Fusion d'informations numériques : panorama méthodologique

Isabelle Bloch
Ecole Nationale Supérieure des Télécommunications, Dép. TSI, CNRS UMR 5141 LTCI
Paris, France - Isabelle.Bloch@enst.fr

Résumé

Dans cet article, nous proposons un panaroma des principales méthodes de fusion numérique. Il s'agit des classiques méthodes probabilistes (en particulier de l'inférence bayésienne), mais aussi de méthodes non probabilistes (théorie des ensembles flous et des possibilités, théorie des fonctions de croyance), apparues plus récemment, mais qui connaissent un essor de plus en plus important. Pour chaque théorie, nous présentons les deux composantes essentielles des systèmes de fusion : la représentation des connaissances et le raisonnement. Pour la première, nous montrons quelles sont les caractéristiques des données imparfaites que chaque théorie permet de modéliser. Pour la seconde, nous décrivons les modes de combinaison et les règles de décision. Nous insistons sur les spécificités de la fusion d'informations spatiales (aussi bien au niveau local qu'à un niveau plus structurel) et sur leur prise en compte dans les modèles probabilistes et non probabilistes.

Mots Clef

Fusion d'informations, représentation des connaissances, décision, fusion numérique, information spatiale.

1 Introduction

La fusion d'informations a beaucoup évolué ces dernières années : dans différents domaines, et en particulier en vision et en robotique, les sources d'informations se sont multipliées, qu'il s'agisse de capteurs, d'informations a priori, de connaissances génériques, etc. Chaque source d'information étant en général imparfaite, il est important d'en combiner plusieurs afin d'avoir une meilleure connaissance du « monde ». La fusion d'informations peut alors se définir comme la combinaison d'informations (souvent imparfaites et hétérogènes) afin d'obtenir une information globale plus complète, de meilleure qualité, et permettant de mieux décider et agir. Parallèlement, les méthodes pour modéliser les connaissances et informations imparfaites et pour les combiner ont connu des développements théoriques importants et leurs champs d'applications se sont étendus. Ces méthodes, souvent issues des théories de la décision, de l'incertain et de l'intelligence artificielle, peuvent être soit numériques, soit symboliques. L'ampleur que prend la fusion d'informations suit celle que prennent les technologies et le traitement de l'information en général.

Dans cet article, nous proposons un panaroma des principales méthodes de fusion numérique. Il s'agit des classiques méthodes probabilistes (en particulier de l'inférence bayésienne), mais aussi de méthodes non probabilistes (théorie des ensembles flous et des possibilités, théorie des fonctions de croyance), apparues plus récemment, mais qui connaissent un essor de plus en plus important. Pour chaque théorie, nous présenterons les deux composantes essentielles des systèmes de fusion : la représentation des connaissances et le raisonnement. Pour la première, nous montrerons quelles sont les caractéristiques des données imparfaites que chaque théorie permet de modéliser. Pour la seconde, nous décrirons les modes de combinaison et les règles de décision. Nous insisterons sur les spécificités de la fusion d'informations spatiales (aussi bien au niveau local qu'à un niveau plus structurel) et sur leur prise en compte dans les modèles probabilistes et non probabilistes. La partie 2 présente une définition générale, les caractéristiques des données à prendre en compte dans un système de fusion, ainsi que les principales étapes. Les principales approches numériques sont ensuite exposées, dans les parties 3 pour les approches probabilistes et bayésiennes, 4 pour la théorie des fonctions de croyance, et 5 pour les méthodes floues et possibilistes. Enfin dans la partie 6, nous discutons du traitement de l'information spatiale en fusion. Les éléments présentés dans cet article sont discutés de manière plus détaillée dans [13, 16, 54].

2 Généralités et définitions

2.1 Définitions

Nous adoptons dans cet article 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 tout au long de cet article est la suivante :

La fusion d'informations consiste à combiner des informations hétérogènes issues de plusieurs sources afin d'améliorer la prise de décision.

Cette définition est suffisamment générale pour englober la diversité des problèmes de fusion que l'on rencontre en particulier en traitement de l'information spatiale. 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 défi-

nition 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 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.

2.2 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.

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. Une attention particulière doit être portée à l'échelle utilisée pour représenter les informations, afin de garantir une commensurabilité.

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.

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 : données fréquentes ou rares, informations factuelles ou génériques, génériques ou spécifiques, etc.

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, 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, d'une manière qui convient bien au problème de la fusion, mais qui n'est sûrement pas universelle. L'**incertitude** est relative à la vérité d'une information, et caractérise son degré de conformité à la réalité [28]. Elle fait référence à la nature de l'objet ou du fait concerné, à sa qualité, à son essence ou à son occurrence. L'**imprécision** concerne le contenu de l'information et mesure donc sur défaut quantitatif de connaissance, sur une mesure [28]. L'**incomplétude** caractérise l'absence d'information apportée par la source sur certains aspects du problème. L'**ambiguïté** exprime la capacité d'une information de conduire à deux interprétations. Elle peut provenir des imperfections précédentes. 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. 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. 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 probabilistes permettant de modéliser et mesurer l'information, la redondance et la complémentarité seront décrits dans la partie 3.1.

2.3 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 la partie 2. Considérons un problème général de fusion pour lequel on dispose de l sources $S_1, S_2, ..., S_l$, et pour lequel le but est de prendre une décision dans un ensemble de n décisions possibles $d_1, d_2, ..., d_n$. 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 four-

nisse une information représentée par M_i^j sur la décision d_i. La forme de M_i^j dépend bien sûr du formalisme choisi.

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.

La manière dont ces étapes sont agencées définit le système de fusion et son architecture. En particulier on distingue les systèmes décentralisés dans lesquels des décisions locales sont prises au niveau de chaque source séparément puis sont fusionnées en une décision globale, et les systèmes centralisés dans lesquels on combine par une opération F tous les M_i^j relatifs à la même décision d_i, pour obtenir une forme fusionnée $M_i = F(M_i^1, M_i^2, ..., M_i^l)$, puis une décision est prise sur le résultat de cette combinaison. Ces deux types de systèmes ont des propriétés différentes sur lesquelles nous ne revenons pas ici. Nous nous restreignons au deuxième type dans ce qui suit.

3 Fusion probabiliste et bayésienne

3.1 Mesures d'information

Lorsqu'on dispose d'un ensemble de l sources S_j, une première tâche consiste souvent à le transformer en un sous-ensemble plus réduit, donc de traitement plus simple, sans perdre d'information. Pour exprimer l'apport d'information dû à l'ajout d'une nouvelle source S_{l+1} à un ensemble déjà connu $\{S_1, ..., S_l\}$, les notions d'information et d'entropie [38, 41] (entropie jointe et entropie conditionnelle) sont bien adaptées. On définit ainsi la redondance entre deux sources par :

$$R(S_1, S_2) = H(S_1) + H(S_2) - H(S_1, S_2), \qquad (1)$$

et la complémentarité de la source S_2 par rapport à la source S_1, c'est-à-dire la quantité moyenne d'information qu'il faut ajouter à S_2 pour retrouver S_1 :

$$C(S_1 \mid S_2) = H(S_1 \mid S_2). \qquad (2)$$

Des approches analogues peuvent être envisagées dans un cadre non probabiliste, en s'appuyant par exemple sur l'entropie floue [40]. Le formalisme est pour l'instant moins développé dans cette direction.
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 mesures en prenant en compte des probabilités d'occurrence de configurations particulières de pixels, soit dans le cadre de classifications [42], soit dans celui de champs markoviens [57, 56], ce qui peut s'étendre à la fusion d'informations spatiales dans un cadre plus général.
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 dont 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.

3.2 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 [31]. L'information y est modélisée par une probabilité conditionnelle. Ainsi, la mesure introduite dans la partie 2.3 s'écrit-elle sous la forme :

$$M_i^j = p(d_i \mid S_j). \qquad (3)$$

Cette probabilité est calculée à partir de caractéristiques de l'information extraites des sources disponibles. L'apprentissage des distributions s'appuie sur des outils statistiques classiques. C'est en général $p(S_j \mid d_i)$ qui peut être estimée, et on en déduit la probabilité de l'équation 3 par application de la règle de Bayes.
L'avantage essentiel des méthodes probabiblites 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 que l'apprentissage des modèles. Elles proposent également des règles d'usage soit théoriques soit heuristiques. Enfin la modélisation probabiliste, soutenue par l'interprétation fréquenciste 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.
Mais les méthodes probabilistes sont également l'objet de critiques. 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, ou, si le problème à traiter est complexe, conduire pratiquement à des incohérences car l'utilisateur ne peut alors prendre en compte tout le réseau des dépendances probabilistes. L'apprentissage de lois de probabilités nécessite, outre les hypothèses, un nombre de données important.

3.3 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 on calcule alors des probabilités de la forme :

$$p(d_i \mid S_1, ..., S_l), \tag{4}$$

à l'aide de la règle de Bayes :

$$p(d_i \mid S_1, ..., S_l) = \frac{p(S_1, ...S_l \mid d_i)\ p(d_i)}{p(S_1, ..., S_l)}, \tag{5}$$

où les différents termes sont estimés par apprentissage ;

– soit par la règle de Bayes elle-même, où l'information issue d'une source vient mettre à jour l'information estimée d'après les sources précédentes :

$$p(d_i \mid S_1, ..., S_l) = \frac{p(S_1 \mid d_i)\ p(S_2 \mid d_i, S_1)\ ...\ p(S_l \mid d_i, S_1, ..., S_{l-1})\ p(d_i)}{p(S_1)\ p(S_2 \mid S_1)\ ...\ p(S_l \mid S_1, ..., S_{l-1})}. \tag{6}$$

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 :

$$p(d_i \mid S_1, ..., S_l) = \frac{\prod_{j=1}^{l} p(S_j \mid d_i)\ p(d_i)}{p(S_1, ..., S_l)}. \tag{7}$$

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.

Malgré les avantages de cette combinaision, elle 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. Elle nécessite de plus l'estimation des probabilités a priori $p(d_i)$, qui est souvent délicate et est primordiale dans les cas où l'on a peu d'informations. La forme conjonctive de la fusion bayésienne conduit souvent en pratique à un effondrement des probabilités des événements qui sont déduits d'une longue chaîne de déduction. Enfin, elle ne permet pas de modéliser l'ignorance pour la prendre en compte dans la combinaison.

3.4 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 [21, 32]. 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 [44], telles que le krigeage ou le krigeage universel, pourraient également être utilisées dans ce cadre.

3.5 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 :

$$d_i \text{ si } p(d_i \mid S_1, ..., S_l) = \max_k p(d_k \mid S_1, ..., S_l), \tag{8}$$

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 démuni devant la justification d'un choix et l'éloigne de l'objectivité recherchée initialement par ces méthodes.

4 Fusion dans la théorie des fonctions de croyance

La théorie des fonctions de croyance (ou théorie de Dempster-Shafer) date des années 70. Nous montrons dans cette partie 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.

4.1 Modélisation

La théorie des fonctions de croyance permet, de manière analogue à la théorie des possibilités comme nous le verrons dans la partie suivante, 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 [50, 35, 51]. Les fonctions de masse sont définies sur tous les sous-ensembles de l'espace de discernement D (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 $D = \{d_1, d_2, ..., d_n\}$ où chaque d_i désigne une hypothèse en faveur de laquelle une décision peut être prise (typiquement une classe dans un problème de classification multi-sources). Une fonction de masse est définie comme une fonction de 2^D (ensemble des parties de D) dans $[0, 1]$. En général on impose $m(\emptyset) = 0$, et une normalisation de la forme :

$$\sum_{A \subset D} m(A) = 1, \tag{9}$$

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.
Une fonction de croyance Bel est une fonction totalement croissante définie de 2^D dans $[0,1]$:

$$\forall A_1...A_k,\ Bel(\cup_{i=1...k}A_i) \geq \sum_{I\subset\{1...k\}, I\neq\emptyset} (-1)^{|I|+1} Bel(\cap_{i\in I}A_i), \tag{10}$$

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 :

$$\forall A \in 2^D,\ Bel(A) = \sum_{B\subset A, B\neq\emptyset} m(B) \tag{11}$$

est une fonction de croyance. Inversement, à partir d'une fonction de croyance définie comme une fonction totalement croissante (inégalité 10) telle que $Bel(\emptyset) = 0$, $Bel(D) = 1$, on peut définir une fonction de masse par :

$$\forall A \in 2^D,\ m(A) = \sum_{B\subset A} (-1)^{|A-B|} Bel(B). \tag{12}$$

Cette fonction de masse vérifie alors l'équation 11. La fonction de croyance mesure la confiance totale que l'on a dans un sous-ensemble A.
Une fonction de plausibilité Pls est également une fonction de 2^D dans $[0,1]$ définie par :

$$\forall A \in 2^D,\ Pls(A) = \sum_{B\cap A\neq\emptyset} m(B) = 1 - Bel(A^C). \tag{13}$$

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 [51] 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). L'intervalle $[Bel(A), Pls(A)]$ est appelée 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. Dans les cas plus complexes, ce n'est pas le cas et il n'y a pas d'équivalence directe avec des probabilités. L'analogue des fonctions de crédibilité et de plausibilité pourrait être obtenue 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.
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, de l'incomplétude, de l'imprécision et de l'incertitude, en analysant les capacités de chaque source à donner une information sur chaque décision possible.

4.2 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 [33, 39]. La manière la plus simple que l'on puisse imaginer consiste à calculer les masses sur les singletons dans une source S_j par $m_j(\{d_i\}) = M_i^j$, où M_i^j 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 l'ensemble des hypothèses composées, ou n'utilisent que certaines hypothèses composées, dans une démarche simplificatrice et souvent très heuristique. Mais d'autres approches peuvent également être envisagées. Des modifications des modèles probabilistes ont été proposées, introduisant une masse sur D par affaiblissement [50], en prenant en compte l'information sur tout ce qui n'est pas d_i [1], etc. Des approches fondée sur des distances à des prototypes ont également été proposées [23], s'inspirant de ce qui est fait en reconnaissance des formes. 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 [6, 55, 47]. Des méthodes d'apprentissage des éléments focaux ont également été proposées, à partir des intersections entre les classes détectées par les différentes sources individuellement [43] ou de de mesures de dissonance et consonance [45], ou d'opérations de morphologie mathématique [7].

4.3 Combinaison conjonctive

Soit m_j ($j = 1...l$) la fonction de masse définie pour la source j. La combinaison conjonctive des fonctions de masse est effectuée selon la règle orthogonale de Dempster [50] :

$$\forall A \subset D,\ (m_1 \oplus ... \oplus m_l)(A) = \sum_{B_1\cap...\cap B_l = A} m_1(B_1)...m_l(B_l). \tag{14}$$

Des justifications axiomatiques de cette règle peuvent être trouvées dans [51]. Les différences entre ces axiomes et ceux de Cox [22] (qui permettent de justifier les règles des

probabilités) expliquent les origines des différences entre les deux théories [4].
Dans cette équation 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. 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 [48, 47].
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 en le divisant par $1 - \sum_{B_1 \cap ... \cap B_l = \emptyset} m_1(B_1) m_2(B_2) ... m_l(B_l)$ (si cette quantité est non nulle) et en forçant la masse de l'ensemble vide à 0. Il est important de prendre en compte la valeur du conflit 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. La forme non normalisée est donc le plus souvent préférable en cas de conflit.
La règle de combinaison est commutative et associative. La fonction de masse définie par $m_0(D) = 1$ et $\forall A \subset D, A \neq D, m_0(A) = 0$ est élément neutre pour la combinaison. Cette masse représente une source complètement non informative, qui ne distingue aucun élément de D. La définition de cette fonction de masse remplace le principe d'indifférence utilisé en probabilités, et représente mieux l'absence d'information. Notons que la loi $\oplus$ n'est pas idempotente. La règle de Demspter 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 $[Bel, Pls]$. Dans le cas général, ainsi que le montre la formule 14, 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 [2], c'est-à-dire où les éléments focaux de chaque source sont uniquement les singletons et les compléments des singletons.

D'autres modes de combinaison, tels que des modes disjonctifs ou de compromis, sont possibles, en remplaçant l'intersection dans la formule 14 par une autre opération ensembliste. Par exemple, une fusion disjonctive est obtenue en prenant la réunion [53]. 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, 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.

4.4 Décision

Une fois calculées les fonctions de masse combinées, les fonctions de croyance et de plausibilité sont déduites par les équations 11 et 13. La dernière étape est celle de la décision, donc du choix d'un sous-ensemble de D maximisant un certain critère.
Dans la théorie des fonctions de croyances, plusieurs règles de décision sont possibles, telle que le maximum de plausibilité, de croyance, de probabilité pignistique [52], auxquelles on peut associer des critères de rejet, de coût [23], de préservation éventuelle des ambiguïtés, etc.

5 Fusion floue et possibiliste

5.1 Modélisation

La théorie des ensembles flous fournit un très bon outil pour représenter explicitement des informations imprécises, sous la forme de fonctions d'appartenance [60]. La mesure M_i^j prend alors la forme $M_i^j = \mu_i^j$, où μ_i^j désigne par exemple le degré d'appartenance d'un élément à la classe d_i selon la source S_j, ou la traduction d'une information symbolique exprimée par une variable linguistique. 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 démuni 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 [61, 28], dérivée des ensembles flous, permet de représenter à la fois l'imprécision et l'incertitude, par l'intermédiaire de distributions de possibilités π et de deux fonctions caractérisant les événements : la possibilité Π et la nécessité N. Dans le cadre de la fusion numérique, une application possible de cette théorie consiste à définir π sur D (l'ensemble des décisions possibles) et la mesure M_i^j par $M_i^j = \pi_j(d_i)$, c'est-à-dire comme le degré de possibi-

lité de la décision d_i, selon la source S_j. Dans un problème de classification, cette modélisation suppose que les classes (ou décisions) sont nettes, alors que le modèle flou suppose que les classes sont floues.

5.2 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 [3]. Des informations spatiales peuvent être prises en compte facilement grâce au développement des techniques floues dans ce domaine, en particulier en traitement d'images [9]. 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 [26, 25, 37]. D'autres méthodes cherchent à estimer directement les fonctions d'appartenance à partir de l'histogramme, afin d'optimiser des critères d'entropie [19], ou de minimum de spécificité et de cohérence [20]. 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 [11] 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.

5.3 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.

Dans la théorie des ensembles flous et des possibilités, de multiples modes de combinaison sont possibles [27, 59]. Parmi les principaux opérateurs on trouve en particulier les t-normes, les t-conormes [46, 49], les moyennes [58, 34], les sommes symétriques, et des opérateurs prenant en compte des mesures de conflit ou encore de fiabilité des sources [29, 24]. Nous n'en rappelons pas les définitions ici, faute de place.

Le choix d'un opérateur peut se faire selon plusieurs critères [5]. 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. Soit 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 [5] 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 :

1. 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 ;
2. aux opérateurs autonomes à comportement variable (ACV) : le comportement dépend des valeurs numériques des informations à fusionner ;
3. 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. Cette classification, qui regroupe tous les opérateurs classiquement utilisés, constitue un premier critère de choix d'un opérateur pour une application spécifique.

Un deuxième critère est donné par les propriétés des opérateurs et leur interprétation en termes de fusion de données incertaines, imprécises, incomplètes ou encore ambiguës.

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 (la t-norme et la t-conorme de Lukasiewicz par exemple) 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 lesquels 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 ordinales sont de plus imposées si on veut qu'elles soient invariantes par une transformation croissante des degrés d'appartenance [30].

5.4 Décision

La règle principalement utilisée en fusion floue est le maximum des degrés d'appartenance :

$$d_i \text{ si } \mu_i(x) = \max\{\mu_k(x),\ 1 \leq k \leq n\}, \qquad (15)$$

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 partie 6).

6 Introduction d'informations spatiales dans la fusion

L'information spatiale est fondamentale en traitement d'images, en robotique, ou plus généralement en traitement et fusion d'informations spatiales. 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 est de garantir que la décision soit spatialement cohérente. Par exemple, en classification multi-sources, on va chercher à éviter les points isolés dans une classe différente.

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 des points, l'information spatiale doit être introduite explicitement. Le contexte spatial considéré est le plus souvent le voisinage local de chaque point. 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 entre les sources ou du positionnement d'un robot), 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 cela [15, 17, 18]. 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 [10] et elles peuvent avantageusement être prises en compte dans la fusion, comme source d'informations supplémentaire [14, 12]. 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.

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. Des règles de type vote majoritaire sont très souvent utilisées pour cela.

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 point par exemple, de nombreux exemples peuvent être trouvés dans la littérature. 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). 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 [8]. Autre exemple, dans [36], 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, encore peu de travaux existent dans ce domaine, qui mérite certainement d'être développé.

7 Conclusion

En fusion numérique, les efforts des dernières années ont permis d'aboutir à une meilleure compréhension des différentes théories. On sait ainsi maintenant quels sont les bons cadres d'application de ces théories, leurs atouts et limites. De nombreuses applications ont vues le jour, entre autres en robotique. L'introduction d'informations spatiales dans la fusion est un point important, pour lequel l'ensemble des méthodes existantes pourrait être encore étoffé. Des travaux sur la combinaison de méthodes sont également 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.

Remerciements

Plusieurs éléments de discussion, en particulier dans la partie 2, sont issus de réunions de deux groupes de travail : European Working Group Fusion, et le GT fusion du GDR ISIS. L'auteur tient à remercier les participants à ces groupes, ainsi que Henri Maître qui a initié nos recherches et y a fortement contribué, et Patrick Rives pour son invitation.

Références

[1] A. Appriou. Formulation et traitement de l'incertain en analyse multi-senseurs. In *Quatorzième Colloque GRETSI*, pages 951–954, Juan les Pins, 1993.

[2] J. A. Barnett. Computational Methods for a Mathematical Theory of Evidence. In *Proc. of 7th IJCAI*, pages 868–875, Vancouver, 1981.

[3] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum, New-York, 1981.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] I. Bloch. Traitement d'images. In B. Bouchon-Meunier and C. Marsala, editors, *Traitement de données complexes et commande en logique floue*, chapter 3, pages 95–152. Hermes, Paris, France, 2003.

[10] I. Bloch. Fuzzy Spatial Relationships for Image Processing and Interpretation : A Review. *Image and Vision Computing*, 23(2) :89–110, 2005.

[11] 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.

[12] I. Bloch, O. Colliot, O. Camara, and T. Géraud. Fusion of Spatial Relationships for Guiding Recognition. Example of Brain Structure Recognition in 3D MRI. *Pattern Recognition Letters*, 26 :449–457, 2005.

[13] I. Bloch and A. Hunter (Eds). Fusion : General Concepts and Characteristics. *International Journal of Intelligent Systems*, 16(10) :1107–1134, oct 2001.

[14] I. Bloch, T. Géraud, and H. Maître. Representation and Fusion of Heterogeneous Fuzzy Information in the 3D Space for Model-Based Structural Recognition - Application to 3D Brain Imaging. *Artificial Intelligence Journal*, 148 :141–175, 2003.

[15] I. Bloch and H. Maître. Fuzzy Mathematical Morphologies : A Comparative Study. *Pattern Recognition*, 28(9) :1341–1387, 1995.

[16] I. Bloch and H. Maître. Fusion d'informations en traitement d'images : spécificités, modélisation et combinaison par des méthodes numériques. *Techniques de l'Ingénieur*, TE 5 230 :1–26, 2002.

[17] I. Bloch, C. Pellot, F. Sureda, and A. Herment. Fuzzy Modelling and Fuzzy Mathematical Morphology applied to 3D Reconstruction of Blood Vessels by Multi-Modality Data Fusion. In D. Dubois R. Yager and H. Prade, editors, *Fuzzy Set Methods in Information Engineering : A Guided Tour of Applications*, chapter 5, pages 93–110. John Wiley & Sons, New-York, 1996.

[18] 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.

[19] 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.

[20] M. R. Civanlar and H. J. Trussel. Constructing Membership Functions using Statistical Data. *Fuzzy Sets and Systems*, 18 :1–13, 1986.

[21] R. Cooke. *Experts in Uncertainty*. Oxford University Press, Oxford, UK, 1991.

[22] R. T. Cox. Probability, Frequency and Reasonable Expectation. *Journal of Physics*, 14(1) :115–137, 1946.

[23] 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.

[24] 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.

[25] B. Bharathi Devi and V. V. S. Sarma. Estimation of Fuzzy Memberships from Histograms. *Information Sciences*, 35 :43–59, 1985.

[26] 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.

[27] D. Dubois and H. Prade. A Review of Fuzzy Set Aggregation Connectives. *Information Sciences*, 36 :85–121, 1985.

[28] D. Dubois and H. Prade. *Possibility Theory*. Plenum Press, New-York, 1988.

[29] 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.

[30] 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.

[31] R. Duda and P. Hart. *Pattern Classification and Scene Analysis*. Wiley, New-York, 1973.

[32] S. French. Group Consensus Probability Distributions : A Critical Survey. In J. Bernardo et al., editor, *Bayesian Statistics*, pages 183–201. Elsevier, The Netherlands, 1985.

[33] 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.

[34] M. Grabisch. Fuzzy Integral in Multicriteria Decision Making. *Fuzzy Sets and Systems*, 69 :279–298, 1995.

[35] J. Guan and D. A. Bell. *Evidence Theory and its Applications*. North-Holland, Amsterdam, 1991.

[36] S. Le Hégarat-Mascle, I. Bloch, and D. Vidal-Madjar. Introduction of Neighborhood Information in Evidenve Theory and Application to Data Fusion of Radar and Optical Images with Partial Cloud Cover. *Pattern Recognition*, 31(11) :1811–1823, 1998.

[37] G. J. Klir and B. Parviz. Probability-Possibility Transformations : A Comparison. *Int. J. General Systems*, 21 :291–310, 1992.

[38] S. Kullback. *Information Theory and Statistics*. Wiley, New York, 1959.

[39] J. D. Lowrance, T. M. Strat, L. P. Wesley, T. D. Garvey, E. H. Ruspini, and D. E. Wilkins. The Theory, Implementation and Practice of Evidential Reasoning. SRI project 5701 final report, SRI, Palo Alto, June 1991.

[40] 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.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] K. Menger. Statistical Metrics. *Proc. National Academy of Siences USA*, 28 :535–537, 1942.

[47] N. Milisavljevic and I. Bloch. Sensor Fusion in Anti-Personnel Mine Detection Using a Two-Level Belief Function Model. *IEEE Transactions on Systems, Man and Cybernetics*, 33(2) :269–283, 2003.

[48] J. Schubert. On Nonspecific Evidence. *International Journal of Intelligent Systems*, 8 :711–725, 1993.

[49] B. Schweizer and A. Sklar. *Probabilistic Metric Spaces*. North Holland, Amsterdam, 1983.

[50] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[51] P. Smets. The Combination of Evidence in the Transferable Belief Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5) :447–458, 1990.

[52] P. Smets. Constructing the Pignistic Probability Function in a Context of Uncertainty. *Uncertainty in Artificial Intelligence*, 5 :29–39, 1990.

[53] P. Smets. Belief Functions : The Disjunctive Rule of Combination and the Generalized Bayesian Theorem. *International Journal of Approximate Reasoning*, 9 :1–35, 1993.

[54] I. Bloch (sous la direction de). *Fusion d'informations en traitement du signal et des images*. Hermès, Paris, France, 2003.

[55] 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.

[56] 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.

[57] E. Volden, G. Giraudon, and M. Berthod. Modeling Image Redundancy. In *IGARSS'95*, volume 3, pages 2148–2150, Firenze, Italy, 1995.

[58] 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.

[59] R. R. Yager. Connectives and Quantifiers in Fuzzy Sets. *Fuzzy Sets and Systems*, 40 :39–75, 1991.

[60] L. A. Zadeh. Fuzzy Sets. *Information and Control*, 8 :338–353, 1965.

[61] L. A. Zadeh. Fuzzy Sets as a Basis for a Theory of Possibility. *Fuzzy Sets and Systems*, 1 :3–28, 1978.