end{aligned} \quad (8.49)$$

qui exprime que la décision doit être suffisamment non ambiguë puisque la condition sera vérifiée si la masse est très focalisée sur C_i .

Le maximum de probabilité pignistique, celle-ci étant définie par [315] :

$$\forall C_j \in D, \text{BetP}(C_j) = \sum_{C_j \in A} \frac{m(A)}{|A|(1 - m(\emptyset))}, \quad (8.50)$$

où $|A|$ désigne le cardinal de A , qui permet de repasser à un contexte probabiliste souvent souhaité pour la prise de décision (ou le pari) ou pour associer cette décision à d'autres critères probabilistes, par exemple dans le cadre des champs de Markov pour des critères de régularisation spatiale [334].

Des règles mixtes ont également été proposées, dans lesquelles la plausibilité est utilisée pour certaines classes et la croyance pour d'autres. Cela permet de favoriser la détection des classes pour lesquelles on considère la plausibilité [243].

La décision peut également être prise en faveur d'une disjonction. Elle est alors imprécise mais permet de prendre en compte des mélanges de classes ou des ambiguïtés subsistant après fusion. Ce type de décision est intéressant par exemple pour tenir compte de l'effet de volume partiel et les voxels qui en sont affectés seront ainsi classés comme des voxels de mélange, plutôt que comme des voxels de classes pures, ce qui correspond bien à l'intuition [37]. La décision permet également d'indiquer les éléments pour lesquels la fusion ne suffit pas à lever les ambiguïtés et donc de suggérer l'acquisition de nouvelles informations, ainsi que l'exploite la fusion active [147, 271].

Enfin, des règles de décision avec coût ont été proposées [107]. Pour toute fonction f de D dans \mathbb{R} , les espérances inférieure et supérieure de f relativement à une fonction de croyance Bel , au sens de Dempster, sont définies par :

$$E_*(f) = \sum_{A \subseteq D} m(A) \min_{C_i \in A} f(C_i), \quad (8.51)$$

$$E^*(f) = \sum_{A \subseteq D} m(A) \max_{C_i \in A} f(C_i). \quad (8.52)$$

Les règles de décision avec coût s'obtiennent alors en prenant pour f une fonction qui exprime le coût d'une action lorsque l'élément sur lequel porte la décision appartient à la classe C_i . Cette fonction de coût peut également être introduite dans une règle de décision avec coût probabiliste classique, en utilisant la probabilité pignistique. Ainsi, la décision peut être optimiste si l'espérance inférieure est minimisée, pessimiste si l'espérance supérieure est minimisée, ou intermédiaire si la probabilité pignistique est utilisée.

8.2.6 Déduction axiomatique de la règle de combinaison de Dempster-Shafer

Il a souvent été reproché à la théorie des croyances de Dempster-Shafer d'imposer une règle de combinaison (la règle orthogonale de Dempster [302], équation 8.37) « ad hoc », sans justification théorique.

Plusieurs travaux récents ont cherché à justifier cette règle, par exemple ceux de Gacôgne à partir du concept d'accentuation [146]⁸, ceux de Dubois et Prade [125], qui justifient mathématiquement l'emploi du produit pour combiner les masses à partir de la notion de séparabilité des sources, ou ceux de Smets à partir d'un modèle de croyance transférable (« the transferable belief model ») [314]. Les travaux de Smets proposent la justification la plus générale à notre connaissance et ce sont ses arguments que nous décrivons ici. De plus sa démarche est similaire à celle de Cox. Des travaux similaires à ceux de Smets ont été effectués par Klawonn et Schwecke [194].

Axiomes de Smets

La première constatation de Smets concerne le principe d'indifférence (ou principe de raison insuffisante). Affecter la même probabilité à tous les événements (simples) entraîne que des probabilités différentes sont affectées à des réunions d'événements, ce qui, à son sens, ne correspond pas à l'indifférence. Celle-ci s'exprimerait plutôt par l'existence d'une constante c positive ou nulle telle que :

$$\forall A \subset D, A \neq D, \text{Bel}(A) = c, \quad (8.53)$$

⁸Gacôgne montre que la règle de Dempster-Shafer (équation 8.37), dans le cas où l'espace de discernement est réduit à une proposition P et à son contraire P^C , peut se déduire de la notion d'accentuation. Une fonction d'accentuation est telle qu'elle diminue les degrés de confiance inférieurs à 0,5 et augmente ceux qui sont supérieurs à 0,5, les rendant ainsi plus proches de degrés binaires (cela correspond à la notion de renforcement que l'on trouve dans la théorie algébrique des semi-groupes ordonnés ainsi que dans celle des ensembles flous). La fonction rationnelle de $[0,1]$ dans $[0,1]$ de plus bas degré qui soit une fonction d'accentuation est définie par $\frac{x^2}{2x^2 - 2x + 1}$, et c'est celle qui permet de montrer l'analogie avec Dempster-Shafer. Puis cette notion est généralisée à des couples (x, y) , caractérisant une proposition P , tels que $0 \leq x \leq y \leq 1$ (appelés « obligation » et « éventualité »), donc proches des notions de crédibilité et plausibilité ou encore de nécessité et possibilité). L'accentuation d'un tel couple est définie par :

$$\#(x, y) = \left(\frac{2xy - x^2}{1 - 2x + 2xy}, \frac{y^2}{1 - 2x + 2xy} \right).$$

Ces deux valeurs correspondent exactement à celles que l'on obtiendrait en combinant, par la règle de Dempster-Shafer (équation 8.37), la crédibilité (resp. la plausibilité) d'une proposition P avec elle-même. Si maintenant on dispose de deux jeux de mesures (x, y) et (x', y') sur P , le même type de raisonnement conduit à la justification de la règle de Dempster-Shafer [146]. Il resterait à généraliser cette approche à des espaces de discernement plus complexes.

où D désigne l'espace de discernement. Cela est à l'évidence impossible avec des probabilités, mais l'est dans le contexte de Dempster-Shafer avec des crédibilités. En effet, on a :

$$A \cap B = \emptyset \Rightarrow Bel(A \cup B) \geq Bel(A) + Bel(B) \quad (8.54)$$

d'où $c \geq 2c$ et donc $c = 0$. La fonction de masse représentant l'indifférence (ou l'ignorance totale) est donc définie par :

$$m(D) = 1 \text{ et } \forall A \neq D, m(A) = 0, \quad (8.55)$$

ce qui cette fois est tout à fait satisfaisant. Cette fonction de masse particulière joue un rôle important dans la règle de combinaison puisqu'elle en est l'élément neutre, ce qui confirme son interprétation en termes d'ignorance totale, qui ne peut modifier aucune autre fonction de masse.

La deuxième idée de Smets est celle du modèle de croyance transférable, qui définit le conditionnement. Le problème se pose de la manière suivante : étant donnée une information nouvelle, permettant d'affirmer que la vérité se trouve dans un sous-ensemble B de l'espace de discernement D , comment modifier un jeu de masses m pour prendre en compte cette nouvelle information ? La formulation que propose Smets est la suivante :

$$\begin{aligned} \forall A, m'(A) &= \sum_{X \subseteq B^C} m(A \cup X) \text{ si } A \subseteq B \\ &= 0 \text{ sinon} \end{aligned} \quad (8.56)$$

où m' désigne le nouveau jeu de masses. Ces formules peuvent éventuellement être modifiées par renormalisation⁹. Cette formule s'interprète de la manière suivante. Si l'on décompose un sous-ensemble A en réunion $A_1 \cup A_2$ avec $A_1 \subseteq B$ et $A_2 \subseteq B^C$, la masse $m(A_1 \cup A_2)$ est entièrement transférée sur A_1 (d'où le nom du modèle). Dans les cas particuliers où $A_2 = \emptyset$ ($A \subseteq B$), la masse de A n'est pas modifiée, et si $A_1 = \emptyset$ ($A \subseteq B^C$), la masse de A devient nulle.

Dans la théorie de Shafer [302] et dans la présentation que nous en avons faite plus haut, la formule de conditionnement est déduite de la règle de combinaison, alors qu'ici elle la précède et est simplement construite par des considérations logiques¹⁰.

Dans une troisième étape, Smets définit des axiomes qu'il veut voir vérifiés par la règle de combinaison, notée \oplus :

A1 : $(Bel_1 \oplus Bel_2)(A)$ doit être fonction seulement des fonctions m_1 et m_2 et de A .¹¹

⁹Notons que Smets considère que la normalisation dans Dempster-Shafer n'est pas indispensable et peut même être nuisible dans la mesure où elle masque le conflit [314]. Elle pose également des problèmes de continuité au voisinage du conflit total [125].

¹⁰La règle de conditionnement sur les plausibilités

$$Pls(A|B) = \frac{Pls(A \cap B)}{Pls(B)}$$

peut elle-même être justifiée en posant

$$Pls(A \cap B) = T[Pls(A|B), Pls(B)]$$

et en appliquant une démarche analogue à celle ayant conduit à la première équation fonctionnelle (partie 3.3.2) [125].

¹¹Dans [194], les axiomes utilisés sont pour la plupart similaires à ceux de Smets. La différence essentielle réside dans l'utilisation de relations entre espaces de discernement inclus les uns dans les autres plutôt que de relations de dépendance comme le fait Smets.

A2 : \oplus doit être commutative.

A3 : \oplus doit être associative.

A4 : Si $m_2(B) = 1$, alors $m_1 \oplus m_2$ doit vérifier la loi du conditionnement, c'est-à-dire :

$$\begin{aligned} \forall A, (m_1 \oplus m_2)(A) &= \sum_{X \subseteq B^C} m_1(A \cup X) \text{ si } A \subseteq B \\ &= 0 \text{ sinon.} \end{aligned} \quad (8.57)$$

A5 : La loi doit vérifier une propriété de symétrie interne (invariance par permutation sur les hypothèses simples).

A6 : Pour $A \neq D$, $(m_1 \oplus m_2)(A)$ ne dépend pas de $m_1(X)$ pour $X \subseteq A^C$ (propriété d'autofonctionnalité).

A7 : Il y a au moins 3 éléments dans D .

A8 : La loi doit vérifier une propriété de continuité :

$$\begin{aligned} m_2(A) &= 1 - \varepsilon, m_2(D) = \varepsilon, m_A(A) = 1 \\ \Rightarrow \forall X, \lim_{\varepsilon \rightarrow 0} (m_1 \oplus m_2)(X) &= (m_1 \oplus m_A)(X), \end{aligned} \quad (8.58)$$

m_1 étant une fonction de masse quelconque ; cette propriété permet d'éliminer des cas dégénérés.

Déduction de la règle de combinaison

À partir des axiomes précédents, Smets déduit la seule règle de combinaison possible satisfaisant ces axiomes. Pour cela, il raisonne sur les fonctions de communalité, définies par :

$$\forall A \subseteq D, q(A) = \sum_{A \subseteq X, X \subseteq D} m(X). \quad (8.59)$$

Sa démonstration, que nous ne détaillerons pas ici, repose sur les propriétés des normes triangulaires et des fonctions absolument monotones et comporte trois parties. Tout d'abord, les axiomes A1 à A4 entraînent l'existence d'une fonction f telle que le résultat de la combinaison ne dépende que de A et des communalités des sous-ensembles inclus dans A :

$$(q_1 \oplus q_2)(A) = f[A, \{q_1(X) ; X \subseteq A, q_2(X) ; X \subseteq A\}]. \quad (8.60)$$

Ensuite, en ajoutant les axiomes A5 et A6, il est possible de préciser la forme de f , qui ne dépend plus que de A et de $q_1(A)$ et $q_2(A)$:

$$(q_1 \oplus q_2)(A) = f[A, q_1(A), q_2(A)]. \quad (8.61)$$

Enfin, l'ensemble des axiomes A1 à A8 permet de déterminer la forme finale de la règle de combinaison :

$$(q_1 \oplus q_2)(A) = q_1(A)q_2(A). \quad (8.62)$$

On retrouve bien la règle de Dempster-Shafer sur les communalités, et on en déduit la combinaison des fonctions de masse ou de crédibilité.

L'avantage de l'approche de Smets est qu'elle repose sur des axiomes dont l'interprétation correspond bien à l'intuition. Il est également plus facile de les réfuter ou de les modifier s'ils ne correspondent pas au problème que l'on se pose.

Relation avec les postulats de Cox

Dans cette partie, nous tâchons d'établir des liens entre les postulats de Cox et les axiomes de Smets, afin de montrer pourquoi ils conduisent à des théories différentes.

Tout d'abord, il convient de préciser dans quel cadre nous nous plaçons pour cette comparaison. En effet, les travaux de Cox et de Smets ne traitent pas exactement du même problème, puisque ceux de Cox s'attachent à justifier les probabilités et leurs propriétés, alors que ceux de Smets s'attachent à justifier une règle de combinaison. Cependant, il est intéressant de constater certaines analogies entre les deux ensembles d'axiomes. De plus, les axiomes de Cox permettent de déduire la règle de Bayes (équation 3.8), qui est utilisée en traitement des images pour fusionner des informations, par l'intermédiaire de probabilités conditionnelles (voir section 8.1). Nous adoptons le point de vue de la fusion de données pour cette comparaison. Il serait également intéressant de comparer les axiomes de Cox avec ceux introduits par Smets pour justifier les fonctions de crédibilité et de plausibilité [317], mais cette comparaison ne porterait que sur les étapes de modélisation du processus de fusion, et non sur les étapes de combinaison elles-mêmes.

L'axiome A1, exprimant la dépendance entre des degrés de confiance et leur combinaison, est moins strict que les postulats de Cox. En effet, le postulat de cohérence entraînait l'existence d'une relation définissant le degré de confiance dans AB n'impliquant que les propositions A et B , sous la forme des degrés de confiance affectés à $[A|B]$ et $[B]$ (ou $[B|A]$ et $[A]$) mais pas à d'autres propositions. L'axiome de Smets, plus général, correspond à la possibilité offerte par la théorie de Dempster-Shafer de travailler sur des sous-ensembles et non plus simplement sur des singletons.

Les axiomes A2, A3 et A5 correspondent à des propriétés de la logique classique des propositions. Les postulats de Cox (en particulier le postulat 4) impliquent également que la logique déductive soit retrouvée comme cas particulier. Les deux approches coïncident donc sur ce point. Ces axiomes sont utilisés dans la démarche de Cox pour éliminer certaines formes de relations fonctionnelles entre $[AB|e]$ et les autres degrés de confiance, pour ne retenir que la seule forme cohérente avec la logique déductive :

$$[AB|e] = T([A|Be], [B|e]) = T([B|Ae], [A|e]). \quad (8.63)$$

De même ces axiomes sont utilisés dans la démonstration de Smets pour éliminer des dépendances et montrer que $(q_1 \oplus q_2)(A)$ ne dépend dans un premier temps que de A , et de $q_1(X)$ et $q_2(X)$ pour $X \subseteq A$, puis, dans un deuxième temps, que de $q_1(A)$ et $q_2(A)$.

L'axiome A4 (conditionnement) traduit une idée très proche de celle du conditionnement hypothétique déduite du cinquième postulat de Cox. La différence essentielle est que le conditionnement y est exprimé plus comme une relation de compatibilité que comme une probabilité conditionnelle.

On ne trouve pas dans les axiomes de Smets d'équivalent au postulat 3 de Cox (universalité). Cela se justifie par la base même de la théorie des croyances, où l'on ne caractérise plus les propositions par un seul nombre mais plutôt par deux (crédibilité et plausibilité), et où l'on accepte de ne pas affecter de degré de confiance à une proposition bien définie ¹². Cette

¹²Cela peut se faire par exemple en affectant une masse nulle à cette proposition A . Cela ne signifie pas pour autant qu'une confiance nulle soit attribuée à A , puisque la crédibilité $Bel(A)$ et la plausibilité $Pls(A)$ ne sont pas nécessairement nulles, des masses non nulles pouvant être affectées à des propositions B telles que $A \cap B \neq \emptyset$. Cela signifie simplement que l'on n'affecte pas de degré de confiance spécifiquement à A .

souplesse permet de résoudre facilement des problèmes liés au manque d'information : si une source n'est pas capable de donner d'informations sur A mais qu'elle en donne par exemple sur $A \cup B$, cette situation est naturellement prise en compte par la théorie des croyances en affectant une masse à $A \cup B$ et pas à A , alors qu'elle nécessite souvent l'introduction d'hypothèses ou de modèles dans la théorie des probabilités pour pouvoir affecter un degré de confiance à A . Du point de vue de la comparaison des degrés de confiance, elle peut être menée dans la théorie des croyances à deux niveaux, soit sur les crédibilités, soit sur les plausibilités, conduisant à des conclusions qui ne sont pas nécessairement équivalentes.

L'axiome A6 de Smets n'entraîne pas que A et A^C soient interchangeables, alors que cette propriété est utilisée explicitement par Cox pour déduire la deuxième équation fonctionnelle (équation 3.6). En effet, des sous-ensembles X peuvent intervenir à la fois dans $(m_1 \oplus m_2)(A)$ et dans $(m_1 \oplus m_2)(A^C)$. On ne peut donc pas en déduire de relation de complémentarité sur m . Celle-ci est remplacée par une relation de dualité entre Bel et Pls .

Enfin, les axiomes A7 et A8 sont considérés par Smets lui-même comme des axiomes techniques permettant de faire les démonstrations. La régularité imposée aux fonctions est à mettre en parallèle avec les hypothèses de régularité faites pour les deux équations fonctionnelles 3.5 et 3.6 de Cox.

Ces différences entre les deux théories ont des conséquences aux trois niveaux qui constituent classiquement le processus de fusion, celui de la modélisation des fonctions de confiance, celui de la combinaison de ces fonctions déterminées à partir d'informations fournies par plusieurs sources, et celui de la décision finale :

- d'abord au niveau de l'étape de modélisation puisque cette étape est fortement contrainte par les deux relations fonctionnelles (équations 3.5 et 3.6) dans la fusion probabiliste alors que la théorie des croyances permet de s'adapter avec beaucoup de souplesse à beaucoup de situations (nous avons cité l'exemple de capteurs ne donnant d'information que sur la réunion de deux classes, sans les différencier) ;
- au niveau de la combinaison des fonctions de confiance, les postulats imposent la règle de Bayes d'une part, la règle de Dempster d'autre part, et leurs différences proviennent en particulier des contraintes plus souples imposées par le conditionnement de Smets que par le conditionnement hypothétique de Cox ;
- enfin, au niveau de la prise de décision, étape ultime du processus de fusion, les différences proviennent surtout de la comparaison des degrés de confiance, laissant la place à plusieurs types de décision dans la théorie de Dempster-Shafer.

8.2.7 Exemple d'application en imagerie médicale

L'application que nous avons choisie ici pour mettre en évidence les potentialités de la théorie des fonctions de croyance est celle de la classification d'images IRM présentant une pathologie appelée adrénoleukodystrophie (ALD), acquises avec deux temps d'écho [37]. L'obtention de mesures significatives pour les médecins nécessite une segmentation à la fois des zones pathologiques et des ventricules, visibles sur des images différentes. Les images initiales sont présentées figure 8.4. Cette figure montre une bonne discrimination entre le cerveau, les ventricules (V) et le liquide céphalo-rachidien (LCR) sur la première image, mais celle-ci ne permet pas de bien distinguer la matière blanche (MB) de la matière grise (MG), ni la MB de l'ALD. Au contraire, la zone d'ALD est bien visible sur la deuxième image (en blanc). Cette image présente de faibles différences entre MB et MG, mais les ventricules

ont presque les mêmes niveaux de gris que la MG et leurs contours sont imprécis.

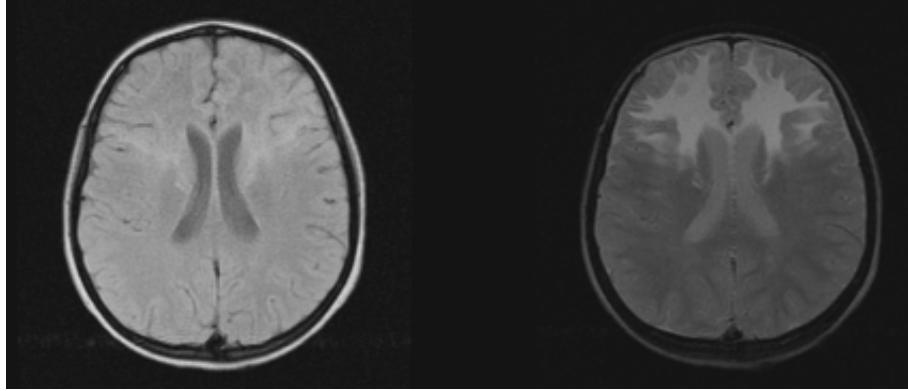


FIG. 8.4 – Exemple d'une coupe IRM du cerveau acquise avec deux temps d'échos (hôpital Saint-Vincent de Paul, service de radiologie, Professeur Catherine Adamsbaum). La zone pathologique correspond aux zones les plus blanches dans le haut de l'image.

Ces deux images constituent donc un exemple typique d'illustration de la fusion par la théorie des fonctions de croyance. Nous ne présentons ici que les résultats obtenus pour trois classes ($C_1 = \text{MB+MG}$, $C_2 = \text{V+LCR}$, and $C_3 = \text{ALD}$).

Les éléments focaux sont définis de manière supervisée, par une méthode de raisonnement prenant en compte les connaissances dont nous disposons et les caractéristiques des images par rapport aux classes d'intérêt. Pour l'exemple décrit ici, les éléments focaux de la fonction de masse m_1 attachée à la première image sont C_2 , $C_1 \cup C_3$, puisque C_1 et C_3 ne sont pas bien discriminés sur cette image. Des fonctions de masse nulles sont affectées aux autres hypothèses composées, puisque les classes correspondantes ne peuvent pas être confondues. Sur la seconde image, il est au contraire difficile de séparer le cerveau et les ventricules, et donc les éléments focaux de m_2 sont C_3 et $C_1 \cup C_2$. Nous discuterons plus loin l'introduction de l'ignorance globale et celle d'une masse représentant explicitement le volume partiel. Les fonctions de masse sont choisies de forme simplement trapézoïdale, dont les paramètres sont déterminés automatiquement sur les histogrammes [50]. Il s'agit bien sûr d'un modèle grossier, mais qui s'est avéré suffisant pour cette application. Puis les fonctions sont normalisées de telle sorte que la contrainte de normalisation $\sum_{A \subseteq D} m(A) = 1$ soit satisfaite. Avec ce modèle, la classification est effectuée seulement d'après les niveaux de gris, et la fusion est effectuée au niveau du pixel, donc sans information spatiale.

La combinaison conjonctive par la règle de Dempster ne donne que des éléments focaux qui sont les singletons C_1 , C_2 , C_3 . Le conflit n'est pas nul dans ce cas.

La dernière étape est celle de la décision. Prendre toujours une décision en faveur d'une hypothèse simple nous constraint en réalité à prendre toujours une décision nette, ce qui n'est pas adapté à toutes les situations réelles en imagerie médicale, où les pixels peuvent appartenir à la réunion de classes mais à aucune d'elles strictement. Cependant, comme $\text{Bel}(A) \geq \text{Bel}(C_i)$ pour tout $C_i \in A$ et que $\text{Bel}(D) \geq \text{Bel}(A)$, un certain nombre de précautions doivent être prises pour pouvoir décider en faveur d'une hypothèses composée. On peut par exemple imaginer de prendre une décision en faveur d'une hypothèse composée si les arguments portant sur les hypothèses simples ne sont pas assez forts.

Ainsi, figure 8.5, la décision a été prise selon le maximum de croyance sur toutes les

hypothèses sauf D . Notons que dans ce cas simple, le maximum de crédibilité est équivalent au maximum de plausibilité, puisque $m_1 \oplus m_2$ est une fonction de masse bayésienne. Avec cette règle, la décision est prise en faveur d'une hypothèse simple aux points où les autres masses sont nulles et en faveur d'une hypothèse composée sinon. On obtient de cette manière des résultats intéressants puisque les points de volume partiel sont détectés comme une hypothèse composée, tandis que les zones sans ambiguïté sont bien segmentées. La figure 8.6 présente les résultats obtenus en prenant une décision en faveur d'une hypothèse simple ayant le maximum de croyance.

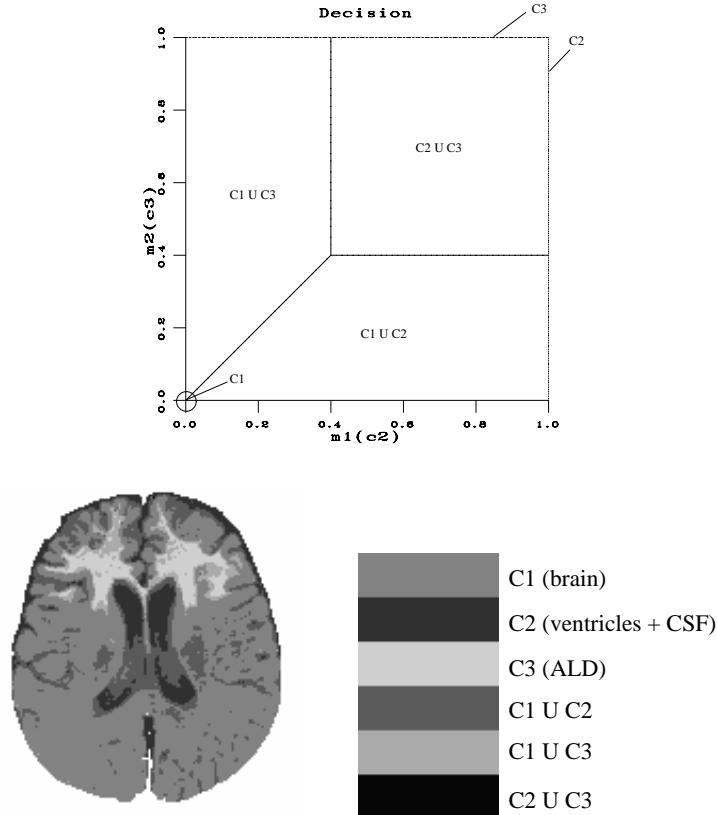


FIG. 8.5 – Les différentes zones de décision en fonction des valeurs de $m_1(C_2)$ et $m_2(C_3)$ et image de décision, en prenant le maximum de croyance sur toutes les hypothèses sauf D

L'étude de l'influence de la pondération de $m_1(C_1 \cup C_3)$ et $m_2(C_1 \cup C_2)$, relativement à $m_1(C_2)$ et $m_2(C_3)$ a montré que le calcul du conflit ne présente pas d'évolution systématique dans le sens de l'augmentation ou de la diminution du conflit. Par exemple, si $m_2(C_3)$ est faible, le conflit diminue si le poids donné à $m_1(C_2)$ augmente (dans ce cas, les deux sources sont en meilleure concordance sur C_2 , c'est-à-dire la classe des ventricules). Le calcul des zones de décision (comme figures 8.5 et 8.6) montre que si le poids de $m_2(C_3)$ et $m_1(C_2)$ augmente, la zone de décision en faveur de C_1 diminue (lorsque les décisions sont prises sur les hypothèses simples). Cela peut également être observé sur les images de décision : la figure 8.7 présente un agrandissement des images de décision, où les ventricules et le LCR sont mieux détectés si le poids sur $m_2(C_3)$ et $m_1(C_2)$ augmente. Des différences sont aussi apparentes sur les petites branches de l'ALD qui sont également mieux détectées. Une remarque très importante est que les zones qui diffèrent sont toutes classifiées en faveur de l'hypothèse $C_1 \cup C_2$ (c'est-à-dire cerveau ou ventricules et LCR) lorsque la décision est prise

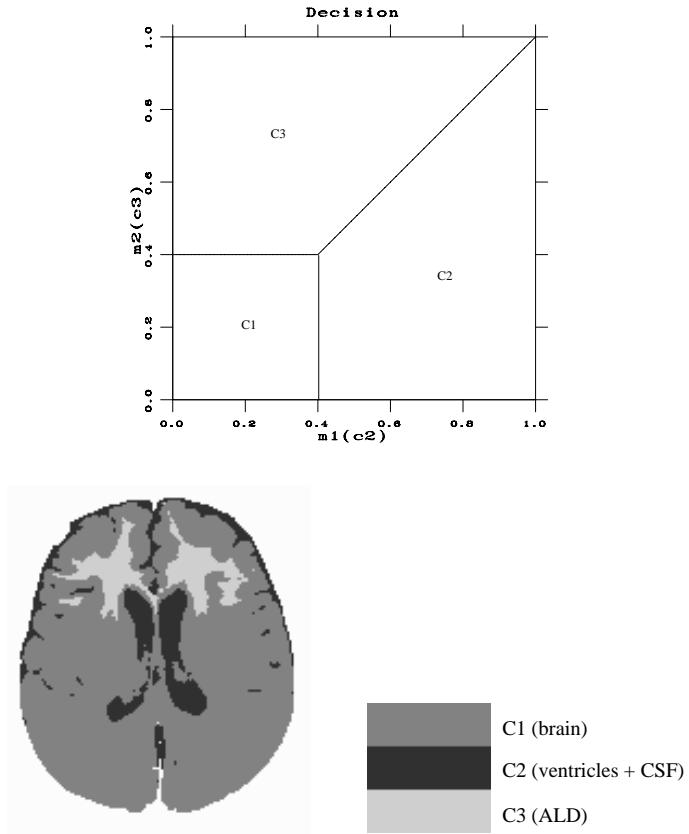


FIG. 8.6 – Les différentes zones de décision en fonction des valeurs de $m_1(C_2)$ et $m_2(C_3)$ et image de décision, en prenant le maximum de croyance sur les hypothèses simples

sur toutes les hypothèses sauf D (respectivement en faveur de l'hypothèse $C_1 \cup C_3$ pour les branches de l'ALD). Avec cette règle de décision, il n'y a pratiquement aucune différence sur les images de décision, montrant donc la robustesse par rapport à la pondération. Cela nous conforte dans l'idée que les hypothèses composées ne devraient pas être si souvent négligées.

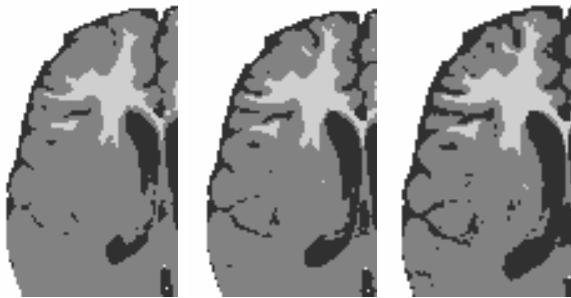


FIG. 8.7 – Décision par le maximum de croyance sur les hypothèses simples, pour un poids croissant sur $m_2(C_3)$ et $m_1(C_2)$

Examinons maintenant l'influence de l'ignorance globale, modélisée par une masse sur D introduite par affaiblissement en fonction de coefficients de fiabilité de chaque source. Cette fois, le résultat de la combinaison $m_1 \oplus m_2$ n'est plus une fonction de masse bayésienne, et

la décision par le maximum de croyance n'est plus équivalente à la décision par le maximum de plausibilité. On observe effectivement des petites différences sur les images de décision. Nous avons observé seulement de très légères différences lorsque la décision est prise sur les hypothèses simples. Cependant, les décisions sur toutes les hypothèses sauf D sont toujours prises en faveur d'une hypothèse composée, comme cela était attendu. Dans notre application, nous n'avons pas d'argument particulier pour dire qu'une image est plus ou moins fiable que l'autre (globalement, car cela n'est pas vrai classe par classe). C'est pourquoi il n'est pas très utile d'affecter une masse à $m(D)$ et cela est confirmé par les résultats obtenus. Au contraire, nous avons des arguments forts pour l'ignorance partielle, dépendant de l'image, qui conduit à des ambiguïtés entre classes, et cela a été introduit par des masses sur les hypothèses composées. Encore une fois, cela justifie la méthode utilisée pour affecter les masses, reposant sur une modélisation du problème, plutôt que la méthode classique à partir des probabilités et de facteurs de fiabilité globale.

Les images utilisées dans cet exemple sont acquises avec des coupes assez épaisses, induisant un fort effet de volume partiel, surtout entre la matière blanche et l'ALD (voir figure 8.4). Nous introduisons maintenant cette connaissance de manière explicite comme une masse sur $C_1 \cup C_3$ dans la deuxième image (fonction en trapèze déduite de l'histogramme). Le conflit est alors réduit. La combinaison $m_1 \oplus m_2$ n'est plus une fonction de masse bayésienne, et les zones de décision sont modifiées, comme le montre la figure 8.8.

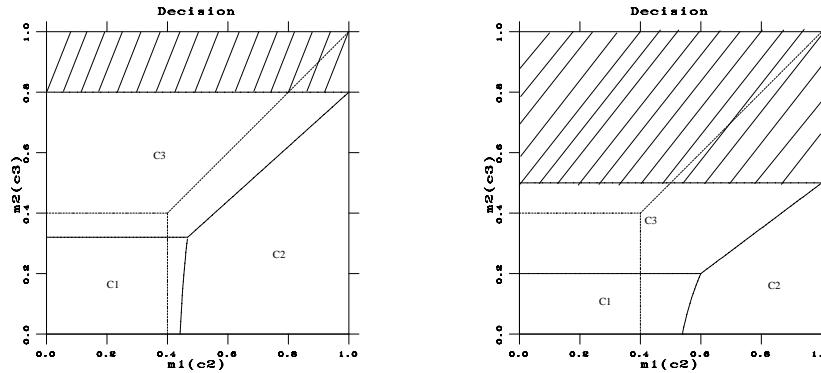


FIG. 8.8 – Zones de décision en fonction des valeurs de $m_1(C_2)$ et $m_2(C_3)$ en introduisant une masse croissante sur $C_1 \cup C_3$ pour la deuxième image ($m_2(C_3)$ doit être inférieur à $1 - m_2(C_1 \cup C_3)$). Les traits en pointillés représentent les limites précédentes (voir figure 8.6).

Les images de décision sur la figure 8.9 présentent les résultats obtenus respectivement sur toutes les hypothèses sauf D et sur les hypothèses simples seulement, d'abord avec $m_2(C_1 \cup C_3) = 0$ et ensuite avec une pondération croissante accordée à $m_2(C_1 \cup C_3)$. Cette figure montre que la décision sur toutes les hypothèses inclut toutes les zones de volume partiel entre la MB et l'ALD dans $C_1 \cup C_3$, et ne change pas si le poids de $m_2(C_1 \cup C_3)$ augmente, illustrant ici encore la robustesse par rapport à la pondération. Au contraire, sur les images de décision sur les hypothèses simples seulement, de plus en plus de points de volume partiel sont inclus dans l'ALD. Cette modélisation permet d'imiter la manière de prendre une décision du médecin, dépendant de son objectif. Sur l'image la plus à gauche, où le volume partiel n'est pas pris en compte, la zone classifiée comme ALD ne présente pas d'ambiguïté (correspondant à l'ALD jj pur jj , sans mélange). Sur l'image la plus à droite au contraire, tout le volume partiel est inclus dans l'ALD (cela correspond à la segmentation

effectivement obtenue manuellement par les médecins), et les zones classifiées dans le cerveau ne contiennent pas de région ambiguë.

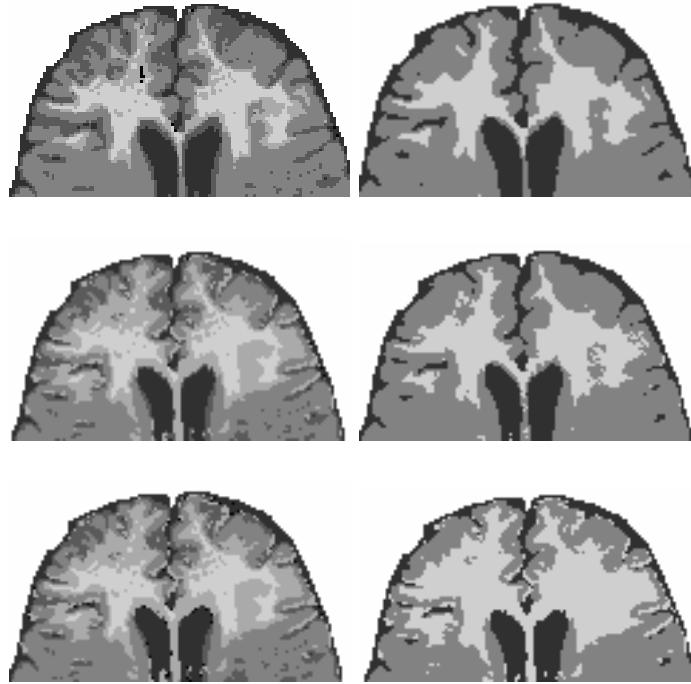


FIG. 8.9 – Décision par le maximum de croyance sur toutes les hypothèses sauf D (première colonne) et sur toutes les hypothèses simples (deuxième colonne), sans masse sur $C_1 \cup C_2$ (en haut) et en introduisant une masse sur $C_1 \cup C_2$ (représentant l’effet de volume partiel entre le cerveau et l’ALD) dans la deuxième image, avec un poids croissant (au milieu et en bas).

Nous avons donc ainsi essayé d’illustrer quelques caractéristiques de la théorie des fonctions de croyance qui peuvent être exploitées en fusion d’images pour la classification, la segmentation ou la reconnaissance, et qui constituent des avantages par rapport aux approches classiques probabilistes et bayésiennes. Elles expriment la grande flexibilité des modélisations possibles, prenant en compte à la fois l’incertitude et l’imprécision, l’ignorance partielle ou globale, la fiabilité des sources, la capacité de chaque source à fournir des informations fiables ou non sur chaque classe, des informations *a priori* qui ne sont pas forcément représentables par des probabilités, etc. L’application présentée illustre bien ces divers avantages. Tout d’abord, une modélisation adaptée au problème est possible, en particulier en affectant des masses sur les hypothèses composées, exprimant par exemple qu’une source ne permet pas de différencier correctement deux classes, ou encore modélisant l’effet de volume partiel. Les probabilités ne sont pas bien adaptées à la modélisation du volume partiel à la limite entre deux classes. Une solution est proposée ici, en affectant une masse directement à la réunion de ces classes, ce qui conduit ici encore à une interprétation satisfaisante. Ensuite, même une définition relativement grossière des fonctions de masse et de leur pondération relative s’est avérée suffisante et robuste. Enfin, la décision a été prise selon deux règles : une règle classique où la décision est toujours prise en faveur d’une hypothèse simple, et une seconde où l’on peut également décider en faveur d’une hypothèse composée. Cette dernière règle correspond mieux à la réalité, en mettant en évidence les zones de volume partiel, et en

s'adaptant au mode de raisonnement du médecin.

8.3 Fusion floue et possibiliste

8.3.1 Modélisation

Parmi les techniques non probabilistes qui ont fait leur apparition depuis une dizaine d'années en fusion, la théorie des ensembles flous fournit un très bon outil pour représenter explicitement des informations imprécises, sous la forme de fonctions d'appartenance [17, 189, 358] comme nous l'avons vu plus haut. La mesure $M_i^j(x)$ introduite chapitre 6 prend alors la forme :

$$M_i^j(x) = \mu_i^j(x), \quad (8.64)$$

où $\mu_i^j(x)$ désigne par exemple le degré d'appartenance de x à la classe C_i selon la source I_j ou la traduction d'une information symbolique exprimée par une variable linguistique (voir par exemple [101]).

On trouve dans la littérature essentiellement deux approches pour l'utilisation des ensembles flous en traitement d'image [34] : la première est de type plutôt symbolique et exprime sous forme de règles floues l'appartenance de certaines structures à une classe en fonction des mesures obtenues par traitement d'image ; la seconde utilise les ensembles flous pour représenter directement les classes ou structures dans l'image, recouvrant spatialement les objets d'une fonction d'appartenance. Considérons l'exemple de la classe « route » dans une image satellitaire. Dans la première approche, on décrira la route sous forme linguistique du type « une route est une structure plutôt allongée ». L'appartenance d'un objet à la classe route sera alors représentée par une fonction associant à sa longueur un degré dans $[0, 1]$. Un algorithme quelconque de détection de contours parallèles permettra alors d'affecter aux objets qu'il détecte un degré d'appartenance à la classe route suivant leur longueur. Dans la deuxième approche, la route sera directement représentée sur l'image par un ensemble flou, avec des degrés d'appartenance forts au centre de la chaussée et voisins de 0 dans les champs ou les forêts.

Ces fonctions ne souffrent pas des contraintes axiomatiques imposées aux probabilités et offrent donc une plus grande souplesse lors de la modélisation. Cette souplesse peut être considérée comme un inconvénient puisqu'elle laisse facilement l'utilisateur démunir pour définir ces fonctions. L'inconvénient des ensembles flous est qu'ils représentent essentiellement le caractère imprécis des informations, l'incertitude étant représentée de manière implicite et n'étant accessible que par déduction à partir des différentes fonctions d'appartenance.

La théorie des possibilités [126, 361], dérivée des ensembles flous, permet de représenter à la fois l'imprécision et l'incertitude, par l'intermédiaire de distributions de possibilités π sur un ensemble S et de deux fonctions caractérisant les événements : la possibilité Π et la nécessité N .

Une distribution de possibilité s'interprète comme une fonction donnant le degré de possibilité pour qu'une variable prenne la valeur s , S étant le domaine des valeurs de la variable. La distribution π s'interprète alors comme la fonction d'appartenance au sous-ensemble flou de S des valeurs possibles pour cette variable. Dans le cadre de la fusion numérique, une application possible de cette théorie consiste à prendre $S = D$ (l'ensemble des classes) et à

définir la mesure M_i^j par :

$$M_i^j(x) = \pi_j^x(C_i), \quad (8.65)$$

c'est-à-dire comme le degré de possibilité pour que la classe à laquelle appartienne x prenne la valeur C_i , selon la source I_j . On définit ainsi une distribution de possibilité par source et par élément x . La possibilité et la nécessité pour une classe C_i s'écrivent alors :

$$\Pi_j(\{C_i\}) = \pi_j(C_i), \quad N_j(\{C_i\}) = \inf \{(1 - \pi_j(C_k)), C_k \neq C_i\}. \quad (8.66)$$

Pour un sous-ensemble quelconque A de D , la possibilité et la nécessité sont calculées d'après les formules 4.55 et 4.57.

Cette modélisation suppose que les classes sont nettes, alors que le modèle flou défini par l'équation 8.64 suppose que les classes sont floues.

De manière plus générale, on trouve dans la littérature trois interprétations pour le flou, en termes de plausibilités, de similarités, de préférences [133]. On retrouve ces interprétations en fusion de signaux et d'images. L'interprétation en termes de plausibilités se retrouve dans l'appartenance à une classe, dans la définition d'un objet spatial flou (objet de limites imprécises). L'interprétation en termes de similarités est celle de la définition d'une classe floue sur un espace de caractéristiques comme fonction de la distance à un prototype par exemple, de variables linguistiques représentant des informations ou des connaissances sur des objets spatiaux, ou encore de degrés de satisfaction d'une relation, d'une contrainte. Enfin, les préférences se retrouvent dans l'expression de critères de choix (par exemple pour des applications de planification en robotique), souvent liés à des contraintes ou des connaissances externes à l'image.

8.3.2 Définition des fonctions d'appartenance ou des distributions de possibilité

La construction des fonctions d'appartenance ou distributions de possibilités peut être effectuée de plusieurs manières.

Dans la plupart des applications, cette construction est faite soit en s'inspirant directement des méthodes d'apprentissage probabiliste, soit par des heuristiques, soit par des méthodes neuromimétiques permettant d'apprendre les paramètres de formes particulières de fonctions d'appartenance, soit enfin par la minimisation de critères de classification [23]. Décrivons maintenant les principales méthodes.

Une première méthode consiste à définir la fonction d'appartenance d'une classe floue à partir de la fonction d'intensité I (les niveaux de gris) de l'image :

$$\mu_i(x) = F_i[(I(x)], \quad (8.67)$$

où F_i est une fonction à déterminer en fonction du problème. Les fonctions les plus utilisées sont des fonctions de normalisation ou des fonctions en S [260] (ce qui revient à considérer que les parties claires de l'image ont une appartenance élevée à la classe), des fonctions Π (monomodales, elles associent la classe à une plage de niveaux de gris aux frontières imprécises), ou encore des fonctions multimodales.

Ces fonctions sont souvent déterminées de manière supervisée, mais elles peuvent également être apprises, par exemple à partir d'algorithmes de classification automatique tels que les

C-moyennes floues [23] ou les C-moyennes possibilistes [201] (voir par exemple [24] pour une revue des algorithmes de classification floue). L'inconvénient principal des C-moyennes floues est que les fonctions d'appartenance ont une forme contre-intuitive : les valeurs d'appartenance à une classe ne sont pas décroissantes en fonction de la distance au centre de la classe. Ce problème est évité avec les C-moyennes possibilités.

D'autres caractéristiques peuvent être utilisées dans ce but. Par exemple, l'ensemble des contours d'une image peut être défini par un ensemble flou spatial dont la fonction d'appartenance est une fonction du gradient de l'image :

$$\mu_i(x) = F_i[(\nabla I(x)], \quad (8.68)$$

où F est une fonction décroissante.

Si l'on dispose de détecteurs d'objets particuliers, la fonction d'appartenance à ces objets est alors définie comme une fonction de la réponse à ces détecteurs (le cas des contours fait partie de cette catégorie). Par exemple, un détecteur de routes peut fournir dans une image satellitaire une réponse d'amplitude croissante avec l'appartenance à la route.

Dans le cas de variables linguistiques, les formes des fonctions d'appartenance et leurs paramètres sont souvent définies par l'utilisateur.

L'imprécision spatiale sur la délimitation des classes (si les fonctions d'appartenance sont définies sur l'espace de l'image) peut être introduite à partir d'une détection préliminaire binaire des classes. On construit alors une fonction d'appartenance qui vaut 1 à l'intérieur de la région binaire à une certaine distance des bords, 0 à l'extérieur de cette région également à une certaine distance des bords, et décroissante entre ces deux limites. Par exemple, on peut modéliser une zone d'imprécision au bord de la classe comme la zone comprise entre l'érodé et le dilaté de cet objet, la taille de ces opérations dépendant de l'extension spatiale de l'imprécision que l'on veut représenter. Si R est la région binaire de départ, $E^n(R)$ son érodé de taille n et $D^m(R)$ son dilaté de taille m , la fonction d'appartenance à la classe floue peut être définie par :

$$\begin{aligned} \mu(x) &= 1 \text{ si } x \in E^n(R), \\ \mu(x) &= 0 \text{ si } x \in D^m(R)^C, \\ \mu(x) &= F[d(x, E^n(R))] \text{ sinon} \end{aligned} \quad (8.69)$$

où F est une fonction décroissante de la distance de x à $E^n(R)$.

La construction de distributions de possibilités peut également être effectuée à partir d'un apprentissage probabiliste, puis par une transformation de probabilité en possibilité. Plusieurs méthodes ont été proposées pour cela. L'avantage essentiel en traitement des images est que l'on dispose souvent d'informations statistiques, en particulier l'histogramme, qui se prêtent bien à l'emploi de méthodes d'apprentissage statistiques. On obtient alors des distributions de probabilités p_k . Leur transformation en distributions de possibilités π_k (les deux distributions sont supposées discrètes, et $1 \leq k \leq K$) est effectuée en fonction de différents critères [113, 123, 196], tels que la préservation de l'ordre, des contraintes de normalisation, la conservation de l'incertain mesuré par l'entropie [196], la cohérence $p - \pi$, exprimée par [100] :

$$\forall k, \pi_k \leq p_k,$$

qui n'est pas très satisfaisante (une classe peu probable peut être possible), ou [361] :

$$\sum_{k=1}^K p_k \pi_k = c$$

où c est une constante dans $[0, 1]$, ou encore une relation plus générale sur tous les sous-ensembles A [120] :

$$N(A) \leq P(A) \leq \Pi(A).$$

Une comparaison de ces approches peut être trouvée dans [196].

D'autres méthodes cherchent à estimer directement les fonctions d'appartenance à partir de l'histogramme, afin d'optimiser des critères d'entropie [80] ou de minimum de spécificité et de cohérence [84].

Dans tous les cas, ces méthodes cherchent une ressemblance entre l'histogramme et les fonctions d'appartenance ou les distributions de possibilité, et ne prennent pas en compte les interprétations spécifiques au flou qui invalident certaines de ces ressemblances. Par exemple, les queues des histogrammes correspondent aux classes peu représentées, donc avec des valeurs qui peuvent être très faibles, même si les points concernés appartiennent bien aux classes correspondantes. La méthode proposée dans [50] permet d'éviter ce problème grâce à un critère combinant la ressemblance des fonctions d'appartenance et de l'histogramme là où elle a un sens, et une forme *a priori* des fonctions correspondant à l'interprétation recherchée. Les paramètres des fonctions d'appartenance sont alors estimés pour optimiser ce critère, par une méthode de recuit simulé.

8.3.3 Combinaison

Un des intérêts de la théorie des ensembles flous et des possibilités, outre qu'elle impose peu de contraintes au niveau de la modélisation, est qu'elle offre une grande variété d'opérateurs de combinaison. Nous en présentons les principaux, puis nous donnons quelques indications permettant de choisir un opérateur de fusion en fonction de ses propriétés et de son comportement.

Une caractéristique importante, commune à toutes les théories, de ces opérateurs de combinaison est qu'ils fournissent un résultat de même nature que les fonctions de départ (propriété de fermeture) et qui a donc la même interprétation en termes d'imprécision et d'incertitude. Ainsi, ils permettent de ne prendre aucune décision binaire partielle avant la combinaison, ce qui pourrait conduire à des contradictions difficiles à lever. La décision n'est prise qu'en dernier lieu, sur le résultat de la combinaison.

Opérateurs

Dans la théorie des ensembles flous et des possibilités, de multiples modes de combinaison sont possibles [124, 351]. Parmi les principaux opérateurs on trouve en particulier les t-normes, les t-conormes [241, 299], les moyennes [350, 166], les sommes symétriques, et des opérateurs prenant en compte des mesures de conflit ou encore de fiabilité des sources [129, 112]. La plupart des définitions de ces opérateurs ont été données dans le chapitre 4. Dans ce qui suit, les lettres $x, y, \text{etc.}$ désignent les valeurs à combiner, valeurs dans $[0, 1]$ représentant μ_i^j ou $\pi_j(C_i)$.

Normes et conormes triangulaires. Voir section 4.4.2 dans le chapitre 4.

Opérateurs de moyenne. Voir section 4.4.3 dans le chapitre 4.

Sommes symétriques. Voir section 4.4.4 dans le chapitre 4.

Opérateurs dépendant d'autres informations. Les opérateurs décrits précédemment se caractérisent par le fait que le résultat d'une combinaison ne dépend que des valeurs à combiner. D'autres opérateurs dépendent d'informations supplémentaires, telles que de facteurs de fiabilité ou du conflit. Tous les opérateurs décrits dans ce paragraphe ont été proposés dans [129].

Considérons par exemple le cas de la combinaison conjonctive de deux distributions de possibilités π_1 et π_2 définies sur D . Ce type de combinaison est bien adapté au cas où les distributions ont un recouvrement au moins partiel, c'est-à-dire que certaines classes sont données comme possibles par les deux sources. Si ce n'est pas le cas, les sources sont en conflit, et une mesure du conflit peut être :

$$conf(\pi_1, \pi_2) = 1 - \max_{c \in D} \min(\pi_1(c), \pi_2(c)), \quad (8.70)$$

qui représente 1 moins la hauteur de l'intersection entre les deux distributions (calculée par un min dans cette équation). La combinaison peut être normalisée par cette hauteur, mais cela masque le conflit : une possibilité de 1 est alors affectée aux classes données comme les plus possibles par les deux sources, même si cette possibilité est faible (on retrouve ici le même type de problème que celui mentionné dans la partie 8.2.3 pour la combinaison conjonctive des fonctions de croyance). L'interprétation de cette quantité en termes de conflit correspond bien à l'intuition pour des distributions de possibilités en triangle ou en trapèze (mono-modales de manière générale), mais n'est pas très bien adaptée à des formes quelconques où un seul point peut engendrer une forte valeur de conflit même si les deux distributions ne diffèrent qu'en ce point.

Dans le cas extrême où les distributions sont complètement conflictuelles, la combinaison conjonctive donne une distribution identiquement nulle. Une combinaison disjonctive est alors préférable, et permet de garder toutes les solutions données comme possibles par au moins une des deux sources. L'hypothèse sous-jacente est qu'au moins une des sources est fiable.

Afin de choisir automatiquement le comportement de la fusion en fonction du conflit, des opérateurs dépendant de celui-ci ont été proposés. En voici quelques exemples :

$$\max\left[\frac{\min(\pi_1, \pi_2)}{1 - conf(\pi_1, \pi_2)}, conf(\pi_1, \pi_2)\right], \quad (8.71)$$

$$\min\left[1, \frac{\min(\pi_1, \pi_2)}{1 - conf(\pi_1, \pi_2)} + conf(\pi_1, \pi_2)\right], \quad (8.72)$$

$$\max\left[\frac{\min(\pi_1, \pi_2)}{1 - conf(\pi_1, \pi_2)}, \min[\max(\pi_1, \pi_2), conf(\pi_1, \pi_2)]\right]. \quad (8.73)$$

Les deux premières formes combinent la conjonction normalisée à une distribution constante de valeur le conflit, alors que la dernière forme permet de passer d'une combinaison strictement conjonctive lorsque le conflit est nul à une combinaison strictement disjonctive lorsque le conflit vaut 1 (voir l'exemple de la figure 4.10 du chapitre 4). Cependant cet opérateur n'est pas associatif. Notons que le min pourrait être remplacé par une autre t-norme.

Lorsque les sources sont inégalement fiables et que l'on a une information sur cette fiabilité, le niveau du conflit entre deux sources indique dans quelle mesure on peut prendre en compte l'information de la source la moins fiable. Si par exemple π_1 est plus fiable que π_2 , on peut considérer que si elles sont concordantes, π_2 peut apporter une information et rendre la fusion plus précise par conjonction. Si au contraire les deux sources sont en conflit, il est préférable de ne pas prendre en compte π_2 . L'opérateur suivant modélise ce comportement :

$$\min[\pi_1, \max[\pi_2, \text{conf}(\pi_1, \pi_2)]]. \quad (8.74)$$

Cela ne suppose que de connaître un ordre entre les fiabilités des sources.

Si on a de plus accès à des valeurs numériques de fiabilité (ce qui est beaucoup plus contraignant que l'hypothèse précédente), on peut alors transformer les distributions de possibilités en distributions ayant des fiabilités équivalentes. Soit w_j le coefficient de fiabilité de π_j . Si la source est complètement fiable, ce coefficient vaut 1, et il vaut 0 si la source n'est pas fiable du tout. La transformation de π_j se fait selon la formule :

$$\max(\pi_j, 1 - w_j) \quad (8.75)$$

ce qui revient à faire une disjonction entre π_j et une distribution constante de valeur $1 - w_j$. Ainsi, si la source est complètement fiable, la distribution correspondante n'est pas modifiée, alors que si elle n'est pas du tout fiable, la distribution devient constante et égale à 1, ce qui représente l'ignorance (tout élément de D est complètement possible). Une fois les distributions transformées, elles peuvent alors être combinées de manière conjonctive.

D'autres opérateurs de ce type peuvent être trouvés dans [133], mais nous ne les détaillons pas plus ici.

Ces opérateurs peuvent également utilisés de manière conditionnelle aux classes, pour prendre en compte les spécificités des sources pour chaque classe. Par exemple, deux sources peuvent être en conflit sur une classe mais pas sur les autres, une source peut être fiable pour certaines classes et pas pour d'autres, etc. Bien que ces idées ne soient pas encore beaucoup exploitées en fusion floue d'images, le cadre théorique le permet.

Choix des opérateurs

Le choix d'un opérateur pour la fusion peut se faire selon plusieurs critères présentés dans [36].

Un premier critère est le comportement de l'opérateur. Des comportements sévères, indulgents ou prudents se traduisent sous forme mathématique de conjonction, disjonction ou compromis. Soient x et y deux réels (dans $[0, 1]$) représentant les degrés de confiance à combiner. La combinaison de x et y par un opérateur F est dite :

- conjonctive si $F(x, y) \leq \min(x, y)$ (correspondant à un comportement sévère) ;
- disjonctive si $F(x, y) \geq \max(x, y)$ (comportement indulgent) ;

- de compromis si $x \leq F(x, y) \leq y$ si $x \leq y$, et $y \leq F(x, y) \leq x$ sinon (comportement prudent).

Cette distinction ne suffit pas à classer les opérateurs dont le comportement n'est pas toujours le même. Ainsi, la classification définie dans [36] ne décrit pas les opérateurs seulement comme conjonctifs ou disjonctifs, mais aussi en fonction de leur comportement selon les valeurs des informations à combiner. Ainsi, les trois classes proposées correspondent :

- aux opérateurs autonomes à comportement constant (ACC) : le résultat ne dépend que des valeurs à combiner (le calcul ne fait intervenir aucune autre information) et le comportement est le même quelles que soient ces valeurs ;
- aux opérateurs autonomes à comportement variable (ACV) : le comportement dépend des valeurs numériques des informations à fusionner ;
- aux opérateurs dépendant du contexte (DC), par exemple d'une connaissance plus globale telle que la fiabilité des capteurs, ou encore le conflit entre les sources.

Les opérateurs de fusion floue se répartissent dans les trois classes. En effet, les t-normes, qui généralisent l'intersection ensembliste aux ensembles flous, sont des opérateurs ACC conjonctifs, puisque pour toute t-norme t , on a :

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

A l'opposé, les t-conormes, généralisant la réunion, sont des opérateurs ACC disjonctifs, puisque pour toute t-conorme T , on a :

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

Les opérateurs de moyenne sont également ACC et ont un comportement de compromis, puisqu'ils vérifient :

$$\forall (x, y) \in [0, 1]^2, \min(x, y) \leq m(x, y) \leq \max(x, y).$$

Notons que la fusion bayésienne, où l'opérateur impliqué est un produit, et la fusion des fonctions de croyances par la somme orthogonale de Dempster sont également conjonctives.

Dans la classe des opérateurs ACV, on trouve par exemple certaines sommes symétriques. De manière générale, toute somme symétrique associative σ (sauf les médianes) a le comportement suivant [126] :

- conjonctif si $\max(x, y) < 1/2$: $\sigma(x, y) \leq \min(x, y)$;
- disjonctif si $\min(x, y) > 1/2$: $\sigma(x, y) \geq \max(x, y)$;
- compromis si $x \leq 1/2 \leq y$: $x \leq \sigma(x, y) \leq y$ (et l'inégalité contraire si $y \leq 1/2 \leq x$).

Les sommes symétriques non associatives ont également un comportement qui varie, mais selon des règles moins simples [36].

On trouve également dans la classe ACV les opérateurs proposés dans le système MYCIN pour combiner des facteurs de certitude [306].

Des exemples d'opérateurs DC se trouvent dans la théorie des possibilités. Nous avons présenté plus haut des opérateurs dépendant d'une mesure globale de conflit entre les deux sources d'information [129], applicables aux cas où l'une des deux informations est fiable sans que l'on sache laquelle, de telle sorte que :

- ils sont conjonctifs si les sources sont consonantes (de faible conflit) : dans ce cas, les deux sources sont nécessairement fiables, et donc l'opérateur peut être sévère ;

- ils sont disjonctifs si les sources sont dissonantes (de fort conflit) : une disjonction favorise alors l'ensemble des possibilités données par les deux sources ;
- ils se comportent comme un compromis dans les cas de conflit partiel : ces cas posant le plus de problèmes, les opérateurs sont alors « prudents ».

La difficulté est alors de trouver une bonne mesure de conflit. Celle proposée comme le maximum de l'intersection entre deux distributions de possibilités [129] n'est pas toujours bien adaptée aux problèmes de traitement d'images, en particulier pour la classification de données multisources. Les distances floues (voir par exemple [43]) peuvent apporter des solutions à ce problème.

L'avantage des opérateurs DC pour le traitement d'images est indéniable. En effet, ils permettent de prendre en compte une grande variété de situations, dont plusieurs se produisent fréquemment simultanément en traitement d'images. En voici quelques exemples :

- les sources peuvent être conflictuelles lorsqu'elles donnent des informations sur un type d'événement (une classe par exemple) et consonantes pour une autre classe ;
- les sources peuvent avoir des fiabilités globales différentes ;
- une source peut être fiable pour une classe et peu fiable pour une autre, etc.

Malheureusement ces opérateurs sont encore, selon nous, trop peu développés en traitement d'images, et mériteraient des recherches spécifiques.

Cette classification, qui regroupe tous les opérateurs classiquement utilisés, constitue un premier critère de choix d'un opérateur pour une application spécifique.

Un second critère est donné par les propriétés des opérateurs et leur interprétation en termes de fusion de données incertaines, imprécises, incomplètes ou encore ambiguës.

Les propriétés de commutativité et d'associativité expriment que le résultat de la combinaison est indépendant de l'ordre dans lequel les informations sont combinées. Si la commutativité est satisfaite par tous les opérateurs couramment utilisés, l'associativité ne l'est pas systématiquement (les moyennes et les sommes symétriques ne sont en général pas associatives). Ces deux propriétés sont souvent posées comme les propriétés minimales que les opérateurs de fusion doivent satisfaire. Pourtant, le mode de raisonnement humain ne les respecte pas toujours. Par exemple, un photo-interprète commence souvent par construire une interprétation primaire d'une scène à partir d'une seule image, puis améliore cette interprétation à l'aide des autres images, selon un processus qui n'est clairement pas commutatif.

L'existence d'un élément neutre signifie qu'une source donnant cette valeur n'aura aucune influence sur le résultat de la combinaison et représente une sorte d'indifférence de la source par rapport à l'information recherchée, ou encore une ignorance totale de celle-ci. Un tel élément existe pour les t-normes et les t-conormes.

Autre élément particulier, un élément absorbant signifie qu'une source donnant cette valeur est complètement déterminante sur le résultat de la fusion. De tels éléments existent également pour les t-normes et les t-conormes.

La propriété de croissance est généralement imposée aux opérateurs et correspond bien à l'intuition.

Des conditions aux limites, définissant le comportement des opérateurs lorsque les informations à combiner prennent des valeurs extrêmes, permettent de garantir la compatibilité avec le cas binaire, où toutes les propositions sont soit justes soit fausses (cela correspond

à la contrainte de compatibilité avec la logique déductive imposée par Cox pour définir une logique inductive [93]).

La propriété de continuité satisfaite par la plupart des opérateurs garantit la robustesse de la fusion. Cependant, cette propriété n'est pas toujours nécessaire, puisque les phénomènes naturels (en particulier les phénomènes temporels) ne sont pas toujours continus.

L'idempotence signifie que la donnée d'une information déjà disponible ne va pas changer le résultat de la fusion. Cette propriété n'est pas systématiquement imposée. Elle est vérifiée par les moyennes, par la t-norme min et la t-conorme max (et ce sont les seules). On peut vouloir au contraire que la combinaison de deux valeurs identiques renforce ou affaiblisse le résultat global. Prenons l'exemple de témoignages simultanés identiques. Si les témoins sont de connivence, il n'est pas surprenant qu'ils disent la même chose et l'on combinera donc les degrés de confiance associés de manière idempotente. Si au contraire ils sont complètement indépendants, on renforcera la crédibilité de ce qu'ils disent si l'on leur fait confiance, ou on l'affaiblira si l'on ne leur fait pas confiance. Notons que des règles de combinaison modélisant ces comportements sont connues depuis Bernoulli. De manière générale, on peut considérer que si les sources sont dépendantes (au sens cognitif), l'idempotence peut être imposée, alors que si elles sont indépendantes, des effets de renforcement peuvent être souhaitables.

Dans le même ordre d'idées, la propriété de nilpotence sera imposée par exemple pour combiner des témoignages successifs, afin de modéliser la dégradation de l'information dans une chaîne de témoins qui ne sont pas complètement fiables. Par exemple pour certaines t-conormes, la satisfaction de cette propriété permet d'aboutir à un résultat égal à 1 en combinant un certain nombre de mesures non toutes nulles. Ce type de comportement peut être utile quand les informations résultent d'une longue chaîne de traitements.

Les propriétés du tiers exclu et de non contradiction, satisfaites pour certains opérateurs seulement, ont une interprétation reconnue en termes de raisonnement, dans le domaine de l'intelligence artificielle et du raisonnement approché. En traitement d'images, on trouve des exemples où le tiers exclu n'est pas souhaitable dès que l'on a besoin d'introduire de l'ignorance sur un événement et son complémentaire, et donc de relâcher la contrainte d'exhaustivité faite par exemple en probabilités.

La généralisation de tout ce qui précède à la combinaison de plus de deux informations ne pose pas de difficulté particulière (on retrouve en particulier les mêmes types de comportement, avec des règles un peu plus compliquées pour les opérateurs ACV), sauf pour les opérateurs non associatifs. La question principale pour de tels opérateurs est de savoir dans quel ordre combiner les informations. Plusieurs situations peuvent se produire :

- dans certaines applications, chaque information doit être combinée aux autres dès qu'elle est disponible (par exemple pour pouvoir prendre des décisions partielles à partir des données présentes à chaque instant) : l'ordre est alors fixé par l'ordre d'arrivée des informations ;
- l'ordre peut être imposé par des priorités sur les informations à prendre en compte, et des opérateurs ont été conçus pour répondre à de tels besoins (par exemple pour combiner des requêtes dans des bases de données) ;
- dans les autres situations, il faudra trouver des critères pour trouver un ordre approprié à l'application, en particulier lorsque les informations sont conflictuelles, car les résultats peuvent être très différents selon que les informations consonantes ou conflictuelles sont combinées en premier.

Enfin, l'étude du comportement des opérateurs en termes de qualité de la décision à

laquelle ils conduisent, et de réaction face aux situations conflictuelles conduit à un dernier critère de choix. Un point important, toutefois, concerne le pouvoir discriminant des opérateurs. Les opérateurs fortement conjonctifs ou fortement disjonctifs (par exemple la t-norme et la t-conorme de Lukasiewicz) saturent très vite à 0 ou à 1 et donc sont souvent peu discriminants. Par exemple, avec la t-conorme $F(a, b) = \min(a + b, 1)$, on a $F(0, 5, 0, 5) = 1$, $F(0, 1, 0, 9) = 1$, ou encore $F(0, 8, 0, 8) = 1$, alors que ces trois situations ont des interprétations bien différentes.

La capacité des opérateurs à combiner des informations quantitatives (numériques) ou qualitatives (pour lesquelles seul un ordre est connu) peut être également un critère de choix. Par exemple, le min, le max et tout filtre de rang sont intéressants à ce titre puisqu'ils peuvent combiner les deux types d'informations. En effet, le calcul de $\min(x, y)$, par exemple, ne nécessite que de connaître un ordre entre x et y , mais ne nécessite pas de connaître leur valeur numérique. Les opérations ordinaires sont de plus imposées si l'on veut qu'elles soient invariantes par une transformation croissante des degrés d'appartenance [133].

8.3.4 Décision

La règle principalement utilisée en fusion floue est le maximum des degrés d'appartenance :

$$x \in C_i \text{ si } \mu_i(x) = \max \{\mu_k(x), 1 \leq k \leq n\}, \quad (8.76)$$

où μ_k désigne la fonction d'appartenance à la classe k résultant de la combinaison.

La qualité de la décision est mesurée essentiellement selon deux critères :

- le premier porte sur la « netteté » de la décision : le degré d'appartenance maximum (ou plus généralement celui correspondant à la décision) est comparé à un seuil, choisi selon les applications (et éventuellement selon l'opérateur de combinaison choisi) ;
- le deuxième porte sur le caractère « discriminant » de la décision, évalué par comparaison des deux valeurs les plus fortes.

Dans le cas où ces critères ne sont pas vérifiés pour un élément x , celui-ci est placé dans une classe de rejet, ou reclassifié en fonction d'autres critères, spatiaux par exemple (voir chapitre 8.4).

8.3.5 Exemples d'application

Dans cette section, nous illustrons les méthodes floues sur deux exemples de classification multisource.

Exemple en imagerie satellitaire

Pour le premier, nous reprenons les images SPOT de l'exemple du chapitre 3 (figure 8.1). Les classes considérées sont toujours les villes ou zones urbaines (classe C_1), les fleuves (classe C_2) et une classe C_3 comportant toutes les autres structures (essentiellement des zones de végétation). Cet exemple a été développé dans [77].

Dans un premier temps, un apprentissage supervisé est effectué à partir des histogrammes conditionnels aux classes, soit par un simple lissage de ces histogrammes, soit par une minimisation de la distance aux histogrammes de fonctions paramétriques telles que des gaussiennes

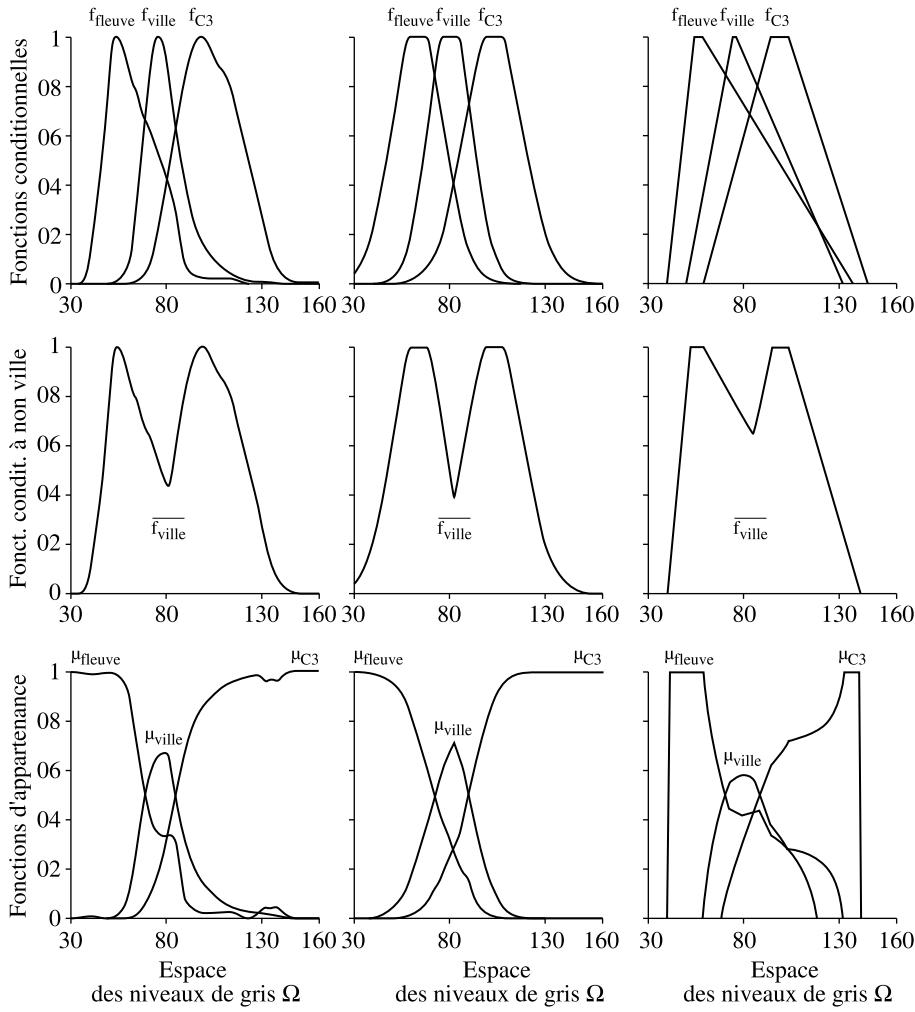


FIG. 8.10 – En haut : fonctions f_i^j conditionnelles aux classes estimées de manière supervisée sur les zones d'apprentissage des images ; au centre, fonctions \bar{f}_{ville}^j estimées par le maximum des fonctions f_{fleuve}^j et $f_{C_3}^j$; en bas, fonctions d'appartenance μ_i^j . A gauche, estimation par les histogrammes normalisés et lissés ; au centre, estimation par les fonctions de type gaussien tronqué ; à droite estimation des fonctions par les fonctions L-R linéaires.

tronquées ou des fonctions de type L-R linéaires par morceaux (fonctions en trapèze). Cette première étape est illustrée sur la première ligne de la figure 8.10.

Cependant, ces fonctions, notées f_i^j pour l'apprentissage de la classe C_i dans l'image j , n'ont pas une interprétation satisfaisante en termes de fonctions d'appartenance. En particulier, les queues de l'histogramme correspondent à des niveaux de gris peu fréquents dans les images, mais les points correspondant appartiennent sans ambiguïté à la classe la plus sombre (respectivement la plus claire). Le passage des fonctions f_i^j aux fonctions d'appartenance μ_i^j se fait par une transformation telle que :

$$\mu_i^j(x) = \lambda_i^j(x) f_i^j(x) \quad (8.77)$$

où x désigne le niveau de gris en un point et les λ_i^j sont des fonctions déterminées de telle sorte que le complémentaire des classes soit défini par la même formule et que l'ordre des fonctions

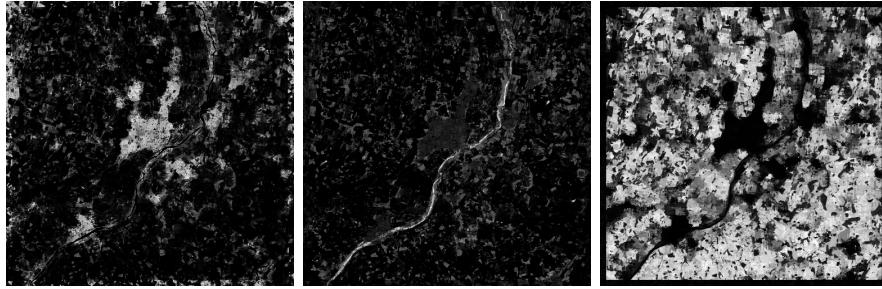


FIG. 8.11 – Images des paramètres H^i des classes ville à gauche, fleuve au centre et C_3 à droite

soit préservé. En prenant par exemple la complémentation la plus simple $c(a) = 1 - a$ pour $a \in [0, 1]$ et en écrivant que le complémentaire d'une classe C_i est également la réunion des classes C_k pour $k \neq i$ (en monde fermé), on obtient :

$$\lambda_i^j(x) = \frac{1}{\overline{\overline{f_i^j}}(x) + f_i^j(x)} \quad (8.78)$$

avec :

$$\overline{\overline{f_i^j}}(x) = \max_{k \neq i} f_k^j(x). \quad (8.79)$$

Les résultats obtenus par cette méthode sont illustrés figure 8.10 et montrent que les fonctions d'appartenance obtenues ont bien le comportement souhaité.

Pour la fusion, un opérateur adaptatif est défini par combinaison d'une t-norme t et d'une t-conorme T , sous la forme $t^{1-\gamma}T^\gamma$. Cet opérateur, appelé parfois opérateur de compensation, peut varier entre la t-norme pour $\gamma = 0$ et la t-conorme pour $\gamma = 1$. L'originalité de l'opérateur défini dans [77] est que γ est défini de manière locale, en chaque point, comme une fonction $H^i(x)$ fonction d'une conjonction normalisée des degrés d'appartenance en x . Cette fonction est illustrée sur la figure 8.11 et donne des valeurs d'autant plus élevées que l'appartenance à la classe est élevée. La t-norme utilisée ici est un min et la t-conorme un max.

La figure 8.12 illustre sur un détail de l'image les résultats obtenus par cet opérateur (dernière ligne), que l'on peut favorablement comparer à ceux obtenus par une simple t-norme. La décision est effectuée par un maximum des degrés d'appartenance en chaque point.

La figure 8.13 montre le résultat de la fusion par l'opérateur adaptatif sur toute l'image, la décision étant toujours simplement prise par un maximum des degrés d'appartenance. Le fleuve est superposé en blanc à l'image originale, ainsi que les contours des zones urbaines. Le reste correspond à la classe C_3 .

Exemple en imagerie médicale

Le deuxième exemple reprend les images de l'exemple de la partie 8.2 sur la classification des images IRM multiéchos du cerveau (figure 8.4). Nous nous plaçons toujours au niveau du pixel et les fonctions d'appartenance sont définies à partir des niveaux de gris, de manière

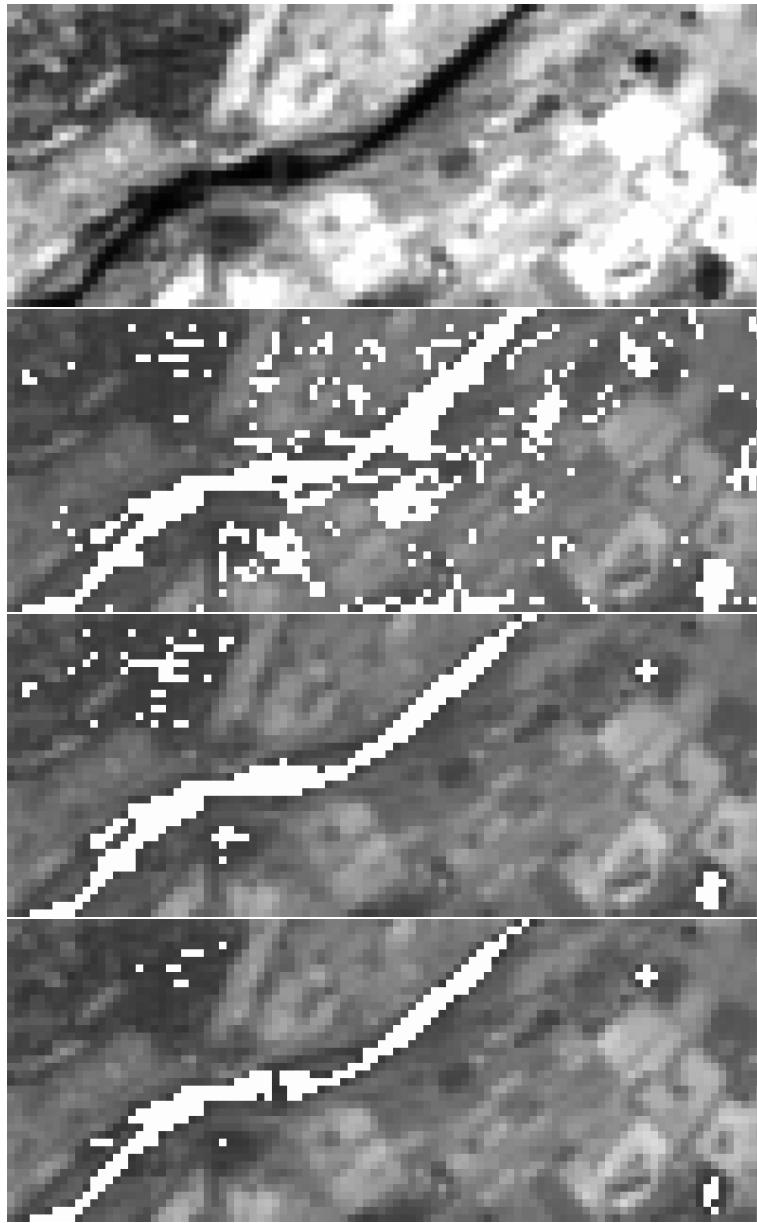


FIG. 8.12 – Détail des images de la détection du fleuve sur la scène de Vignola : image du canal 3 SPOT originale, décision à partir des images H^i , à partir de la t-norme à comportement constant, à partir de l'opérateur de compensation à comportement variable

non supervisée [50]. Trois classes sont considérées : cerveau, ventricules et liquide céphalo-rachidien, zone pathologique.

Ici les opérateurs de combinaison sont choisis de manière adaptative, non plus localement en chaque point, mais en fonction des classes. Comme les deux images fournissent des informations similaires sur les ventricules, les fonctions d'appartenance sont fusionnées par un opérateur de moyenne. Le cerveau sain et la pathologie ne peuvent être distingués dans la première image et l'on ne peut donc apprendre qu'une seule classe, notée μ_c^1 . Dans la seconde image, il est possible d'apprendre les deux classes μ_c^2 (cerveau) et μ_{path}^2 (patholo-



FIG. 8.13 – Résultat de la fusion à partir de l'opérateur de compensation à comportement variable sur la scène de Vignola

gie). Les fonctions μ_c^1 et μ_c^2 sont également combinées par une moyenne arithmétique. Pour la pathologie, on combine μ_c^1 et μ_{path}^2 par une somme symétrique définie par : $\frac{ab}{1-a-b+2ab}$. Cela garantit que la pathologie n'est pas détectée dans les zones où $\mu_{path}^2 = 0$ et renforce l'appartenance à cette classe sinon. Cela permet d'inclure dans la zone pathologique toute la zone de volume partiel où les points sont composés en partie de tissu pathologique.

Après la combinaison, la décision est prise en fonction du maximum d'appartenance. Le résultat est illustré figure 8.14.

Cet exemple illustre l'intérêt des opérateurs choisis de manière adaptative en fonction des informations apportées par les images sur les différentes classes.

8.4 Introduction de l'information spatiale

L'information spatiale est fondamentale en traitement d'images. Son introduction dans les méthodes de fusion est cruciale, et nécessite souvent des développements spécifiques pour adapter les méthodes issues d'autres domaines. Un des objectifs les plus fréquents de ces développements est de garantir que la décision soit spatialement cohérente. Par exemple, en

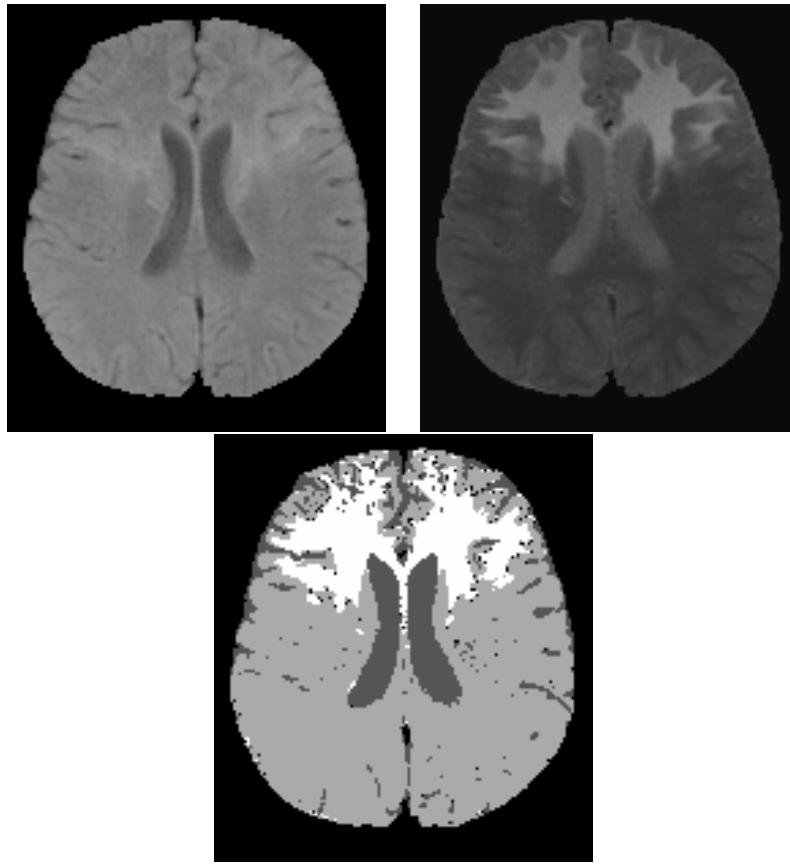


FIG. 8.14 – Deux acquisitions IRM du cerveau et résultat de la classification par fusion floue (la décision n'est prise que localement en chaque point, sans régularisation spatiale)

classification multisource, on cherchera à éviter que des points isolés ou éparpillés dans une classe homogène soient attribués à une classe différente.

8.4.1 Au niveau de la modélisation

L'introduction de l'information spatiale au niveau de la modélisation est plus ou moins implicite suivant le niveau de représentation auquel on se place. Si l'on raisonne au niveau du pixel, l'information contenue dans un pixel ne contient pas d'information spatiale, et celle-ci doit alors être introduite explicitement. Le contexte spatial considéré est le plus souvent le voisinage local de chaque point. Une manière simple de le prendre en compte est de définir la mesure $M_i^j(x)$ (voir chapitre 6) en fonction des caractéristiques de x et également de ses voisins. Si l'on note $\mathcal{V}(x)$ le voisinage de x (contenant x), on définira $M_i^j(x)$ comme une fonction du type :

$$M_i^j(x) = F_i[f_j(y), y \in \mathcal{V}(x)], \quad (8.80)$$

où $f_j(y)$ désigne les caractéristiques de y dans la source j . Ce type d'approche peut être vu comme un problème de filtrage spatial. Dans le cas d'un filtrage linéaire, F s'exprime par une convolution et le noyau de convolution défini sur \mathcal{V} est classiquement une gaussienne ou une fonction porte. Si le filtrage n'est pas linéaire, de nombreuses solutions sont proposées en traitement des images [321], dont les plus classiques sont le filtre médian, le filtre sigma ou

les filtrages morphologiques. Enfin, des techniques plus élaborées procèdent par relaxation, comme les champs de Markov qui agissent soit au niveau de la mesure (on parle alors de restauration), soit au niveau de la classification comme on le verra plus loin.

Si l'on raisonne au niveau de primitives (segments, contours, régions) ou au niveau des objets ou structures de la scène, l'information spatiale locale est implicitement prise en compte dans la représentation. Si la détection de ces éléments n'est pas précise ou si leur localisation ne l'est pas (par exemple à cause de l'imperfection du recalage), il est souvent souhaitable d'introduire cette imprécision spatiale explicitement dans la représentation, avant la fusion. La dilatation floue est une opération bien adaptée à cette fin [57, 61, 72]. Cela permet de réduire le conflit au moment de la fusion, et donc de choisir simplement et sans risque un mode de combinaison conjonctif.

De manière moins locale, les relations spatiales entre primitives constituent une information importante sur la structure de la scène [60, 41, 42, 43, 45] et elles peuvent avantageusement être prises en compte dans la fusion, comme source d'informations supplémentaire [44, 48, 154]. Dans ce cas, le contexte spatial $\mathcal{V}(x)$ d'un élément x est un ensemble de primitives ou d'objets dont on connaît les relations spatiales par rapport à x .

8.4.2 Au niveau de la décision

L'introduction de l'information spatiale au niveau de la décision est la plus facile. La méthode la plus courante consiste à établir dans un premier temps une règle de rejet (en fonction de la netteté et du caractère discriminant de la décision) puis à reclasser les éléments rejetés en fonction de leur contexte spatial. Par exemple, la reclassification peut être faite suivant la règle suivante (majorité absolue) :

$$x \in C_i \text{ si } |\{y \in \mathcal{V}(x), y \in C_i\}| \geq \frac{|\mathcal{V}|}{2} \quad (8.81)$$

qui exprime qu'au moins la moitié des éléments du voisinage doit être dans C_i pour classer x dans C_i . Cette règle ne permet pas toujours d'affecter x à une classe. Une autre règle moins sévère considère juste la classe la plus représentée dans le voisinage (règle majoritaire) :

$$x \in C_i \text{ si } |\{y \in \mathcal{V}(x), y \in C_i\}| = \max_k |\{y \in \mathcal{V}(x), y \in C_k\}|. \quad (8.82)$$

Ces règles s'appliquent quel que soit le niveau de représentation des éléments considérés.

Un exemple en classification floue peut être trouvé dans [67], mais la méthode est générale et peut être appliquée de manière similaire aux autres théories.

8.4.3 Au niveau de la combinaison

L'introduction de l'information spatiale au niveau de la combinaison est plus rare et plus délicate.

En fusion probabiliste, les champs de Markov offrent un cadre naturel pour cela. Dans l'expression de la règle de Bayes, c'est dans la probabilité *a priori* qu'intervient l'hypothèse markovienne. Cette probabilité est combinée aux probabilités conditionnelles aux classes par un produit. Cette remarque nous conduit à considérer que l'information spatiale constitue dans ce modèle une source de données au même titre que les autres.

C'est l'approche la plus courante, et elle a été appliquée à plusieurs niveaux de représentation. Au niveau local, du pixel, de nombreux exemples peuvent être trouvés dans la littérature (par exemple [13, 110]). A un niveau plus structurel, les champs de Markov sont définis sur des graphes plus généraux que les graphes de pixels (les nœuds sont des primitives ou même des objets), et l'on trouve des exemples pour la détection de routes dans des images SAR [335], pour la segmentation d'images IRM du cerveau [156], pour la reconnaissance de structures du cortex cérébral [229, 228], pour l'interprétation d'images aériennes [249], etc.

Dans les autres théories, il serait également possible de développer des approches similaires, toujours en considérant l'information spatiale comme une source de données supplémentaire.

C'est par exemple le cas des relations spatiales mentionnées plus haut considérées comme source supplémentaire d'informations : la reconnaissance d'un objet peut résulter de la fusion d'informations sur cet objet, et d'informations sur les relations qu'il doit avoir par rapport à d'autres objets. Le cadre des ensembles flous permet à la fois la représentation et la fusion de telles informations [48].

Citons un autre exemple : dans [177], une fonction de masse est définie pour représenter le contexte spatial et combinée à des fonctions de masse représentant les informations extraites des images par la règle de Dempster. Toutefois, peu de travaux existent encore dans ce domaine, qui mérite certainement d'être développé.

8.4.4 Exemples d'applications

Au niveau de la combinaison : classification markovienne multisource

Nous reprenons ici l'exemple de la fusion des images IRM du cerveau des parties 8.2 et 8.3, pour segmenter le cerveau sain, la pathologie et les ventricules, cette fois par une approche markovienne. Cette méthode a été développée dans [14, 12, 13].

La règle de Bayes permet de calculer la probabilité *a posteriori* de chaque classe conditionnellement aux deux images. Le terme de probabilité *a priori* est modélisé à l'aide d'une hypothèse markovienne sur l'image des classes et tient lieu de régularisation spatiale. La probabilité *a posteriori* s'exprime donc comme le produit de trois termes : deux termes exprimant les probabilités des niveaux de gris de chacune des deux images conditionnellement aux classes (sous l'hypothèse d'indépendance conditionnelle), et un terme exprimant la régularité spatiale des classes. Le cadre markovien permet d'exprimer le problème d'optimisation du maximum *a posteriori* comme la minimisation d'une énergie comprenant :

1. des termes d'attache aux données, dépendant des niveaux de gris dans chaque image, de coefficients pondérant l'importance de chaque image suivant les classes, et d'une connaissance *a priori* sur la position des ventricules dans le cerveau ;
2. un terme de régularisation, sous la forme d'un potentiel de Potts, prenant en compte les pixels voisins de chaque point.

L'énergie s'écrit donc en chaque point x affecté à la classe C_i dans l'étape courante :

$$\Phi_i^1(f_1(x)) + \Phi_i^2(f_2(x)) + \lambda \sum_{y \in \mathcal{V}(x)} \omega(C_i, C_y) \quad (8.83)$$

où Φ_i^1 (Φ_i^2) représente le potentiel d'attache aux données caractérisant la classe C_i dans la première (seconde) image, fonction du niveau de gris $f_1(x)$ ($f_2(x)$) au point x dans cette

image, $\mathcal{V}(x)$ représente le voisinage spatial de x , C_y la classe à laquelle est affecté le voisin y à l'itération courante, et $\omega(C_i, C_y)$ représente les contraintes de régularisation entre les classes C_i et C_y . Le facteur λ permet de pondérer l'influence de la régularisation par rapport à l'attache aux données. Les potentiels d'attache aux données sont déterminés automatiquement à partir des histogrammes de niveaux de gris, dont les modes significatifs sont sélectionnés par une approche multiéchelle. Ici la régularisation est simple : elle favorise l'appartenance à la même classe de points voisins ($\omega(C_i, C_i) = 0$) et défavorise l'appartenance de points voisins à des classes différentes ($\omega(C_i, C_y) = 1$ si $C_y \neq C_i$).

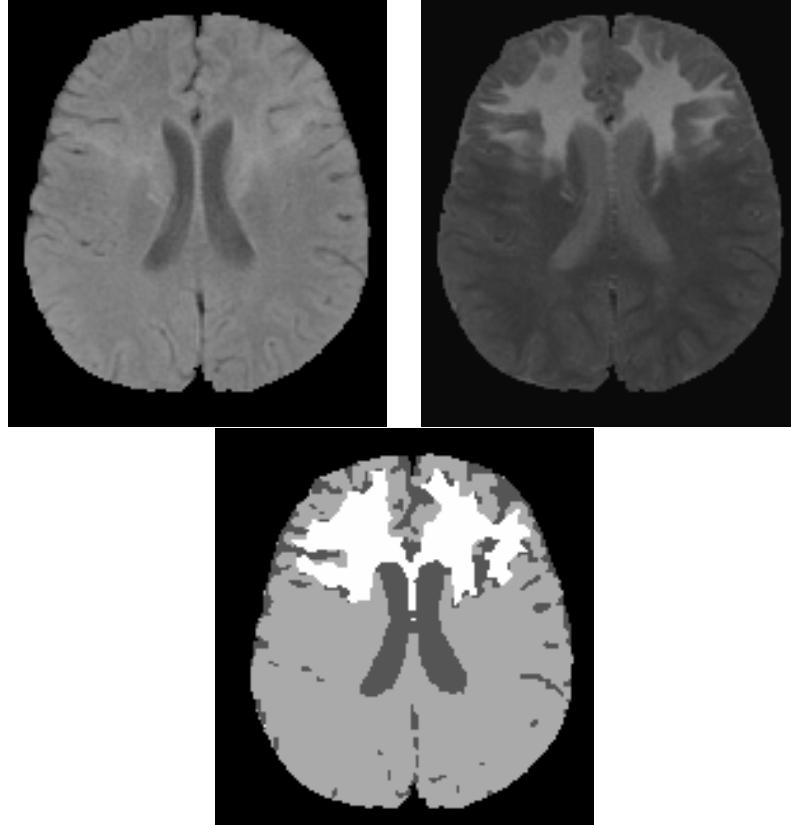


FIG. 8.15 – Deux acquisitions IRM du cerveau et résultat de la classification par fusion markovienne

Les résultats obtenus (voir figure 8.15) montrent l'homogénéité spatiale de la segmentation obtenue. L'information spatiale introduite dans cet exemple reste relativement locale, puisqu'elle ne concerne qu'un petit voisinage autour de chaque point.

Au niveau de la modélisation et de la décision : fusion de détecteurs de structures par la théorie des fonctions de croyance

Dans cet exemple, développé dans [334], l'objectif est d'interpréter une image radar en fusionnant les résultats de plusieurs détecteurs de structures (routes, relief, villes, etc.). La fusion est effectuée à un niveau intermédiaire, sur des primitives obtenues par ces détecteurs (segments ou régions), et donc comportant des informations spatiales.

La figure 8.16 illustre un exemple d'image du radar ERS-1 de la région d'Aix-en-Provence.

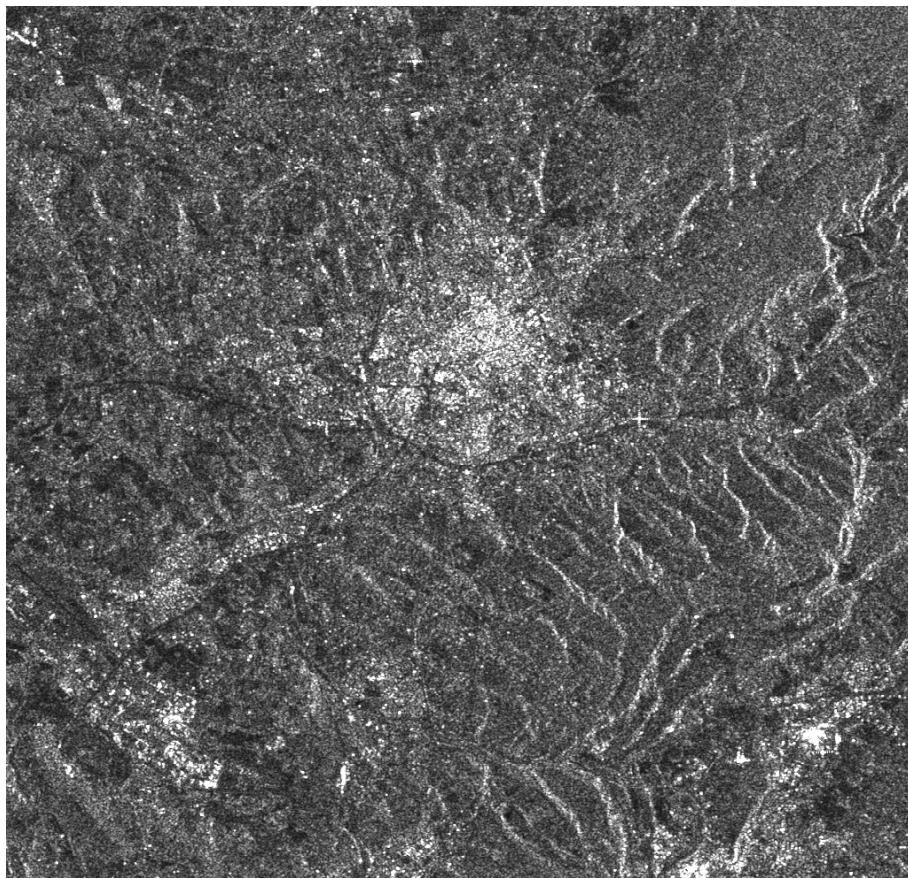


FIG. 8.16 – Image du radar ERS-1 de la région d’Aix-en-Provence

La figure 8.17 montre les résultats de trois des détecteurs appliqués sur cette image.

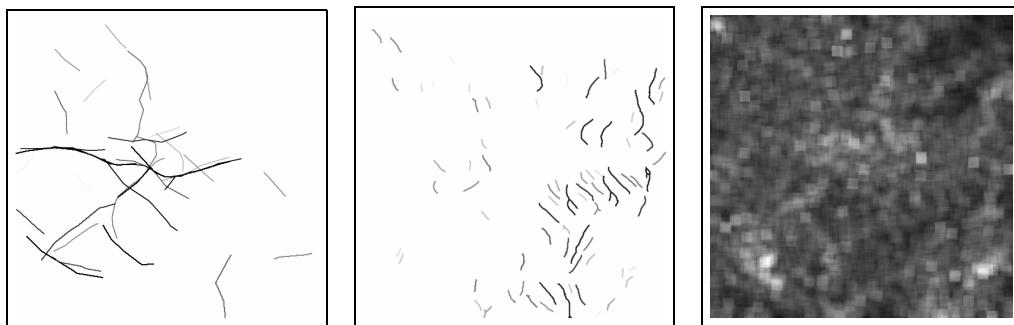


FIG. 8.17 – Résultats de trois détecteurs : routes, relief et zones homogènes. Les niveaux de gris indiquent la confiance fournie par chaque détecteur dans le résultat obtenu

Les résultats de ces détecteurs servent de base à une modélisation par des fonctions de croyance. Cette théorie est particulièrement bien adaptée ici puisqu’elle permet de modéliser le comportement de chaque détecteur. Ainsi, les détecteurs de routes, de relief, ou encore de fleuves sont précis pour les objets auxquels ils sont dédiés, mais ne sont pas capables de distinguer les autres classes. Ainsi, les éléments focaux d’un détecteur de routes sont la classe « routes » et l’ensemble « non routes » (soit la réunion de toutes les autres classes

considérées). Au contraire, le détecteur de zones homogènes n'est pas très précis car il utilise des filtrages sur des fenêtres de taille 9×9 , ce qui fait disparaître les routes traversant des zones homogènes. Il est alors naturel de prendre pour éléments focaux l'ensemble « routes \cup zones homogènes » et l'ensemble « non homogène ». C'est un exemple de détecteur pour lequel aucun élément focal n'est un singleton. Les détecteurs de zones urbaines sont modélisés de manière similaire. On voit dans cet exemple que ce sont les connaissances sur les détecteurs et leurs comportements qui permettent d'effectuer la modélisation et le choix des éléments focaux.

Les fonctions de masse sont ensuite apprises sur les histogrammes des réponses des différents détecteurs, par minimisation d'une distance entre ces histogrammes et des fonctions paramétriques trapézoïdales.

La fusion est ensuite effectuée par la règle orthogonale de Dempster non normalisée (fusion conjonctive) puisque toutes les imprécisions et ambiguïtés sur les détecteurs sont explicitement prises en compte dans la modélisation. Cela permet de réduire les éléments focaux à des singletons ou des réunions de deux classes seulement. De plus, il s'agit typiquement d'une application où l'on est en monde ouvert : il n'est pas possible de prévoir toutes les classes qui peuvent apparaître dans l'image, et seules celles pour lesquelles des détecteurs ont été construits peuvent être détectées. La combinaison non normalisée permet de représenter dans la masse de l'ensemble vide tout ce qui n'est pas prévu.

Enfin, l'étape de décision est effectuée dans un cadre markovien, permettant d'ajouter une cohérence spatiale entre les différentes régions, donc un niveau supplémentaire d'information spatiale. Les probabilités pignistiques (voir partie 8.2) permettent de revenir à des probabilités sur des singletons, qui sont ensuite combinées à un terme de régularisation spatiale.

Un résultat d'interprétation est illustré figure 8.18.

Au niveau de la modélisation : fusion floue de relations spatiales

Dans ce dernier exemple, les informations spatiales considérées sont des informations structurelles, portant non plus sur la cohérence locale des classes ou régions, mais sur les relations entre les objets recherchés. L'application concerne la reconnaissance des structures internes du cerveau dans des images IRM, guidée par un atlas anatomique [48, 155].

Une coupe extraite du volume 3D de l'atlas est montrée figure 8.19 ; la vue de la coupe correspondante dans l'acquisition IRM 3D qui doit être traitée est représentée figure 8.20.

La reconnaissance est effectuée de manière progressive, une structure étant détectée à chaque étape. Chaque étape s'appuie sur les objets obtenus lors des étapes précédentes et sur des connaissances anatomiques de différentes natures. Les informations de localisation et de morphologie de cet objet sont apportées par l'atlas et des informations symboliques sur cet objet sont exprimées relativement à des objets identifiés lors des étapes précédentes. Ces informations symboliques concernent aussi bien des relations spatiales (ensemblistes, directionnelles ou de distances) que des informations de constitution (substance grise, substance blanche ou liquide) ou des connaissances radiométriques relatives au type d'imagerie. Le cadre des ensembles flous a été choisi pour plusieurs raisons : il permet d'exprimer dans un formalisme unique des informations de sémantiques variées, nombreuses pour ce problème ; il permet également de modéliser l'imprécision et l'incertain, ce qui est important dans ce

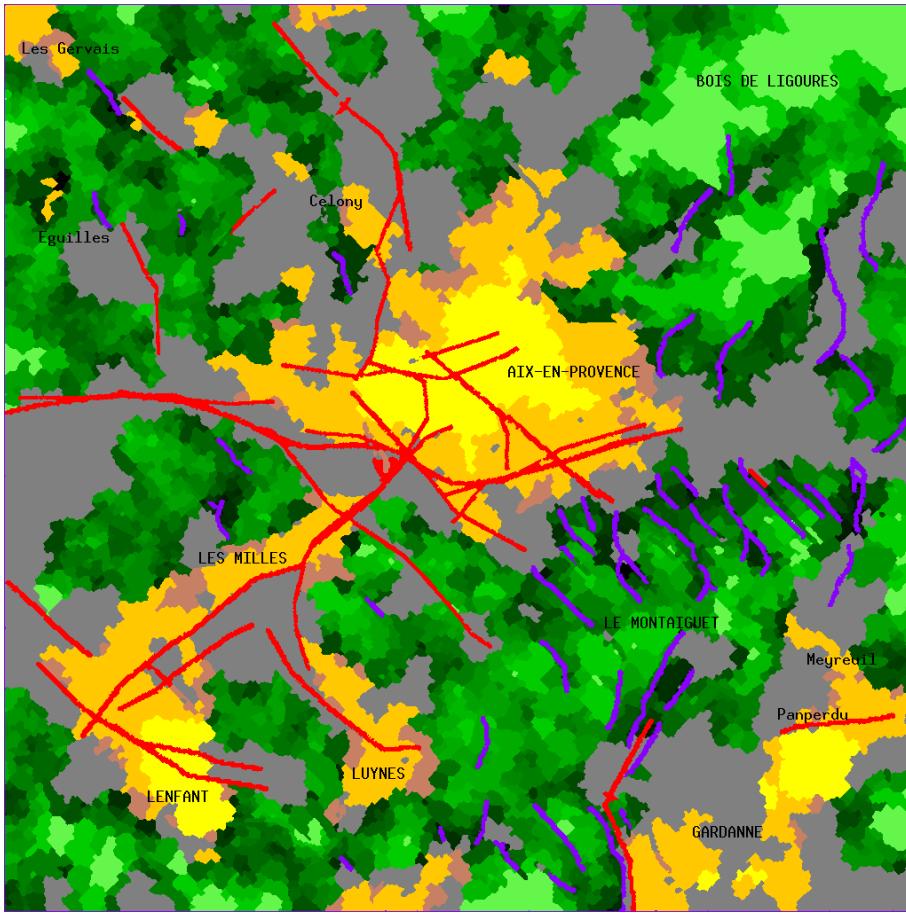


FIG. 8.18 – Résultat obtenu sur l'image de la figure 8.16 par fusion de détecteurs de diverses structures. Les zones les plus claires représentent les zones urbaines. Les traits sombres correspondent aux routes, les traits moins sombres aux zones de relief. Les zones les plus sombres (avec différents degrés) sont les zones de forêt. Enfin, les zones de gris moyen correspondent à l'ensemble vide, donc aux zones non reconnues.

domaine en particulier à cause de la variabilité anatomique interindividuelle ; enfin, la fusion floue qui mène à la reconnaissance peut tirer parti d'un nombre important d'opérateurs qui permettent de bien modéliser les relations que l'on connaît entre ces informations.

La modélisation s'effectue en représentant, dans l'espace de l'image, chaque élément d'information par un ensemble flou. Des zones d'intérêt floues sont ainsi définies pour chaque type de connaissance et leur fusion permet de focaliser la recherche dans une zone de plus en plus restreinte. Nous illustrons cette démarche sur l'étape de reconnaissance du noyau caudé. A cette étape de la procédure de reconnaissance, trois objets anatomiques ont été segmentés : le cerveau et les deux ventricules latéraux. La figure 8.21 illustre la représentation des connaissances relatives au noyau caudé (information de localisation et de forme approximative provenant de l'atlas, localisation dans le cerveau et à l'extérieur des ventricules latéraux, radiométrie et position relative à gauche des ventricules latéraux). La définition formelle de ces représentations peut être trouvée dans [45, 48].

La reconnaissance de certaines structures utilise des informations de distance, qui sont également représentées par des ensembles flous dans l'espace de l'image. La figure 8.22 illustre

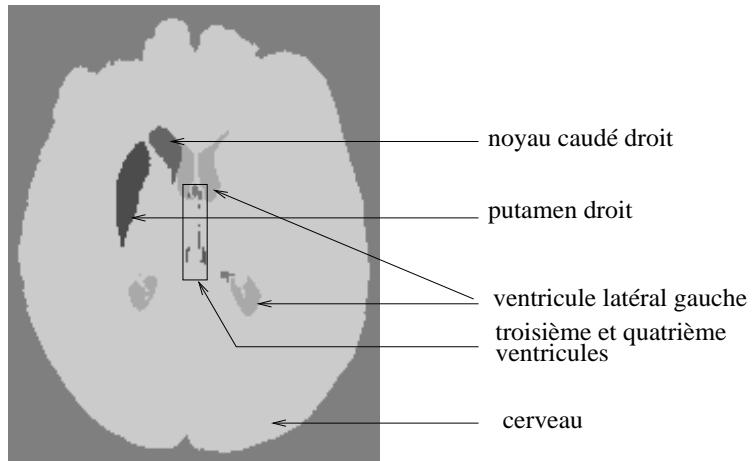


FIG. 8.19 – Une coupe de l'atlas représentant les structures que nous cherchons à reconnaître dans l'image

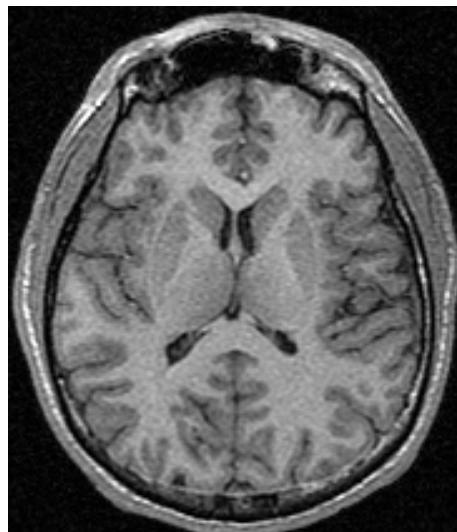


FIG. 8.20 – Une coupe du volume cérébral à reconnaître dans un plan proche de celui de l'atlas de la planche précédente. C'est une image acquise en IRM

ce type de représentation, pour trois types de connaissances.

La fusion des connaissances spatiales est effectuée par un opérateur conjonctif (une t-norme) afin de réduire la zone d'intérêt au fur et à mesure que de nouvelles connaissances sont apportées, de façon à focaliser la recherche. Puis ces connaissances spatiales fusionnées sont combinées aux informations radiométriques par un opérateur de moyenne, réalisant cette fois un compromis entre ces deux types d'informations. La zone obtenue permet alors une segmentation très facile (figure 8.23).

La figure 8.24 illustre six objets de l'atlas ainsi que ces mêmes objets reconnus dans une acquisition IRM. Ils sont correctement segmentés bien que leurs taille, localisation et morphologie dans l'image diffèrent significativement de celles de l'atlas qui sert de modèle. Il est notable que les troisième et quatrième ventricules, particulièrement difficiles à segmenter dans des images IRM, sont ici correctement reconnus et segmentés, grâce à l'utilisation de

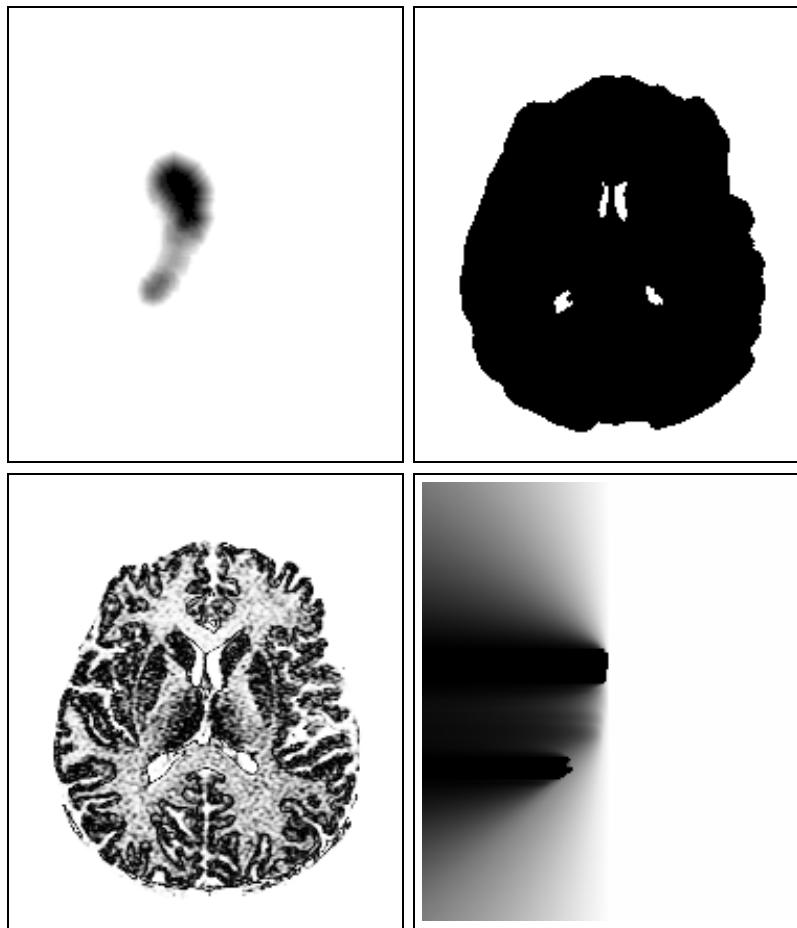


FIG. 8.21 – Informations exprimées dans l'espace de l'image. Le blanc et le noir correspondent respectivement aux valeurs d'appartenance minimale et maximale. En haut à gauche : information de localisation et de forme approximative provenant de l'atlas. En haut à droite : contrainte binaire de localisation exprimant que le noyau caudé se trouve dans le cerveau (zone de recherche, en noir) et à l'extérieur des ventricules latéraux en blanc. En bas à gauche : connaissance radiométrique *a priori*. En bas à droite : relation directionnelle relative « à gauche des ventricules latéraux ».

relations spatiales par rapport à d'autres structures du cerveau.

8.5 Conclusion

Après ce tour d'horizon des principales méthodes numériques de fusion et de leur utilisation en traitement et interprétation des images, nous résumons ici quelques acquis et conclusions, ainsi que quelques points qui posent encore des difficultés.

8.5.1 Quelques acquis

Les développements précédents font état d'une grande variété de techniques numériques utilisées pour la fusion d'informations imprécises et incertaines. Ce foisonnement est dû à la

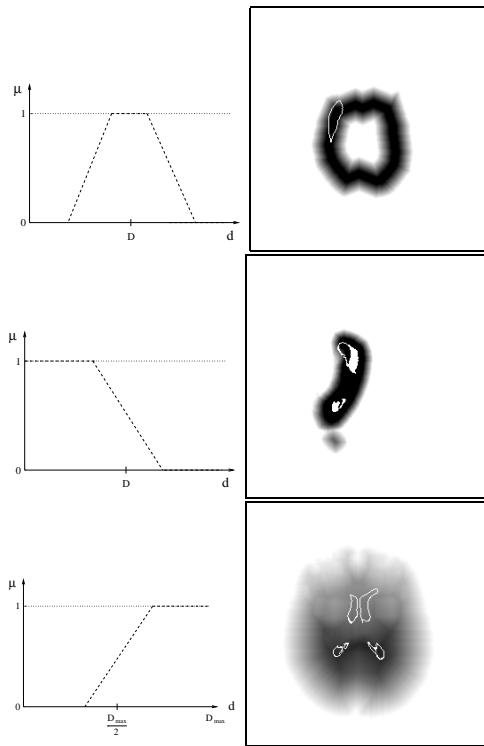


FIG. 8.22 – Exemples de représentation de connaissances sur les distances. A gauche : représentations sous forme d’intervalles flous sur \mathbb{R}^+ . A droite : représentations spatiales correspondantes. Le putamen a une distance approximativement constante à la surface du cerveau (en haut), le noyau caudé est à une distance inférieure à environ D des ventricules latéraux (représentés en blanc) (au milieu), les ventricules latéraux sont dans le cerveau à une distance plus grande que $D_{\max}/2$ environ de la surface du cerveau (en bas). Les contours des objets recherchés (putamen, noyau caudé et ventricules respectivement) sont montrés en blanc, et sont bien dans les zones satisfaisant les relations avec un fort degré.

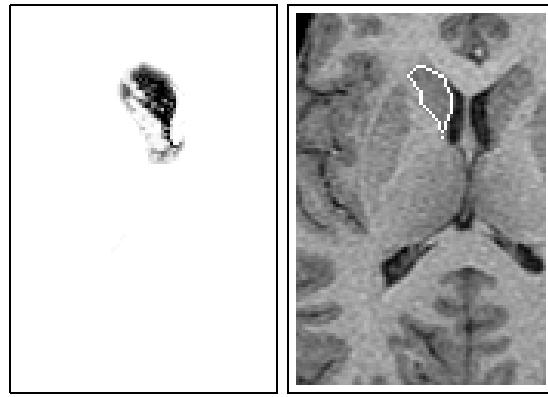


FIG. 8.23 – Résultat de la fusion de toutes les connaissances disponibles sur le noyau caudé (à gauche) et segmentation de ce noyau (à droite)

variété des tâches qui contribuent à la décision dans un système d’informations multisources. Les approches probabilistes restent les plus utilisées, surtout parce qu’elles ont donné lieu au développement d’outils aboutis et d’un grand savoir-faire, résultant d’une longue pratique. Au niveau de la décision, ces outils se révèlent particulièrement efficaces, alors qu’au niveau de la modélisation plusieurs aspects restent limités voire criticables. La théorie des ensembles

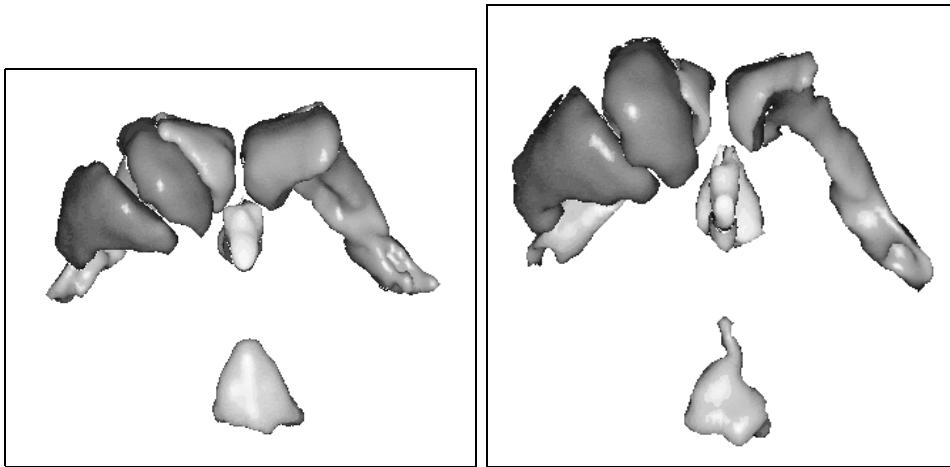


FIG. 8.24 – Résultats de la procédure reconnaissance. Six structures sont représentées : les ventricules latéraux (gris moyen), les troisième et quatrième ventricules (gris clair), un noyau caudé et un putamen (gris foncé). A gauche, les structures de l'atlas, et à droite celles reconnues dans une image IRM 3D.

fous repose sur une modélisation proche de l'intuition. Dans les applications de fusion dans les domaines traités ici, l'étape de décision reste encore peu formalisée et peu développée. En revanche, l'étape de combinaison est très riche et permet d'introduire nombre de connaissances, de tous types. La théorie des fonctions de croyance offre les outils de modélisation les plus puissants, permettant simplement et efficacement l'introduction de connaissances, de l'imprécision et de l'incertitude. La combinaison, telle qu'elle est utilisée en traitement des images, est restreinte au mode conjonctif.

En fusion numérique pour le traitement d'images, les efforts des dernières années ont permis d'aboutir à une meilleure compréhension des différentes théories issues d'autres domaines telles que les fonctions de croyance et le flou. On sait ainsi maintenant quels sont les bons cadres d'application de ces théories, leurs atouts et limites dans ces domaines d'application, on sait également représenter et modéliser l'information et les données numériques, symboliques ou structurelles dans chacun des formalismes, et effectuer leur combinaison. De nombreux développements ont vu le jour, en particulier pour les applications typiques de classification multisource, de reconnaissance de structures ou d'objets dans les images, de suivi, de localisation et de planification.

8.5.2 Quelques perspectives

Malgré ces avancées, certains points restent obscurs ou nécessitent plus de développement.

La prise en compte de l'origine des données et des connaissances, ainsi que des relations entre les sources est encore souvent effectuée de manière supervisée et elle nécessite donc un peu d'expérience. Une des questions importantes concerne l'indépendance entre les sources et le conditionnement (en particulier dans le cas de la fusion séquentielle, dans des processus dynamiques de mise à jour). Le cadre probabiliste propose des méthodes pour tester l'indépendance statistique, et ce sont à peu près les seuls outils dont on dispose. Mais ce type d'indépendance est souvent considéré comme trop contraignant et dans d'autres cadres, tels que celui de la théorie des fonctions de croyance, on préfère parler d'indépendance cognitive,

plus reliée à la manière dont les connaissances et les données sont acquises qu'à leur nature [302, 314]. Dans le choix des opérateurs des théories floues et possibilistes, l'indépendance se traduit par des propriétés d'idempotence des opérateurs, alors que la dépendance nécessite le renforcement.

Une question très difficile est celle de la gestion du conflit. Dans la mesure du possible, on cherchera à identifier les sources du conflit et à les expliciter pour éviter les incohérences au moment de la décision. En particulier il n'est pas toujours facile de différencier le conflit de la complémentarité des sources, ni de savoir s'il doit être résolu ou non. Le conflit, que l'on peut appeler « apparent », est en réalité une forme de complémentarité. Par exemple, si une source inclut systématiquement la classe B dans la classe A alors qu'une autre source permet de bien les distinguer, ces deux sources apparaissent en conflit. La reconnaissance de la complémentarité passe souvent par des connaissances *a priori* (ou apprises) sur les faibles possibilités de la première source à distinguer ces deux classes. La résolution de tels conflits est facile lorsqu'ils ont été bien identifiés. Une deuxième forme de conflit, réel cette fois, vient d'incohérences entre les sources, dues à leur fiabilité limitée, aux changements survenus dans la scène entre les acquisitions, ou encore au fait qu'elles ne parlent pas de la même chose. Ce type de conflit est plus difficile à identifier et à résoudre. Il vaut même mieux parfois se limiter à le signaler et ne pas tenter de le résoudre car il correspond bien souvent à une insuffisance fondamentale de nos connaissances sur le problème.

Le choix des méthodes et leur évaluation est aussi crucial que difficile. Là encore, il n'existe pas de solution générale pour le choix des méthodes adaptées aux types d'informations et de connaissances manipulées, ainsi qu'aux applications visées. L'évaluation des méthodes peut être plus ou moins facile suivant que l'on a accès à la vérité ou non. Les essais de comparaison des approches numériques de fusion, appliquées aux mêmes problèmes, ont souvent donné des résultats contradictoires, et ont donc échoué. Nous pensons que la raison essentielle est que chaque problème s'exprime plus facilement dans une théorie que dans une autre, et que sa résolution par des outils non adaptés nécessite donc des distorsions de ces techniques et n'a pas beaucoup de sens.

Enfin, dans le cas des applications en traitement d'images, et également pour certaines applications en robotique, l'introduction d'informations spatiales dans la fusion est un point important, pour lequel l'ensemble des méthodes existantes pourrait être encore étouffé. En particulier, les succès récents de la prise en compte d'informations structurelles montrent l'intérêt de la combinaison d'informations spatiales de différents niveaux.

Nous avons constaté dans les chapitres précédents que chacune des approches est adaptée à un ensemble limité d'imperfections des informations à fusionner. Il est rare que toutes les imperfections soient modélisées simultanément et simplement dans une théorie unique. Ainsi, un champ d'investigations encore ouvert concerne la fusion de méthodes ou l'utilisation conjointe de différents formalismes complémentaires. De tels travaux sur la combinaison de méthodes sont prometteurs puisqu'ils visent à exploiter les avantages des différentes théories pour les faire coopérer. Cette combinaison peut s'appuyer sur les liens qui existent entre les différentes approches. Par exemple, une probabilité peut être interprétée comme une fonction de masse particulière, une fonction de croyance dont les éléments focaux sont emboîtés peut s'interpréter comme une distribution de possibilités, une distribution de possibilités peut être interprétée comme des intervalles de confiance ou comme une famille de probabilités (voir par exemple [133] pour le détail des liens entre possibilités et probabilités), etc. Ainsi, des travaux ont déjà été effectués pour combiner l'imprécision représentée par des ensembles flous

à l'incertitude probabiliste (par exemple [74, 270, 291] dans une approche de classification markovienne où les classes sont floues), pour combiner les champs de Markov et les fonctions de croyance [21, 177], ou pour raisonner avec des fonctions de croyance dont les éléments focaux sont flous [313, 348, 355, 362].

Chapitre 9

Les systèmes à base de connaissance en traitement des images

L'évolution du traitement des images depuis le bas niveau de traitement du signal jusqu'à l'interprétation des scènes complexes conduit tout naturellement l'utilisateur à prendre de plus en plus en compte des connaissances extérieures au simple signal d'image. Cette évolution amène donc spontanément les techniques de gestion de ces connaissances en contact du traitement des images [247, 330]. Ce sont les méthodes développées en intelligence artificielle et qui se sont appelées successivement : systèmes experts, KBS (*Knowledge based systems*), systèmes multi-agents, etc.

Ces techniques ont eu une forte influence sur le développement des techniques de traitement des images. Elles ont en particulier alimenté de très importants projets dans le cadre des programmes européens. Nous nous proposons de voir ici quels sont les enseignements que l'on peut tirer aujourd'hui de ces approches. Une littérature abondante est présentée plus loin, mais quelques revues permettent d'aborder le domaine [94, 117].

Les objectifs de ces méthodes sont les suivants :

1. représenter les connaissances de façon déclarative et non procédurale comme dans les algorithmes les plus classiques du traitement d'images. Exemple de connaissance déclarative : « Le riz pousse en région chaude et humide. ». Exemple de déclaration procédurale :
 - if $signal(i, j) < seuil$ then $\{i, j\} \in Classe_{route}$ endif
2. séparer les connaissances en divers types :
 - les connaissances factuelles séparées des connaissances opératoires. Exemple de connaissance factuelle : « Les nuages d'altitude et la neige ont la même température ». Exemple de connaissance opératoire : « En imagerie Radar, l'eau se détecte comme une zone de très faible signal ». Il est parfois délicat de distinguer l'une et l'autre connaissance et c'est très heureux car alors le passage de la connaissance à la mise en œuvre est très simple. Mais une donnée peut-être opératoire dans un domaine (contexte limité, champ étroit d'application) tandis qu'elle ne sera que factuelle ailleurs. (Par exemple si l'on fait de la surveillance urbaine, la connaissance « Un chien est un quadrupède » est très rapidement opératoire pour reconnaître les chiens,

tandis qu'elle n'est que factuelle s'il s'agit de faire de l'analyse de scène dans une ferme ou un zoo).

- les connaissances particulières séparées des connaissances générales. Exemple de connaissance générale : « Les contours sont portés par les hautes fréquences du spectre ». Exemple de connaissance particulière : « Les champs de blé se distinguent des champs de maïs par leur texture en imagerie SPOT-5 mais pas en SPOT-1 ». Là encore, il faut souvent examiner le domaine d'intérêt pour décider si une information est particulière ou générale.
3. utiliser les mêmes connaissances pour servir des objectifs différents. ainsi. Dans la référence [214], où l'on se propose de faire de l'expertise de textes anciens, l'assertion : « L'usage de la plume d'oie crée des jambages plus fins que l'usage du calame¹ », peut être utilisée pour discriminer des styles dans un objectif de datation d'un document inconnu (parce que cette assertion sera associée à une autre qui précise que telle période a utilisé tel instrument), mais aussi pour authentifier l'instrument d'un texte connu mais d'une période où les deux instruments étaient en concurrence.
 4. créer des mécanismes de déduction universels, donc réutilisables. Ainsi, la détection des contours en imagerie médicale pourra servir pour des organes différents (donc dans d'autres services d'un hôpital), voire en contrôle de qualité en imagerie industrielle.
 5. garder des traces (justificatifs) des déductions obtenues ; il s'agit alors de « justifier » le raisonnement appliqué, afin, soit de le corriger dans une étape de construction du système où des règles nouvelles sont nécessaires à la cohérence, soit de mettre en place des techniques automatiques de gestion de conflit si le système abouti par des voies différentes à des conclusions incompatibles.

9.1 Un système à base de connaissance en bref

9.1.1 Schéma d'ensemble

Le schéma le plus souvent utilisé est représenté sur la figure 9.1. On y trouve les principales fonctionnalités :

1. un système d'interfaçage avec l'utilisateur qui permet de fournir les données à traiter pour un problème donné, spécifier les règles particulières, fixer les objectifs d'une session ;
2. le système de supervision chargé de la gestion globale du système. C'est lui qui fixe en particulier la stratégie qui sera adoptée en fonction du problème à traiter, qui résoud des conflits s'ils se présentent et qui évalue l'adéquation de la solution apportée et de l'objectif proposé pour arrêter les traitements. Le système de supervision utilise des « métarègles », c'est-à-dire des règles sur les règles, indiquant comment enchaîner celles-ci pour assurer l'objectif.
3. un moteur d'inférence qui active les règles de traitement, selon un ordre qui dépend de la stratégie adoptée. Ce moteur d'inférence peut s'appeler différemment selon les contextes, système expert, système multi-agent, communauté d'agents en interaction, etc.

¹le calame est un bambou taillé.

4. et bien sûr une base de connaissance. Elle contient deux familles différentes de connaissances :

- les faits, d'une part, (pour le traiteur d'images on trouvera ici les images, mais aussi les éléments particuliers du contexte : date et heure de la prise de vue, nom, age et antécédents du patient dans un suivi médical, etc.),
- les règles, d'autre part, qui indiquent comment ces faits doivent être combinés pour déduire de nouvelles connaissances.

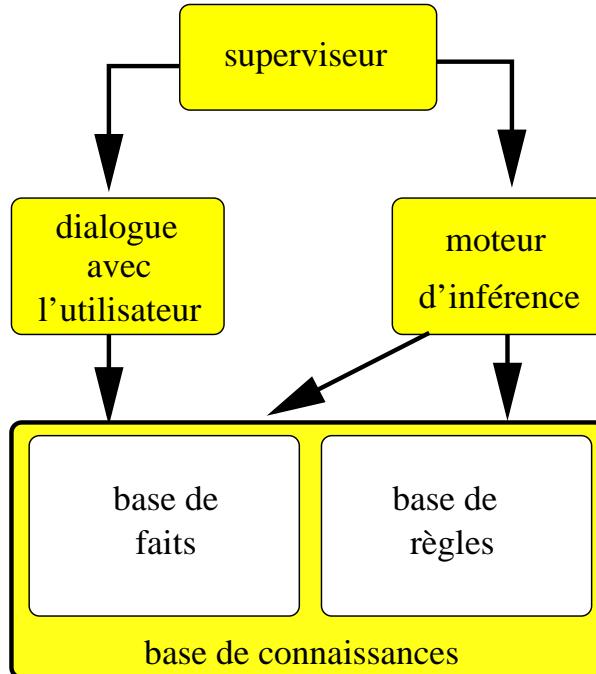


FIG. 9.1 – Les grandes fonctionnalités d'un système à base de connaissances.

Le modèle de système expert le plus utilisé en traitement d'images reprend le schéma du « tableau noir » (*black-board* [174]). Le tableau noir est une zone partagée par tous les opérateurs du système pour lire l'information à traiter et déposer l'information élaborée au fur et à mesure du fonctionnement du système. L'information affichée sur le tableau noir est disponible pour toute opération ultérieure. Le tableau noir d'un système expert de traitement d'images est souvent l'image elle-même, ou du moins une extension de l'image en de multiples couches, toutes référencées sur l'image, et chacune portant les résultats d'une famille d'opérateurs ou d'experts appliqués à une zone précise de l'image (cf. fig. 9.2). Cette assimilation simplifie considérablement la représentation des connaissances dans les SBC de traitement d'images.

9.1.2 Les ingrédients des systèmes à base de connaissances

De leur origine en intelligence artificielle, les systèmes à base de connaissance en traitement des images ont hérité un outillage très classique. Les entités manipulées sont par exemple des *frames* ou des prototypes, entités constituées des composantes élémentaires de l'image, associées à des attributs et liées par des relations.

$$\text{prototype} = \{\text{composantes} + \text{attributs} + \text{relations}\}$$

Les composantes sont assez généralement :

- des points isolés de l'image,
- des contours,
- des régions.

Les relations sémantiques sont elles aussi issues de l'IA : "Part-Of", "Is-a", "Instance-of" sont les plus classiques. Il est beaucoup plus compliqué de mettre en œuvre des structures plus complexes pourtant fort utiles en traitement d'images comme les notions de "partie cachée", d'"ombre" de "représentation multi-échelle", de "composante fréquentielle".

La partie la plus noble du système à base de connaissance est l'étape d'inférence, c'est à dire de création de faits nouveaux à partir de ceux que l'on connaît déjà. Les techniques de raisonnement étaient initialement issues de la logique des prédictats (*If ... then ...*). mais ces outils qui n'admettent que des implications logiques rigoureuses sont trop limitées en traitement des images. La nécessité de faire appel à des techniques plus complètes de gestion de l'incertain, prenant en compte l'incertitude et l'imprécision des données, ainsi que les exceptions dans les règles de déduction ont poussé à l'emploi de techniques plus élaborées [230]. Les règles de probabilité sont naturellement les plus utilisées, et l'inférence bayésienne avec sa rigueur deductive est un excellent candidat pour le maintien de la cohérence. Les probabilités ont conduit à des extensions fort utiles comme les réseaux bayésiens [204]. Malheureusement c'est un candidat exigeant (il faut connaître les probabilités jointes ou conditionnelles, ce qui n'est généralement pas le cas) ; c'est également un candidat limité car il gère mal la dualité incertitude/ignorance. Tout naturellement des techniques différentes de gestion de l'incertain ont été utilisées dans les SBC : ensembles flous, puis théorie des croyances, mais aussi gestion *ad hoc* comme dans [323]. Ces techniques sont largement expliquées dans les premiers chapitres de ce cours.

Néanmoins toutes ces techniques sont avant tout des méthodes de « déduction » et s'opposent ainsi au raisonnement humain qui utilise spontanément des techniques « d'induction » qui permettent véritablement d'innover dans la déduction. Un gros effort restent à faire pour doter les systèmes experts de techniques qui innovent véritablement.

Enfin le contrôle du système est aussi un point très discriminant des divers systèmes à base de connaissance. Les systèmes les plus récents proposent des techniques de contrôle réparti, c'est à dire où les divers experts disposent d'un certain libre-arbitre et ne partagent pas un but commun. C'est de la conjugaison des objectifs propres de chaque expert que la solution finale résulte, et son adéquation à notre intention découle d'un choix judicieusement fait des intérêts des experts. Ainsi ces systèmes multi-agents que l'on dit «en interaction» parce qu'ils peuvent souvent collaborer (par exemple un suiveur de ligne exploitera le résultat d'un détecteur de contours) mais aussi travailler en compétition (par exemple on tentera une segmentation par zone et une segmentation par contours et on gardera le meilleur des deux résultats) [150]. Dans les solutions plus classiques, un système de supervision enchaîne les tâches des experts en fonction des objectifs et des résultats déjà obtenus. On distingue alors des grandes familles de systèmes qui se partagent en :

- les systèmes guidés par les données, qui assemblent par exemple des morceaux de contours pour former des lignes, puis des lignes pour former des objets, puis etc. jusqu'à retrouver un objet connu. Clairement un tel système est aisément à programmer mais laisse assez incertain sur l'objet de ses découvertes.
- les systèmes guidés par les buts, qui partent de l'hypothèse du résultat et recherchent récursivement les étapes antérieures nécessaires à la présence de cette hypothèse.

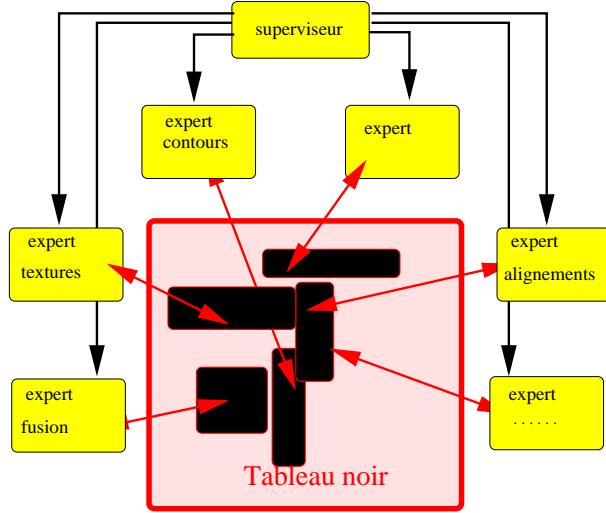


FIG. 9.2 – Le tableau noir est le lieu où sont affichés les résultats du moteur obtenus par la mise en œuvre des règles. Ces résultats sont à la disposition des autres experts pour des utilisations ultérieures. Ici le système expert est figuré comme un système multi-experts, chacun travaillant en parallèle des autres.

Ces derniers systèmes sont bien adaptés à résoudre des tâches précises (détecter un objet précis dans une scène, détecter une pathologie donnée en imagerie médicale), ils sont moins bien adaptés à répondre à des tâches ouvertes (détecter des incidents sur un quai de métro, construire une analyse de scène). Expérimentalement les systèmes comportent souvent des étapes guidées par les données et des étapes guidées par les buts.

9.2 La structure d'un SBC pour la vision

Le prototype d'un tel système est bien représenté par le système VISION de Hanson et Rieseman [173]. Un schéma assez semblable est repris par de nombreux autres systèmes.

Le processus de traitement est séparé en niveaux (cf. figure 9.3) : classiquement un haut niveau (niveau 3) qui contient les descriptions symboliques et les objets reconnus, puis un niveau intermédiaire (niveau 2) qui contient des régions, des lignes, des formes, enfin le bas niveau (niveau 1) qui contient les images et leurs pixels. Des opérateurs utilisent des données du niveau $n - 1$ pour créer l'information du niveau n . En sens inverse des requêtes sont faites par le niveau n au niveau $n - 1$ pour accomplir un certain nombre de tâches nécessaires à sa prise de décision. En permanence, des processus fonctionnent également au sein de chaque niveau pour organiser et ordonner l'information.

9.2.1 Réseaux sémantiques

L'un des outils les plus employés pour raisonner dans l'univers du traitement des images est les réseaux sémantiques qui dressent des structures relationnelles entre les divers types d'objets que l'on rencontre dans la scène et entre leurs constituants.

Un réseau sémantique est constitué d'un ensemble d'objets (*frames* ou prototypes) qui

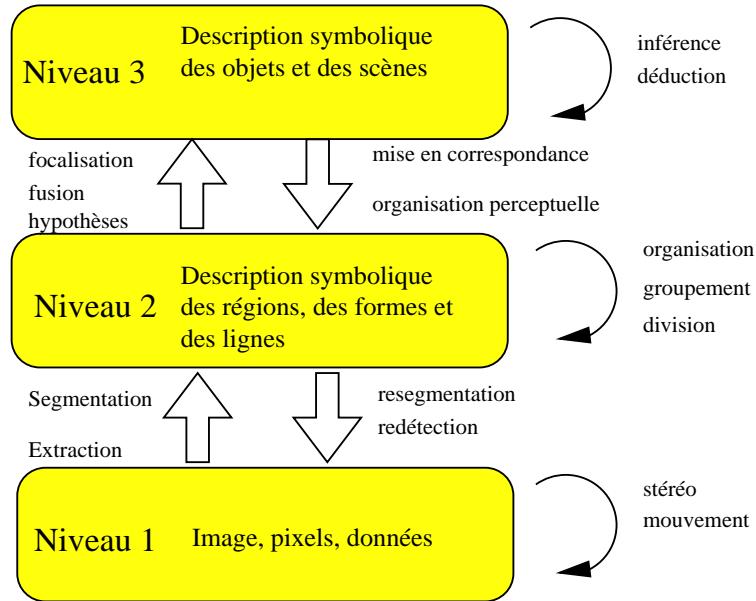


FIG. 9.3 – Schéma de principe d'un système à base de connaissances. Cette représentation est inspirée de VISION.

portent la sémantique de la scène (voir figure 9.4). Chaque objet a des entrées (généralement nourries par d'autres objets) il associe ces entrées au moyen d'un opérateur plus ou moins élaboré (cela va de la formule logique la plus simple à des opérations complexes comme peuvent le faire les agents). Le résultat de cette opération est ensuite transmis pour les autres objets qui s'en servent autant que de besoin.

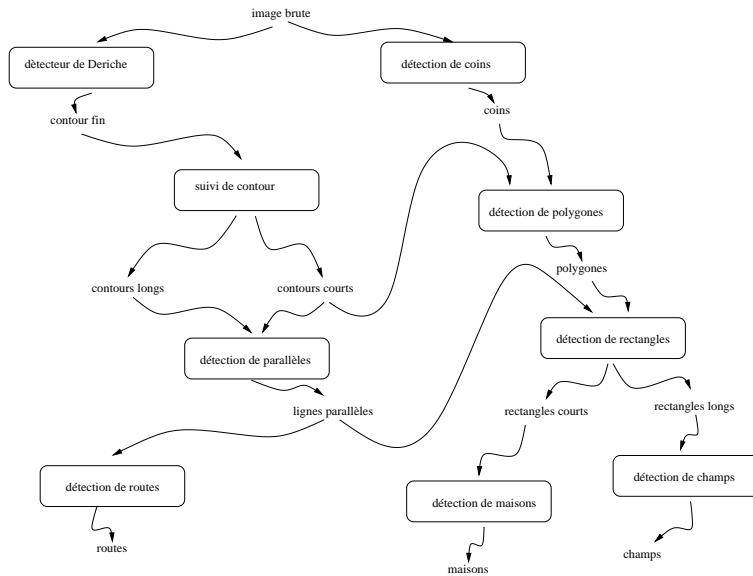


FIG. 9.4 – Exemple de réseau sémantique pour l'analyse d'images aériennes. Les objets élémentaires sont des opérateurs classiques de traitement des images, les "prototypes" sont des éléments plus ou moins aboutis de l'interprétation de la scène.

Lorsque l'on est dirigé par les buts, on recherche les chemins permettant de passer de

l'information disponible (initialement l'image brute) aux buts (par exemple détecter des maisons). Une évaluation des divers chemins permet de proposer la stratégie du système expert (c'est-à-dire l'enchaînement des règles à appliquer), ainsi que des solutions de rechange en cas d'échec.

Lorsqu'on est poussé par les données, le réseau sémantique permet de maintenir l'arborescence des hypothèse et d'éviter les bouclages ou les parcours inutiles.

9.2.2 Réseaux bayésiens

Les réseaux bayésiens sont une structure enrichie des réseaux sémantiques qui remplacent les relations logiques entre objets par des relations probabilistes qui permettront de les gérer au sein d'un système de déduction bayésienne [187, 75]. Dans un réseau bayésien, l'entrée E d'un objet, de valeur e est reliée à la sortie S de valeur s par une probabilité (probabilité que la sortie S prenne la valeur s sachant que l'entrée E a la valeur e) :

$$P(S = s | E = e)$$

Les réseaux bayésiens permettent de calculer les probabilités des diverses objets, aussi bien de façon ascendante (à partir des prémisses, en appliquant les probabilités ci-dessus) que de façon descendante (à partir des conclusions) en utilisant les probabilités inverses déduites de la règle de Bayès :

$$P(E = e | S = s)$$

Au cours du fonctionnement du système expert, chaque nouvelle déduction issue de l'application d'un opérateur a donc pour conséquence de remettre à jour l'ensemble des confiances que nous avons dans les autres prototypes déjà détectés.

Le fonctionnement des réseaux sémantiques posent plusieurs problèmes délicats :

- tout d'abord il convient de fournir au système les probabilités jointes, ceci est une tâche très lourde et délicate car on peut aisément biaiser les probabilités par des choix erronés.
- d'autre part un réseau sémantique a un comportement naturellement combinatorial comme nous venons de le voir, ce qui entraîne des couts inacceptables en temps de calcul [90].

9.3 Bibliographie thématique commentée

9.3.1 Les grandes familles de systèmes experts pour le traitement d'images

On peut distinguer 3 grandes classes de systèmes à base de connaissance en traitement d'image, en fonction du type d'expertise qu'ils convoient :

1. les experts d'un domaine technique : Ils disposent d'une bibliothèque de fonctions de traitement d'images (détection de contours, seuillage, classification, etc.) qui est mise au service d'une problématique thématique spécifique : l'imagerie sismique (Pitas), l'analyse des radiographies pulmonaires (Angy [323]), l'imagerie cardiaque [323, 258, 274, 6, 343], l'analyse des galaxies, (Système OCAPI [331]) la reconnaissance des

textes hébreux anciens [213, 212, 214], la surveillance des forêts (Goldberg), les dessins industriels ([188] ou Redraw [8] ou Celestin [342] ou Mirabelle [232]). La base de connaissances porte essentiellement sur le domaine thématique et a été recueilli auprès d'experts (médecins [66], géologues, linguistes, etc.). Les enchaînements d'algorithmes de traitement des images sont assez figés en fonction des scénarios envisagés, mais les connaissances spécialisées sont très poussées et conduisent à des conclusions très proches d'un diagnostic et généralement au delà des compétences d'un traiteur d'images.

2. Les systèmes d'interprétation de scène : Ils sont spécialisés dans la perception de notre environnement, c'est à dire dans des tâches qui allient vision et description de scène. Ils disposent donc d'expertise en géométrie tridimensionnelle, géométrie projective, description des objets, en particulier pour les univers construits par l'homme, formés d'objets polyèdriques (vision proche) ou les paysages (imagerie aérienne et satellitaire) fusion de capteurs, recalage : système Vision de Hanson [174, 173], système Messie de Giraudon [151, 85, 215, 182, 181], système SPAM de McKeown [239, 238], système Sigma de Matsuyama [236], système ICARE de Desachy [108, 364]. La thématique n'est pas très précise de ces systèmes. Les experts d'imagerie aérienne [319, 210, 292] peuvent par exemple être utilisés pour la cartographie [170, 295] ou pour la surveillance [82] ou pour la planification de trajectoire. Les connaissances thématiques sont réduites, mais les connaissances expertes sont celles d'un observateur expérimenté, doué d'une solide expérience de l'observation du type d'image particulier considéré. Il y a également une bonne base de traitement des images avec descomposantes spécifiques pour les images multispectrales, pour le Radar et l'imagerie IR.
3. Les systèmes experts de traitement des images : Ces systèmes là sont experts des opérations de base de traitement d'image [290, 87] : ils connaissent très précisément comment choisir les seuils de décision en fonction du bruit dans l'image et comment évaluer ce bruit, comment enchaîner les algorithmes de base, comment sortir des situations d'échec pour des tâches élémentaires comme la segmentation des images [148] ou la reconnaissance de textures. Ce sont par exemple les systèmes de Lévine et Nazif (Segmentation par contours et régions [254, 255]), de Lévy-Véhel : Arthur (classification des textures [242]), LLVE (segmentation, [236]).

9.3.2 Articles généraux sur la gestion des connaissances et leur représentation

Représentation des connaissances et raisonnement : [5, 59, 79, 83, 102, 99, 140, 139, 173, 176, 210, 221, 264, 290, 331, 175, 318, 267, 65, 257, 296, 157].

Raisonnement spatial : [7, 69, 91, 102, 364, 40, 39, 58, 179, 326].

Systèmes multi- agents : [20, 68, 63, 62, 103, 331, 142, 251, 319].

Apprentissage : [76, 237, 331, 73].

9.3.3 Applications thématiques

Imagerie aérienne et satellitale : [76, 151, 181, 182, 210, 215, 236, 238, 239, 86, 85, 108, 143, 293, 170].

Imagerie médicale, biologie : [70, 307, 258, 323, 343, 6, 66].

Reconnaissance des caractères, traitement des documents : [232, 213, 259, 212, 342, 188, 8].

Vision pour l'industrie, vision en robotique : [20, 69, 221, 331, 251, 222, 157, 292].

Traitement des images : [254, 255, 148, 292]

Bibliographie

- [1] *Data Fusion Lexicon, Data Fusion Subpanel of the Joint Directors of Laboratories Technical Panel for C³*. F. E. White, Code 4202, NOSC, San Diego, CA, 1991.
- [2] J. Aczél. Sur les opérations définies pour nombres réels. *Bull. Soc. Math. Franç.*, 76 :59–64, 1948.
- [3] J. Aczél. *Lectures on Functional Equations and Their Applications*. Academic Press, New York, 1966.
- [4] C. E. Alchourrón, P. Gärdenfors, and D. Makinson. On the Logic of Theory Change : Partial Meet Contraction and Revision Functions. *Journal of Symbolic Logic*, 50 :510–530, 1985.
- [5] S. Amarel. On representation of problems of reasoning about actions. *Machine Intelligence*, 3 :131–170, 1968.
- [6] M. Anakok, A. Aurengo, H. Tavernier, and L. Marechal. Un système expert pour l’analyse des angiographies numérisées. In *Congrès Cesta Mari 87*, pages 315–319, Paris, 1987.
- [7] K .M. Andress and A. C. Kak. Evidence Accumulation and Flow Control in a Hierarchical Spatial Reasoning System. *AI Magazine*, pages 75–94, 1988.
- [8] D. Antoine, S. Collin, and K. Tombre. Analysis of technical documents, the redraw system. In *Structured documents Analysis*. Baird, Bunke and Yamamoto, 385-402 edition, 1992.
- [9] A. Appriou. Probabilités et incertitude en fusion de données multi-senseurs. *Revue Scientifique et Technique de la Défense*, (11) :27–40, 1991.
- [10] A. Appriou. Formulation et traitement de l’incertain en analyse multi-senseurs. In *Quatorzième Colloque GRETSI*, pages 951–954, Juan les Pins, 1993.
- [11] 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.
- [12] L. Aurdal. *Analysis of Multi-Image Magnetic Resonance Acquisitions for Segmentation and Quantification of Cerebral Pathologies*. PhD thesis, École Nationale Supérieure des Télécommunications, ENST, mar 1997.
- [13] L. Aurdal, I. Bloch, H. Maître, C. Graffigne, and C. Adamsbaum. Continuous Label Bayesian Segmentation, Applications to Medical Brain Images. In *ICIP’97*, volume II, pages 128–131, Santa Barbara, CA, October 1997.
- [14] L. Aurdal, X. Descombes, H. Maître, I. Bloch, C. Adamsbaum, and G. Kalifa. Fully Automated Analysis of Adrenoleukodystrophy from Dual Echo MR Images : Automatic

- Segmentation and Quantification. In *Computer Assisted Radiology CAR'95*, pages 35–40, Berlin, Germany, June 1995.
- [15] J. F. Baldwin. Inference for Information Systems Containing Probabilistic and Fuzzy Uncertainties. In L. Zadeh and J. Kacprzyk, editors, *Fuzzy Logic and the Management of Uncertainty*, pages 353–375. J. Wiley, New York, 1992.
 - [16] H. Bandemer and W. Näther. *Fuzzy Data Analysis*. Theory and Decision Library, Serie B : Mathematical and Statistical Methods. Kluwer Academic Publisher, Dordrecht, 1992.
 - [17] G. Banon. Distinction entre plusieurs sous-ensembles de mesures floues. In *Colloque International sur la Théorie des Ensembles Flous*, Marseille, France, 1978.
 - [18] Y. Bar-Shalom and T. E. Fortmann. *Tracking and Data Association*. Academic Press, San Diego, 1988.
 - [19] J. A. Barnett. Computational Methods for a Mathematical Theory of Evidence. In *Proc. of 7th IJCAI*, pages 868–875, Vancouver, 1981.
 - [20] O. Baujard. *Un environnement de programmation multi-agents et son application à la vision pr ordinateur*. PhD thesis, Université J. Fourier, Grenoble, 1992.
 - [21] A. Bendjebour and W. Pieczynski. Segmentation d’images multisensor par fusion évidentielle dans un contexte markovien. *Traitemet du Signal*, 14(5) :453–464, 1997.
 - [22] E. Benoit and L. Fouloy. Capteurs flous multicompasantes : applications à la reconnaissance des couleurs. In *Les Applications des Ensembles Flous*, pages 167–176, Nîmes, France, October 1993.
 - [23] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum, New-York, 1981.
 - [24] 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.
 - [25] J. C. Bezdek and S. K. Pal. *Fuzzy Models for Pattern Recognition*. IEEE Press, New-York, 1992.
 - [26] 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.
 - [27] 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.
 - [28] I. Bloch. About Properties of Fuzzy Mathematical Morphologies : Proofs of Main Results. Technical report, Télécom Paris 93D023, December 1993.
 - [29] I. Bloch. Distance floue géodésique en traitement d’images. In *Rencontres francophones sur la Logique Floue et ses Applications*, pages 17–24, Paris, France, November 1995.
 - [30] I. Bloch. Fondements des probabilités et des croyances : une discussion des travaux de Cox et Smets. In *15ème Colloque GRETSI*, pages 909–912, Juan-les-Pins, France, September 1995.
 - [31] 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.

- [32] 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.
- [33] 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.
- [34] I. Bloch. Image Information Processing using Fuzzy Sets. In *World Automation Congress, Soft Computing with Industrial Applications*, pages 79–84, Montpellier, France, May 1996.
- [35] I. Bloch. Incertitude, imprécision et additivité en fusion de données : point de vue historique. *Traitemet du Signal*, 13(4) :267–288, 1996.
- [36] 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.
- [37] I. Bloch. Some Aspects of Dempster-Shafer Evidence Theory for Classification of Multi-Modality Medical Images Taking Partial Volume Effect into Account. *Pattern Recognition Letters*, 17(8) :905–919, 1996.
- [38] 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.
- [39] 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 (submitted)*, 1998.
- [40] I. Bloch. Representation of Structural Information in Images using Fuzzy Set Theory (invited conference). In *IEEE SMC*, San Diego, CA, October 1998.
- [41] 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.
- [42] 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.
- [43] I. Bloch. On Fuzzy Distances and their Use in Image Processing under Imprecision. *Pattern Recognition*, 32(11) :1873–1895, 1999.
- [44] I. Bloch. Fusion of Numerical and Structural Image Information in Medical Imaging in the Framework of Fuzzy Sets. In P. Szczepaniak et al., editor, *Fuzzy Systems in Medicine*, Series Studies in Fuzziness and Soft Computing, pages 429–447. Springer Verlag, 2000.
- [45] 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.
- [46] I. Bloch. Geodesic Balls in a Fuzzy Set and Fuzzy Geodesic Mathematical Morphology. *Pattern Recognition*, 33(6) :897–905, 2000.
- [47] I. Bloch. On Links between Mathematical Morphology and Rough Sets. *Pattern Recognition*, 33(9) :1487–1496, 2000.

- [48] 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.
- [49] I. Bloch. Using Mathematical Morphology Operators as Modal Operators for Spatial Reasoning. In *ECAI 2000, Workshop on Spatio-Temporal Reasoning*, pages 73–79, Berlin, Germany, 2000.
- [50] I. Bloch, L. Aurdal, D. Bijno, and J. Müller. Estimation of Class Membership Functions for Grey-Level Based Image Fusion. In *ICIP'97*, volume III, pages 268–271, Santa Barbara, CA, October 1997.
- [51] I. Bloch and A. Hunter (Eds). Fusion : General Concepts and Characteristics. *International Journal of Intelligent Systems*, 16(10) :1107–1134, oct 2001.
- [52] I. Bloch and J. Lang. Towards Mathematical Morpho-Logics. In *8th International Conference on Information Processing and Management of Uncertainty in Knowledge based Systems IPMU 2000*, volume III, pages 1405–1412, Madrid, Spain, 2000.
- [53] 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.
- [54] 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.
- [55] I. Bloch and H. Maître. Fusion de données en traitement d’images : modèles d’information et décisions. *Traitemet du Signal*, 11(6) :435–446, 1994.
- [56] I. Bloch and H. Maître. Fuzzy Mathematical Morphology. *Annals of Mathematics and Artificial Intelligence*, 10 :55–84, 1994.
- [57] I. Bloch and H. Maître. Fuzzy Mathematical Morphologies : A Comparative Study. *Pattern Recognition*, 28(9) :1341–1387, 1995.
- [58] I. Bloch and H. Maître. Extending Adjacency to Fuzzy Sets for Coping with Imprecision Image Objects. In *Accepted to ICIAP'97*, volume I, pages 30–37, Florence, Italy, September 1997.
- [59] I. Bloch and H. Maitre. *Fusion of Image Information under Imprecision*. B. Bouchon Meunier Ed. , Series Studies in Fuzziness, Physica Verlag, Springer, 1997.
- [60] 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.
- [61] I. Bloch, C. Pellet, 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.
- [62] O. Boissier. *Problème du contrôle dans un système de vision intégré de vision : Utilisation d'un système multi-agents*. PhD thesis, Inst. Nat. Polytech. de Grenoble, January 1993.

- [63] O. Boissier and Y. Demazeau. A distributed AI view on a general purpose vision systems. In *3rd Modeling Autonomous Agent in Multi-Agent World Workshop*, Kaiserslautern (Germany), 1991.
- [64] 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.
- [65] B. Bouchon-Meunier and H.T. Nguyen. *Les incertitudes dans les systèmes intelligents*. Number 3110. Presses Universitaires de France, Que Sais-je ?, 1996.
- [66] 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.
- [67] N. Boujemaa, G. Stamon, J. Lemoine, and E. Petit. Fuzzy Ventricular Endocardium Detection with Gradual Focusing Decision. In *14th IEEE EMBS Conference*, pages 1893–1894, Paris, France, 1992.
- [68] T. Bouron. *Structure de communication et d'organisation pour la coopération dans un univers multi-agents*. PhD thesis, Université Pierre et Marie Curie, Paris VI, 1992.
- [69] R. Brooks. Symbolic Reasoning Among 3-D Models and 2-D Images. *Artificial Intelligence*, 17 :285–348, 1981.
- [70] B.G. Buchanan, G. Sutherland, and E.A. Feigenbaum. Heuristic dendral : a program for generating explanatory hypotheses in organic chemistry. In *Machine Intelligence*, pages 205–254, 1969.
- [71] J. J. Buckley and E. Eslami. Fuzzy Plane Geometry I : Points and Lines. *Fuzzy Sets and Systems*, 86 :179–187, 1997.
- [72] P. Buschka, A. Saffiotti, and Z. Wasik. Fuzzy Landmark-Based Localization for a Legged Robot. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1205–1210, Takamatsu, Japan, 2000.
- [73] H. Buxton and R. Howarth. Watching behaviour : the role of context and learning. In *IEEE-ICIP'96*, volume II, pages 797–800, Lausanne, 1996.
- [74] H. Caillol, A. Hillion, and W. Pieczynski. Fuzzy Random Fields and Unsupervised Image Segmentation. *IEEE Trans. on Geoscience and Remote Sensing*, 31(4) :801–810, 1993.
- [75] E. Castillo, J.M. Gutteriez, and A.S. Hadi. *Expert systems and probabilistic network models*. Springer-Verlag, 1997.
- [76] D. Charlebois, D. Goodenough, and S. Matwin. Machine learning from remote sensing. *IGARSS'93, Tokyo*, 1 :165–172, August 1993.
- [77] S. Chauvin. *Evaluation des théories de la décision appliquées à la fusion de capteurs en imagerie satellitaire*. PhD thesis, Ecole Nationale Supérieure des Télécommunications et Université de Nantes, December 1995.
- [78] B. Chellas. *Modal Logic, an Introduction*. Cambridge University Press, Cambridge, 1980.
- [79] S. Y. Chen, W. C. Lin, and C. T. Chen. Evidential Reasoning based on Dempster-Shafer Theory and its Application to Medical Image Analysis. In *SPIE*, volume 2032, pages 35–46, 1993.
- [80] H. D. Cheng and J. R. Chen. Automatically Determine the Membership Function based on the Maximum Entropy Principle. In *2nd Annual Joint Conf. on Information Sciences*, pages 127–130, Wrightsville Beach, NC, 1995.

- [81] 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.
- [82] N. Chleq and M. Thonnat. Real-time image sequence interpretation for video-surveillance applications. In *IEEE-ICIP'96*, volume II, pages 801–804, Lausanne, 1996.
- [83] E. Chouraqui and E. Godbert. Representation of definite descriptions in a semantic network. In *7eme AFCET, Paris*, volume 2, pages 855–868, November 1989.
- [84] M. R. Civanlar and H. J. Trussel. Constructing Membership Functions using Statistical Data. *Fuzzy Sets and Systems*, 18 :1–13, 1986.
- [85] V. Clément and M. Thonnat. A knowledge-based approach to integration of image processing procedures. *CVGIP : Image Understanding*, 2 :166–184, 1993.
- [86] V. Clément and M. Thonnat. A knowledge-based approach to integration of image processing procedures. *CVGIP : Image Understanding*, 2 :166–184, 1993.
- [87] R. Clouard, M. Revenu, A. Elmoataz, and C. Porquet. Un atelier logiciel pour la conception d'applications de segmentation d'images. In *Gretsi*, Juan les Pins (France), 1995.
- [88] R. Cooke. Uncertainty in Risk Assessment : A Probabilist's Manifesto. *Reliability in Engineering and Systems Safety*, 23 :277–283, 1988.
- [89] R. Cooke. *Experts in Uncertainty*. Oxford University Press, Oxford, UK, 1991.
- [90] G.F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42 :393–405, 1990.
- [91] P. R. Cooper, L. A. Birnbaum, and M. E. Brand. Causal scene understanding. *Computer Vision and Image Understanding*, 62(2) :215–231, September 1995.
- [92] T.M. Cover and P.E. Hart. Nearest Neighbor Pattern Classification. *IEEE Trans on Information Theory*, IT-13(1) :21–27, 1967.
- [93] R. T. Cox. Probability, Frequency and Reasonable Expectation. *Journal of Physics*, 14(1) :115–137, 1946.
- [94] D. Crevier and R. Lepage. Knowledge based image understanding systems : A survey. *Computer Vision and Image Understanding*, 67(2) :161–185, 1997.
- [95] P. Cucka and A. Rosenfeld. Evidence-based Pattern Matching Relaxation. Technical Report CAR-TR-623, Center of Automation Research, University of Maryland, May 1992.
- [96] B. V. Dasarathy. Fusion Strategies for Enhancing Decision Reliability in Multi-Sensor Environments. *Optical Engineering*, 35(3) :603–616, March 1996.
- [97] B. de Baets. Fuzzy Morphology : a Logical Approach. In B. Ayyub and M. Gupta, editors, *Uncertainty in Engineering and Sciences : Fuzzy Logic, Statistics and Neural Network Approach*, pages 53–67. Kluwer Academic, 1997.
- [98] B. de Finetti. La prévision : ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7(1) :1–68, 1937.
- [99] Ph. Dejean and P. Dalle. Modèle symbolique de la donnée de traitement d'images. In *10 ème Congrès RFIA (Rennes)*, pages 746–755, 1996.
- [100] M. Delgado and S. Moral. On the Concept of Possibility-Probability Consistency. *Fuzzy Sets and Systems*, 21(3) :311–318, 1987.

- [101] S. Dellepiane, G. Venturi, and G. Vernazza. Model Generation and Model Matching of Real Images by a Fuzzy Approach. *Pattern Recognition*, 25(2) :115–137, 1992.
- [102] Y. Demazeau. *Niveaux de représentation pour la vision par ordinateur*. PhD thesis, Inst. Nat. Polytechnique de Grenoble, 1986.
- [103] Y. Demazeau and J.P. Müller. *Decentralized Artificial Intelligence*. Elsevier Science (North-Holland), 1990.
- [104] 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.
- [105] G. Demoment. Probabilités, modélisation des incertitudes, inférence logique, et traitement des données expérimentales. Technical report, Cours de l'Université de Paris-Sud, Orsay, France, 1993.
- [106] T.-Q. Deng and H. Heijmans. Grey-Scale Morphology Based on Fuzzy Logic. Technical Report PNA-R0012, CWI, Amsterdam, NL, 2000.
- [107] T. Denœux. A k-nearest Neighbor Classification Rule based on Dempster-Shafer Theory. *IEEE Transactions on Systems, Man and Cybernetics*, 25(5) :804–813, 1995.
- [108] J. Desachy. ICARE : an expert system for automatic mapping from satellite imagery. In L. F. Pau Springer Verlag Berlin Heidelberg, editor, *Mapping and spatial modelling for navigation*, volume F65 of *NATO-ASI*, 1990.
- [109] X. Descombes. *Champs markoviens en analyse d'images*. PhD thesis, Télécom-Paris ENST 93E026, dec 1993.
- [110] X. Descombes, M. Moctezuma, H. Maître, and J.-P. Rudant. Coastline Detection by a Markovian Segmentation in SAR Images. *Signal Processing*, 55(1) :123–132, November 1996.
- [111] C. Desrousseaux and D. Pomorski. Optimisation entropique des systèmes de détection distribuée. *Traitements du Signal*, 16(4), 1999.
- [112] S. Deveughele and B. Dubuisson. Using Possibility Theory in Perception : An Application in Artificial Vision. In *Second IEEE International Conference on Fuzzy Systems*, pages 821–826, San Francisco, California, 1993.
- [113] B. Bharathi Devi and V. V. S. Sarma. Estimation of Fuzzy Memberships from Histograms. *Information Sciences*, 35 :43–59, 1985.
- [114] V. di Gesu. Mathematical Morphology and Image Analysis : A Fuzzy Approach. In *Workshop on Knowledge-Based Systems and Models of Logical Reasoning*, Reasoning, 1988.
- [115] 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.
- [116] P. Dokládal. *Grey-Scale Image Segmentation : A Topological Approach*. PhD thesis, Université Marne la Vallée, jan 2000.
- [117] B.A. Draper, A. R. Hanson, and E.M. Riseman. Knowledge directed vision : Control, learning and integration. *Proc of IEEE*, 84(11) :1625–1637, 1996.
- [118] A. Dromigny-Badin, S. Rossato, and Y. M. Zhu. Fusion de données radioscopiques et ultrasonores via la théorie de l'évidence. *Traitements du Signal*, 14(5) :147–160, 1997.

- [119] D. Dubois and M.-C. Jaulent. A General Approach to Parameter Evaluation in Fuzzy Digital Pictures. *Pattern Recognition Letters*, 6 :251–259, 1987.
- [120] D. Dubois and H. Prade. *Fuzzy Sets and Systems : Theory and Applications*. Academic Press, New-York, 1980.
- [121] 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.
- [122] 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.
- [123] D. Dubois and H. Prade. Unfair Coins and Necessity Measures : Towards a Possibilistic Interpretation of Histograms. *Fuzzy Sets and Systems*, 10(1) :15–20, 1983.
- [124] D. Dubois and H. Prade. A Review of Fuzzy Set Aggregation Connectives. *Information Sciences*, 36 :85–121, 1985.
- [125] D. Dubois and H. Prade. On the Unicity of Dempster Rule of Combination. *International Journal of Intelligent Systems*, 1 :133–142, 1986.
- [126] D. Dubois and H. Prade. *Possibility Theory*. Plenum Press, New-York, 1988.
- [127] D. Dubois and H. Prade. Representation and Combination of Uncertainty with Belief Functions and Possibility Measures. *Computational Intelligence*, 4 :244–264, 1988.
- [128] D. Dubois and H. Prade. Fuzzy Sets in Approximate Reasoning, Part I : Inference with Possibility Distributions. *Fuzzy Sets and Systems*, 40 :143–202, 1991.
- [129] D. Dubois and H. Prade. Combination of Information in the Framework of Possibility Theory. In M. Al Abidi, editor, *Data Fusion in Robotics and Machine Intelligence*. Academic Press, 1992.
- [130] 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.
- [131] D. Dubois and H. Prade. La problématique scientifique du traitement de l'information. *Information-Interaction-Intelligence*, 1(2), 2001.
- [132] D. Dubois, H. Prade, and C. Testemale. Weighted Fuzzy Pattern Matching. *Fuzzy Sets and Systems*, 28 :313–331, 1988.
- [133] 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.
- [134] B. Dubuisson. *Diagnostic et reconnaissance des formes*. Hermès, Paris, 1990.
- [135] R. Duda and P. Hart. *Pattern Classification and Scene Analysis*. Wiley, New-York, 1973.
- [136] R.O. Duda and O.E. Hart. *Pattern Classification and Scene Analysis*. John Wiley and Sons, New York (USA), 1973.
- [137] A. Edwards and L. Cavalli-Sforza. A Method for Cluster Analysis. *Biometrics*, 21 :362–376, 1965.
- [138] B. Efron. Why isn't Everyone a Bayesian ? *The American Statistician*, 40(1) :1–11, 1986.

- [139] M. A. Eshera and King-Sun Fu. An image understanding system using attributed symbolic representation and inexact graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(5) :604–618, 1986.
- [140] O. Faugeras and K. Price. Semantic description of aerial images using stochastic labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3(6) :633–642, November 1981.
- [141] W. Feller. *An Introduction to Probability Theory and its Applications*. Wiley, New York, 1966.
- [142] J. Ferber and L. Gasser. *Intelligence Artificielle Distribuée*. J.C. Rault Ed., 1991.
- [143] W. Forstner and T. Locherbach. Fusiong information in remote sensing. *ISPRS'92 (Washington D. C.)*, 1992.
- [144] M. J. Frank. On the Simultaneous Associativity of $f(x, y)$ and $x + y - f(x, y)$. *Aequationes Mathematicae*, 19 :194–226, 1979.
- [145] S. French. Group Consensus Probability Distributions : A Critical Survey. In J. Bernardo et al., editor, *Bayesian Statistics*, pages 183–201. Elsevier, The Netherlands, 1985.
- [146] L. Gacôgne. About a Foundation of Dempster’s Rule. Technical report, Laforia 93/27, 1993.
- [147] H. Ganster and A. Pinz. Active Fusion using Dempster-Shafer Theory of Evidence. In A. Pinz, editor, *Proceedings of 20th ÖAGM / AAPR Workshop*, Schriftenreihe der OCG. Oldenbourg, 1996.
- [148] C. Garbay. Quelques propositions pour la réalisation d’un système expert de segmentation d’image. *Traitemet du Signal*, 4(3) :229–237, 1987.
- [149] C. Garbay. Architectures logicielles et contrôle dans les systèmes de vision. In J. M. Jolion, editor, *Les systèmes de Vision*, chapter 7, pages 197–252. Hermès, 2001.
- [150] C. Garbay, F. Bellet, and A. Boucher. Des agents situés pour l’interprétation de scènes. *revue d’Intelligence Artificielle*, 1(12) :11–36, 1998.
- [151] P. Garnesson, G. Giraudon, and P. Montesinos. Detecting buildings and roads in aerial images by a multi expert system. *Rapport de recherche INRIA*, 1988.
- [152] P. Garnesson, G. Giraudon, and P. Montesinos. Messie : un système multi-spécialistes en vision, application à l’interprétation en imagerie aérienne. *RFIA-AFCET’89 (Paris)*, pages 817–832, 1989.
- [153] T. D. Garvey. Evidential Reasoning for Land-Use Classification. In *Analytical Methods in Remote Sensing for Geographic Information Systems, International Association of Pattern Recognition, Technical Committee 7 Workshop*, Paris, October 1986.
- [154] 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.
- [155] T. Géraud, I. Bloch, and H. Maître. Reconnaissance de structures cérébrales à l’aide d’un atlas at par fusion d’informations structurelles floues. In *RFIA 2000*, volume I, pages 287–295, Paris, France, 2000.
- [156] T. Géraud, J.-F. Mangin, I. Bloch, and H. Maître. Segmenting Internal Structures in 3D MR Images of the Brain by Markovian Relaxation on a Watershed Based Adjacency Graph. In *ICIP-95*, Washington DC, October 1995.

- [157] R. Gerber and H.H. Nagel. Knowledge representation for the generation of quantified natural language descriptions of vehicle traffic in image sequences. In *IEEE-ICIP'96*, volume II, pages 805–808, Lausanne, 1996.
- [158] G. Gerla and L. Scarpati. Extension Principles for Fuzzy Set Theory. *Journal of Information Sciences*, 106 :49–69, 1998.
- [159] 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.
- [160] V. Goetcherian. From Binary to Grey Tone Image Processing using Fuzzy Logic Concepts. *Pattern Recognition*, 12 :7–15, 1980.
- [161] R. Goetschel and W. Voxman. Topological Properties of Fuzzy Numbers. *Fuzzy Sets and Systems*, 10 :87–99, 1983.
- [162] R. Goetschel and W. Voxman. Elementary Fuzzy Calculus. *Fuzzy Sets and Systems*, 18 :31–43, 1986.
- [163] J. A. Goguen. The Logic of Inexact Concepts. *Synthese*, 19 :325–373, 1969.
- [164] I. J. Good. Kinds of Probability. *Science*, 129(3347) :443–447, 1959.
- [165] J. Gordon and E. H. Shortliffe. A Method for Managing Evidential Reasoning in a Hierarchical Hypothesis Space. *Artificial Intelligence*, 26 :323–357, 1985.
- [166] M. Grabisch. Fuzzy Integral in Multicriteria Decision Making. *Fuzzy Sets and Systems*, 69 :279–298, 1995.
- [167] M. Grabisch, T. Murofushi, and M. Sugeno. Fuzzy Measures of Fuzzy Events Defined by Fuzzy Integrals. *Fuzzy Sets and Systems*, 50 :293–313, 1992.
- [168] J.-F. Grandin and M. Marques. Robust Data Fusion. In *Fusion 2000*, pages MoC3–1–9, Paris, France, 2000.
- [169] J. Guan and D. A. Bell. *Evidence Theory and its Applications*. North-Holland, Amsterdam, 1991.
- [170] M.E. De Gunst. *Knowledge based interpretation of aerial images for Updating of road maps*. PhD thesis, Faculty of Geodesy, Delft University, 1996.
- [171] D. L. Hall and J. Llinas. An Introduction to Multisensor Data Fusion. *Proceedings of the IEEE*, 85(1) :6–23, 1997.
- [172] H. Hamacher. Ueber logische Verknupfungen Unscharfer Aussagen und deren Zugehörige Bewertungsfunktionen. *Progress in Cybernetics and System Research*, 3 :276–287, 1978.
- [173] A. Hanson, E. Riseman, and T. Williams. Sensor and information fusion from knowledge based constraints. *SPIE Sensor Fusion*, 931 :186–196, 1988.
- [174] A.R. Hanson and E.M. Rieseman. *Visions : a computer system for interpreting scenes*, pages 303–333. A.R. Hanson and E.M. Rieseman, Academic Press, 1978.
- [175] J.P. Haton, N. Bouzid, F. Charpillet, M.C. Haton, B. Laasri, H. Laasri, P. Marquis, T. Mondot, and A. Napoli. *Le raisonnement en intelligence artificielle*. InterEditions, Paris, 1991.
- [176] B. Hayes-Roth. The blackboard architecture : A general framework for problem solving. Technical report, Computer Science Department, Stanford University, 1983.

- [177] S. Le Hégarat-Mascle, I. Bloch, and D. Vidal-Madjar. Introduction of Neighborhood Information in Evidence Theory and Application to Data Fusion of Radar and Optical Images with Partial Cloud Cover. *Pattern Recognition*, 31(11) :1811–1823, 1998.
- [178] J. D. Holland. The Reverend Thomas Bayes, F.R.S. (1702-61). *J. Roy. Stat. Soc. (A)*, 125 :451–461, 1962.
- [179] R. Horaud and T. Skordas. Stereo correspondence through feature grouping and maximal cliques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-11(11) :1168–1180, November 1989.
- [180] E. J. Horvitz, D. E. Heckerman, and C. P. Langlotz. A Framework for Comparing Alternative Formalisms for Plausible Reasoning. In *National Conference on Artificial Intelligence*, pages 210–214, 1986.
- [181] S. Houzelle. *Extraction automatique d'objets cartographiques par fusion d'informations extraites d'images satellites*. PhD thesis, Ecole Nationale Supérieure des Télécommunications (93 E 001), 1993.
- [182] S. Houzelle and G. Giraudon. Contribution to multisensor fusion formalization. *Robotics and Autonomous Systems*, 13 :69–85, 1995.
- [183] G. E. Hughes and M. J. Cresswell. *An Introduction to Modal Logic*. Methuen, London, UK, 1968.
- [184] M. Ishizuka. Inference Methods based on Extended Dempster Shafer's Theory for Problems with Uncertainty / Fuzziness. *New Generation Computing*, 1 :159–168, 1983.
- [185] E. T. Jaynes. Information Theory and Statistical Mechanics. *Physical Review*, 106(4) :620–630, 1957.
- [186] R. Jeffreys. *Theory of Probability*. Oxford University Press, 1961.
- [187] F. V. Jensen. *Bayesian networks and decision graphs*. Springer-Verlag, 2001.
- [188] S.H. Joseph and T.P. Pridmore. Knowledge directed interpretation of mechanical engineering drawings. *IEEE-PAMI*, 14(9) :928–940, 1992.
- [189] A. Kaufmann. *Introduction to the Theory of Fuzzy Subsets*. Academic Press, New-York, 1975.
- [190] A. Kaufmann and M. M. Gupta. *Fuzzy Mathematical Models in Engineering and Management Science*. North-Holland, Amsterdam, 1988.
- [191] 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.
- [192] E. C. Kemble. Is the Frequency Theory of Probability Adequate for All Scientific Purposes? *Am. J. Physics*, 10 :6–16, 1942.
- [193] J. M. Keynes. *A Treatise on Probability*. Macmillan, London, 1929.
- [194] F. Klawonn and E. Schwecke. On the Axiomatic Justification of Dempster's Rule of Combination. *International Journal of Intelligent Systems*, 7 :469–478, 1992.
- [195] G. J. Klir and T. A. Folger. *Fuzzy Sets, Uncertainty, and Information*. Englewood Cliffs, 1988.
- [196] G. J. Klir and B. Parviz. Probability-Possibility Transformations : A Comparison. *Int. J. General Systems*, 21 :291–310, 1992.
- [197] L. T. Koczy. On the Description of Relative Position of Fuzzy Patterns. *Pattern Recognition Letters*, 8 :21–28, 1988.

- [198] M. Koshelev. Fuzzy Logic Explains the Golden Proportion. *International Journal of Intelligent Systems*, 12(5) :415–417, 1997.
- [199] B. Kosko. Fuzziness vs. Probability. *International Jounal of General Systems*, 17 :211–240, 1990.
- [200] 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.
- [201] R. Krishnapuram and J. M. Keller. A Possibilistic Approach to Clustering. *IEEE Transactions on Fuzzy Systems*, 1(2) :98–110, 1993.
- [202] 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.
- [203] S. Kullback. *Information Theory and Statistics*. Wiley, New York, 1959.
- [204] V.P. Kumar and U.B. Desai. Image interpretation using Bayesian networks. *IEEE trans on PAMI*, 18(1) :74–77, 1996.
- [205] 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.
- [206] 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.
- [207] 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.
- [208] 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.
- [209] T. Lee, J. A. Richards, and P. H. Swain. Probabilistic and Evidential Approaches for Multisource Data Analysis. *IEEE Transactions on Geoscience and Remote Sensing*, GE-25(3) :283–293, 1987.
- [210] V. Lefèvre. *Système Multi-agents d'aide à la photo-interprétation sur architectures multi-processeur*. PhD thesis, Université Paris XI (Orsay), 1994.
- [211] H. Leung. Neural Networks Data Association with Application to Multiple-Target Tracking. *Optical Engineering*, 35(3) :693–700, March 1996.
- [212] L. Likforman-Sulem, H. Maître, and C. Sirat. Un système de vision pour l'expertise de textes. In *7ème Congrès AFCET RFIA*, pages 989–998, Paris, France, November 1989.
- [213] L. Likforman-Sulem, H. Maître, and C. Sirat. An expert vision system for analysis of Hebrew characters and Authentication of Manuscripts. *Pattern Recognition*, 24(2) :121–137, 1991.
- [214] Laurence Likforman-Sulem. *Mise en œuvre d'un système expert pour l'analyse de caractères hébreux en vue de l'authentification de manuscrits*. PhD thesis, Télécom-Paris ENST 89E003, March 1989.

- [215] W. Liu, J. Hong, M. F. McTear, and J. G. Hughes. An Extended Framework for Evidential Reasoning Systems. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(3) :441–457, 1993.
- [216] X. Liu, S. Tan, V. Srinivasan, S. H. Ong, and Z. Xie. Fuzzy Pyramid-based Invariant Object Recognition. *Pattern Recognition*, 27(5) :741–756, 1994.
- [217] R. Lowen and W. Peeters. Distances between Fuzzy Sets Representing Grey Level Images. *Fuzzy Sets and Systems*, 99(2) :135–150, 1998.
- [218] J. D. Lowrance, T. M. Strat, L. P. Wesley, T. D. Garvey, E. H. Rusconi, and D. E. Wilkins. The Theory, Implementation and Practice of Evidential Reasoning. SRI project 5701 final report, SRI, Palo Alto, June 1991.
- [219] 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.
- [220] R. C. Luo and M. G. Kay. Multisensor Integration and Fusion in Intelligent Systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(5) :901–931, 1989.
- [221] A. Lux and V. Souvignier. Pv_v - un système de vision appliquant une stratégie de prédiction - vérification. *RFIA-AFCET'84 (Paris)*, pages 223–234, 1984.
- [222] A. K. Mackworth. Quick and clean : constraints based vision for situated robots. In *IEEE-ICIP*, volume II, pages 789–792, Lausanne, 1996.
- [223] H. Maître. Utilisation de l'imagerie aérienne et satellitaire pour l'aménagement du territoire. Cours du 8ème Congrès AFCET de Lyon, Tutorial 4, 26 nov 1991, 38 pages, 1991.
- [224] H. Maître. Entropy, Information and Image. In H. Maître and J. Zinn-Justin, editors, *Progress in Picture Processing, Les Houches Session LVIII*, pages 881–1115. Springer Verlag, 1996.
- [225] H. Maître, I. Bloch, and M. Sigelle. Spatial entropy : a tool for contextual classification control. In *IEEE Int. Conf on Image Processing*, volume II, pages 212–216, Austin, Texas, November 1994.
- [226] 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.
- [227] J. F. Mangin, V. Frouin, I. Bloch, J. Lopez-Krahe, and B. Bendriem. Fast Nonsupervised 3D Registration of PET and MR Images of the Brain. *Journal of Cerebral Blood Flow and Metabolism*, 14 :749–762, 1994.
- [228] J.-F. Mangin, V. Frouin, J. Régis, I. Bloch, P. Belin, and Y. Samson. Towards Better Management of Cortical Anatomy in Multi-Modal Multi-Individual Brain Studies. *Physica Medica*, XII :103–107, 1996.
- [229] J.-F. Mangin, J. Regis, I. Bloch, V. Frouin, Y. Samson, and J. Lopez-Krahe. A Markovian Random Field based Random Graph Modelling the Human Cortical Topography. In *CVRMed'95*, pages 177–183, Nice, France, April 1995.
- [230] M. Marengoni, A. R. Hanson, S. Zilberstein, and E.M. Riseman. Decision making and uncertainty management in 3D reconstruction system. *IEEE trans on PAMI*, 25(7) :852–858, 2003.
- [231] S. Mascle, I. Bloch, and D. Vidal-Madjar. Application of Dempster-Shafer Evidence Theory to Unsupervised Classification in Multisource Remote Sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 35(4) :1018–1031, 1997.

- [232] G. Masini and R. Mohr. Mirabelle a system for structural analysis of drawings. *Pattern Recognition*, 16(4) :363–372, 1988.
- [233] G. Matheron. La théorie des variables régionalisées et ses applications. Technical Report 5, Ecole des Mines de Paris, Centre de Morphologie Mathématique de Fontainebleau, 1970.
- [234] G. Matheron. Estimer et choisir - essai sur la pratique des probabilités. Technical report, Ecole Nationale Supérieure des Mines de Paris, Centre de Géostatistique, Fontainebleau, France, 1978.
- [235] 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.
- [236] T. Matsuyama. Knowledge-based aerial image understanding systems and expert systems for image processing. *International Geoscience and Remote Sensing Symposium (Zurich)*, pages 1026–1038, 1986.
- [237] 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.
- [238] D. M. McKeown, W. A. Harvey, and J. McDermott. Rule-based interpretation of aerial imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7(5) :570–585, 1985.
- [239] D. M. McKeown, W.A. Harvey, and L. E. Wixson. Automatic knowledge acquisition for aerial image interpretation. *Computer Vision, Graphics and Image Processing*, 46 :37–81, 1989.
- [240] M. Ménard, E. H. Zahzah, and A. Shahin. Mass Function Assessment : Case of Multiple Hypotheses for the Evidential Approach. In *Europto Conf. on Image and Signal Processing for Remote Sensing*, Taormina, Italy, September 1996.
- [241] K. Menger. Statistical Metrics. *Proc. National Academy of Sciences USA*, 28 :535–537, 1942.
- [242] P. Mignot and J. Lévy-Véhel. Arthur : un système d’analyse de textures. *Traitemet du Signal*, 9(6) :507–516, 1993.
- [243] N. Milisavljević and I. Bloch. A Two-Level Approach for Modeling and Fusion of Humanitarian Mine Detection Sensors within the Belief Function Framework. In *Applied Stochastic Models and Data Analysis*, volume 2, pages 743–748, Compiègne, France, jun 2001.
- [244] N. Milisavljević, I. Bloch, and M. Acheroj. Characterization of Mine Detection Sensors in Terms of Belief Functions and their Fusion, First Results. In *3rd International Conference on Informatin Fusion - Fusion 2000*, volume II, pages ThC3 15–22, Paris, France, jul 2000.
- [245] 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.
- [246] 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.

- [247] S. Moisan. *Une plate-forme pour une programmation par composants de systèmes à base de connaissances*. PhD thesis, Habilitation à Diriger des Recherches de l'Université de Nice, April 1998.
- [248] H. Moissinac, H. Maître, and I. Bloch. Graph Based Urban Scene Analysis Using Symbolic Data. In *SPIE Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision II*, volume 2486, pages 93–104, Orlando, Florida, April 1995.
- [249] H. Moissinac, H. Maître, and I. Bloch. Markov Random Fields and Graphs for Uncertainty Management and Symbolic Data Fusion in a Urban Scene Interpretation. In *SPIE/EUROPTO Conference on Image and Signal Processing for Remote Sensing*, volume 2579, pages 298–309, Paris, France, September 1995.
- [250] 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.
- [251] P. Levi N. Oswald. Cooperative vision in a multi-agent architecture. In Del Bimbo, editor, *Image Analysys and Processing - 9th ICIAP'97*, volume 1, pages 709–716. Lecture Notes in Computer Science , Springer,1311, 1997.
- [252] M. Nachtegael and E. E. Kerre. Classical and Fuzzy Approaches towards Mathematical Morphology. In E. E. Kerre and M. Nachtegael, editors, *Fuzzy Techniques in Image Processing*, Studies in Fuzziness and Soft Computing, chapter 1, pages 3–57. Physica-Verlag, Springer, 2000.
- [253] M. Nakatsuyama. Fuzzy Mathematical Morphology for Image Processing. In *ANZIIS-93*, pages 75–79, Perth, Western Australia, 1993.
- [254] A. Nazif. *A ruled based expert system for image segmentation*. PhD thesis, McGill University (Canada), 1983.
- [255] A. Nazif and M.D. Levine. Low level image segmentation : an expert system. *IEEE PAMI*, 9(4) :555–577, 1984.
- [256] R. E. Neapolitan. A Survey of Uncertain and Approximate Inference. In L. Zadeh and J. Kaprzyk, editors, *Fuzzy Logic for the Management of Uncertainty*, pages 55–82. J. Wiley, New York, 1992.
- [257] R.E. Neapolitan. *Probabilistic reasonning in intelligent systems*. J. Wiley, New-York, 1990.
- [258] H. Niemann, H. Bunke, I. Hofmann, G. Sagerer, F. Wolf, and H. Feistel. A knowledge based system for analysis of gated blood pool studies. *IEEE-PAMI*, 7(3) :246–259, 1985.
- [259] D. Niyogi and S.N. Shrihari. Using a blackboard architecture for control in a kb document understanding system. In *Application of Artificial Intelligence V*, volume 786, pages 319–326. SPIE, 1987.
- [260] S. K. Pal. Fuzzy Set Theoretic Measures for Automatic Feature Evaluation. *Information Science*, 64 :165–179, 1992.
- [261] 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.
- [262] S. K. Pal and A. Rosenfeld. Image Enhancement and Thresholding by Optimization of Fuzzy Compactness. *Pattern Recognition Letters*, 7 :77–86, 1988.
- [263] J. B. Paris. *The Uncertain Reasoner's Companion, a Mathematical Perspective*. Cambridge University Press, 1995.

- [264] L. Pau. Knowledge representation approaches in sensor fusion. *Automatica*, 25(2) :207–214, 1989.
- [265] J. Pearl. Fusion, Propagation, and Structuring in Belief Networks. *Artificial Intelligence*, 29 :241–288, 1986.
- [266] J. Pearl. On Evidential Reasoning in a Hierarchy of Hypotheses. *Artificial Intelligence*, 28 :9–15, 1986.
- [267] J. Pearl. *Probabilistic reasonning in Intelligent Systems*. Morgan-Kaufmann, San-Mateo, USA, 1988.
- [268] C. Pellet, 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.
- [269] 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.
- [270] W. Pieczynski and J.M. Cahen. Champs de Markov cachés flous et segmentation d'images. *Rev. Statistique Appliquée*, XLII(3) :13–31, 1994.
- [271] A. Pinz and M. Prantl. Active Fusion for Remote Sensing Image Understanding. In *European Symposium on Satellite Remote Sensing, Paris*, volume 2579, pages 67–77. EOS/SPIE, 1995.
- [272] A. Popov. Convexity Indicators based on Fuzzy Morphology. *Pattern Recognition Letters*, 18 :259–267, 1997.
- [273] A. T. Popov. Morphological Operations on Fuzzy Sets. In *IEE Image Processing and its Applications*, pages 837–840, Edinburgh, UK, July 1995.
- [274] J. Puentes. *Analyse symbolique du mouvement cardiaque en angiographie vasculaire*. PhD thesis, Université Rennes 1, 1996.
- [275] P. Quinio. Representation and Accumulation of Uncertain Informations : A Theoretical Comparison of Probabilistic and some Non-Probabilistic Formalisms. Technical report, Ito Lab., Tohoku University, 1989.
- [276] P. Quinio and T. Matsuyama. Random Closed Sets : A Unified Approach to the Representation of Imprecision and Uncertainty. In R. Kruse and P. Siegel, editors, *Symbolic and Quantitative Approaches to Uncertainty, ECSQARU*, pages 282–286, Marseille, 1991. Springer Verlag.
- [277] H. Rasoulian, W. E. Thompson, L. F. Kazda, and R. Parra-Loera. Application of the Mathematical Theory of Evidence to the Image Cueing and Image Segmentation Problem. In *SPIE Signal and Image Processing Systems Performance Evaluation*, volume 1310, pages 199–206, 1990.
- [278] H. Reichgelt. *Knowledge Representation : An AI Perspective*. Ablex Publishing, 1991.
- [279] M. Rombaut. Fusion de données images segmentées à l'aide du formalisme de Dempster-Shafer. In *GRETISI'99*, pages 655–658, Vannes, France, 1999.
- [280] J. B. Romine and E. W. Kamen. Modeling and Fusion of Radar and Imaging Sensor Data for Target Tracking. *Optical Engineering*, 35(3) :659–673, March 1996.
- [281] 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.
- [282] A. Rosenfeld. Fuzzy Digital Topology. *Information and Control*, 40 :76–87, 1979.

- [283] A. Rosenfeld. The Fuzzy Geometry of Image Subsets. *Pattern Recognition Letters*, 2 :311–317, 1984.
- [284] A. Rosenfeld. Distances between Fuzzy Sets. *Pattern Recognition Letters*, 3 :229–233, 1985.
- [285] 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.
- [286] A. Rosenfeld and R. Klette. Degree of Adjacency or Surroundness. *Pattern Recognition*, 18(2) :169–177, 1985.
- [287] F. Russo and G. Ramponi. Introducing the Fuzzy Median Filter. In *Signal Processing VII : Theories and Applications*, pages 963–966, 1994.
- [288] 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.
- [289] A. Saffiotti. Artificial Intelligence fo Embedded Systems. Technical report, AASS, University of Örebro, Sweden, 2002.
- [290] M. Salotti. *Gestion des informations dans les premières étapes de la vision par ordinateur*. PhD thesis, INPG Grenoble, 1994.
- [291] F. Salzenstein and W. Pieczynski. Unsupervised Bayesian Segmentation using Hidden Fuzzy Markov Fields. In *IEEE Int. Conf. on Acoustics, Speech and Signal Procesing*, Detroit, Michigan, 1995.
- [292] F. Sandakly. *Contribution à la mise en oeuvre d'une architecture à base de connaissance pour l'interprétation de scènes 2D et 3D*. PhD thesis, Université de Nice-Sophia-Antipolis, 1995.
- [293] K.J. Schilling and T. Vógtle. Satellite image analysis using integrated knowledge processing. In *ISPRS'96 Commission III*, pages 752–757, Vienne (Autiche), 1996.
- [294] M. Schmitt and J. Mattioli. *Morphologie mathématique*. Masson, Paris, 1994.
- [295] R. A. Schowengerdt and H. Wang. A general purpose expert system for image processing. *Photogrammetric Engineering and Remote Sensing*, 55(9) :1277–1284, 1989.
- [296] C. Schröder and amm B. Neu. On the Logics of Image Interpretation : Model Construction in a Formal Knowledge Representation Framework. In *IEEE International Conference on Image Processing ICIP 96*, volume II, pages 785–789, Lausanne, 1996.
- [297] J. Schubert. On Nonspecific Evidence. *International Journal of Intelligent Systems*, 8 :711–725, 1993.
- [298] B. Schweizer and A. Sklar. Associative Functions and Abstract Semigroups. *Publ. Math. Debrecen*, 10 :69–81, 1963.
- [299] B. Schweizer and A. Sklar. *Probabilistic Metric Spaces*. North Holland, Amsterdam, 1983.
- [300] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, London, 1982.
- [301] J. Serra. *Image Analysis and Mathematical Morphology, Part II : Theoretical Advances*. Academic Press (J. Serra Ed.), London, 1988.
- [302] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

- [303] G. Shafer. Non-Additive Probabilities in the Work of Bernoulli and Lambert. *Archive for History of Exact Sciences*, 19 :309–370, 1978.
- [304] G. Shafer. The Combination of Evidence. *International Journal of Intelligent Systems*, 1 :155–179, 1986.
- [305] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, 1959.
- [306] E. H. Shortliffe and B. G. Buchanan. A Model of Inexact Reasoning in Medicine. *Mathematical Biosciences*, 23 :351–379, 1975.
- [307] E.H. Shortliffe. *Computer Based medical consultation : Mycin*. Elsevier, 1976.
- [308] J. C. Simon. *From Pixels to Features*. V-X, North Holland, Amsterdam, 1989.
- [309] D. Sinha and E. Dougherty. Fuzzy Mathematical Morphology. *Journal of Visual Communication and Image Representation*, 3(3) :286–302, 1992.
- [310] D. Sinha and E. R. Dougherty. Fuzzification of Set Inclusion : Theory and Applications. *Fuzzy Sets and Systems*, 55 :15–42, 1993.
- [311] 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.
- [312] P. Smets. Medical Diagnosis : Fuzzy Sets and Degree of Belief. In *Colloque International sur la Théorie et les Applications des Sous-Ensembles Flous*, Marseille, September 1978.
- [313] P. Smets. The Degree of Belief in a Fuzzy Event. *Information Sciences*, 25 :1–19, 1981.
- [314] P. Smets. The Combination of Evidence in the Transferable Belief Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5) :447–458, 1990.
- [315] P. Smets. Constructing the Pignistic Probability Function in a Context of Uncertainty. *Uncertainty in Artificial Intelligence*, 5 :29–39, 1990.
- [316] P. Smets. Belief Functions : The Disjunctive Rule of Combination and the Generalized Bayesian Theorem. *International Journal of Approximate Reasoning*, 9 :1–35, 1993.
- [317] P. Smets. Quantifying Beliefs by Belief Functions : an Axiomatic Justification. In *13th Int. Joint. Conf. on Artificial Intelligence*, pages 598–603, Chaméry, France, 1993.
- [318] P. Smets, E.H. Mandani, D. Dubois, and H. Prade. *Non standard logics for automated reasoning*. Academic Press, 1988.
- [319] G. Socher, G. A. Fink, F. Kummert, and G. Sagerer. A Hybrid Approach to Identifying Objects from Verbal Descriptions. In *Workshop on Multi-Lingual Spontaneous Speech Recognition in Real Environments*, Nancy, France, 1996.
- [320] Léa Sombé. *Raisonnements sur des informations incomplètes en intelligence artificielle*. Teknea, Marseille, 1989.
- [321] H. Maître (sous la direction de). *Le traitement des images*. Hermès, Traité IC2, Paris, France, 2002.
- [322] I. Bloch (sous la direction de). *Fusion d'informations en traitement du signal et des images*. Hermès, Paris, France, 2003.
- [323] S. Stansfield. Angy : a ruled based expert system for identifying and isolating coronary vessels in digital angiograms. *IEEE PAMI*, 8(2) :188–199, 1986.

- [324] S. M. Stigler. Thomas Bayes's Bayesian Inference. *J. Roy. Stat. (A)*, 145 :250–258, 1982.
- [325] S. M. Stigler. Who Discovered Bayes's Theorem? *The American Statistician*, 37(4) :290–296, 1983.
- [326] T.M. Strat and M. Fischler. The role of context in computer vision. In *ICCV'95 Context based Vision*, pages 30–37, Cambridge, 1995.
- [327] M. Sugeno. *Theory of Fuzzy Integral and its Applications*. Ph.d. thesis, Tokyo Institute of Technology, Tokyo, 1974.
- [328] S. C. A. Thomopoulos. Sensor Integration and Data Fusion. *Journal of Robotics Systems*, 7(3) :337–372, 1990.
- [329] S. C. A. Thomopoulos, R. Viswanathan, and D. K. Bougoulias. Optimal Distributged Decision Fusion. *IEEE Transactions on Aerospace and Electronic Systems*, 25(5) :761–765, 1989.
- [330] M. Thonnat. *Vers une vision cognitive : mise en œuvre de connaissances et de raisonnements pour l'analyse et l'interprétation d'images*. PhD thesis, Habilitation à Diriger des Recherches de l'Université de Nice, October 1998.
- [331] M. Thonnat, V. Clément, and J. Van den Elst. Supervision of perception tasks for autonomous systems : the ocapı approach. Technical Report 2000, INRIA, Sophia Antipolis (France), 1993.
- [332] 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.
- [333] M. Tribus. *Rational, Descriptions, Decisions and Designs*. Pergamon Press Inc., 1972.
- [334] F. Tupin, I. Bloch, and H. Maître. A First Step Towards Automatic Interpretation of SAR Images using Evidential Fusion of Several Structure Detectors. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3) :1327–1343, 1999.
- [335] F. Tupin, H. Maître, J-F. Mangin, J-M. Nicolas, and E. Pechersky. Linear Feature Detection on SAR Images : Application to the Road Network. *IEEE Transactions on Geoscience and Remote Sensing*, 36(2) :434–453, 1998.
- [336] F. Tupin, M. Sigelle, and H. Maître. Definition of a Spatial Entropy and its Use for Texture Discrimination. In *IEEE ICIP'2000*, volume I, pages 725–728, Vancouver, Canada, 2000.
- [337] S. Turpin-Dhilly. *Adaptation des outils de la morphologie floue à l'analyse de données multidimensionnelles*. PhD thesis, Université de Lille, 2000.
- [338] 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.
- [339] 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.
- [340] J. van Cleynenbreugel, S. A. Osinga, F. Fierens, P. Suetens, and A. Oosterlinck. Road Extraction from Multi-temporal Satellite Images by an Evidential Reasoning Approach. *Pattern Recognition Letters*, 12 :371–380, 1991.
- [341] P. K. Varshney. *Distributed Detection and Data Fusion*. Springer Verlag, New York, 1997.

- [342] P. Vaxivière and K. Tombre. Celesttin iv : Knowledge based analysis of mechanical engineering drawings. In *IEE Conf on Syst. Engineering*, pages 242–245, Kobe (Japan), 1992.
- [343] G. Vernazza, S. Serpico, and S. Dellepiane. A knowledge based system for biomedical image processing and recognition. *IEEE T. Circuits and Systems*, 34(11) :1399–1416, 1987.
- [344] E. Volden, G. Giraudon, and M. Berthod. Modeling Image Redundancy. In *IGARSS'95*, volume 3, pages 2148–2150, Firenze, Italy, 1995.
- [345] L. Wald. Some Terms of Reference in Data Fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3) :1190–1193, 1999.
- [346] R. R. Yager. On the Measure of Fuzziness and Negation. *International Journal of General Systems*, 5 :221–229, 1979.
- [347] R. R. Yager. On a General Class of Fuzzy Connectives. *Fuzzy Sets and Systems*, 4 :235–242, 1980.
- [348] R. R. Yager. Generalized Probabilities of the Degree of Fuzzy Events from Fuzzy Belief Structures. *Information Sciences*, 28 :45–62, 1982.
- [349] R. R. Yager. On the Dempster-Shafer Framework and New Combination Rules. *Information Science*, 41 :93–137, 1987.
- [350] 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.
- [351] R. R. Yager. Connectives and Quantifiers in Fuzzy Sets. *Fuzzy Sets and Systems*, 40 :39–75, 1991.
- [352] R. Y. Yager. Entropy Measures under Similarity Relations. *Int. J. General Systems*, 20 :341–358, 1992.
- [353] M. Yeddanapudi, Y. Bar-Shalom, and K. Pattipati. IMM Estimation for Multi-Target-Multisensor Air Traffic Surveillance. *Proceedings of the IEEE*, 85(1) :80–94, 1997.
- [354] 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.
- [355] J. Yen. Generalizing the Dempster-Shafer Theory to Fuzzy Sets. *IEEE Transactions on Systems, Man and Cybernetics*, 20(3) :559–569, 1990.
- [356] V. R. Young. Fuzzy Subsethood. *Fuzzy Sets and Systems*, 77 :371–384, 1996.
- [357] 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.
- [358] L. A. Zadeh. Fuzzy Sets. *Information and Control*, 8 :338–353, 1965.
- [359] L. A. Zadeh. Similarity Relations and Fuzzy Orderings. *Information Sciences*, 3 :177–200, 1971.
- [360] L. A. Zadeh. The Concept of a Linguistic Variable and its Application to Approximate Reasoning. *Information Sciences*, 8 :199–249, 1975.
- [361] L. A. Zadeh. Fuzzy Sets as a Basis for a Theory of Possibility. *Fuzzy Sets and Systems*, 1 :3–28, 1978.

- [362] L. A. Zadeh. A Theory of Approximate Reasoning. *Machine Intelligence*, 9 :149–194, 1979.
- [363] L. A. Zadeh. Fuzzy Logic = Computing with Words. *IEEE Transactions on Fuzzy Systems*, 4(2) :103–111, 1996.
- [364] E. Zahzah. *Contribution à la représentation des connaissances et à leur utilisation pour l'interprétation automatique des images satellites*. Thèse de doctorat, Université Paul Sabatier, Toulouse, 1992.
- [365] H. J. Zimmermann. *Fuzzy Set Theory and its Applications*. Kluwer Academic Publisher, Boston, 1991.
- [366] H. J. Zimmermann and P. Zysno. Latent Connectives in Human Decision Making. *Fuzzy Sets and Systems*, 4 :37–51, 1980.
- [367] 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.