

Fusion de données

Fusion de données en traitement d'images : modèles d'information et décisions

Data Fusion in Image Processing : Information Models and Decisions

par **I. BLOCH, H. MAITRE**

*Ecole Nationale Supérieure des Télécommunications, département Images, 46 rue Barrault, 75634 Paris Cedex 13
Tél : 45 81 75 85, Fax : 45 81 37 94, E-mail : bloch@ima.enst.fr*

Résumé

Nous présentons les définitions et les étapes nécessaires à la fusion de données en traitement des images. Nous introduisons ensuite les divers outils utilisés à ce jour pour modéliser les informations issues de divers capteurs, combiner ces informations puis décider, en nous focalisant sur les théories probabilistes et bayésiennes, des croyances, des ensembles flous et des possibilités. Nous insistons sur les points centraux de la fusion d'images : l'incertain et l'imprécis. Nous montrons comment la prise en compte de ces deux caractéristiques intrinsèques du traitement de l'image induit des choix dans les modèles et les opérateurs de fusion.

Mots clés : Fusion de données, traitement d'image, théorie de la décision, théorie de l'information, décision bayésienne, ensembles flous, possibilités, croyances, imprécision, incertitude.

Abstract

We present the definitions and necessary steps for data fusion in image processing. We introduce then the different tools used for modeling information issued from various sensors, for combining these pieces of information, and then for taking a decision. We focus on probabilistic and Bayesian inference, evidence theory, fuzzy sets and possibility theory. We underline the key features of data fusion in image processing : uncertainty and imprecision. We show how the management of these two characteristics, intrinsic to image processing, induces choices of models and fusion operators.

Key words : Data fusion, image processing, decision theory, information theory, Bayesian decision, fuzzy sets, possibility theory, evidence theory, imprecision, uncertainty.

1. Introduction

La fusion de données a fait irruption dans le domaine du traitement d'images il y a cinq ans environ et retient depuis l'attention de nombreuses équipes dans deux domaines principalement, l'imagerie satellitaire et aérienne d'une part, l'imagerie médicale d'autre part. On la considère cependant comme potentiellement utilisable dans de nombreux autres champs : le contrôle non destructif, la vision des véhicules autonomes ou la surveillance des sites sensibles.

La fusion d'images hérite naturellement des techniques et méthodes développées depuis plus de 20 ans dans le champ du contrôle des processus complexes (usine, réacteur chimique, écosystème) et qui utilisaient abondamment la fusion de données issues de capteurs très hétérogènes (mesures de pression, de débit, de température, etc.). Mais, par la spécificité des données d'images (leur

taille, leur structure, leur faible cadence), par l'absence, dans les problèmes d'imagerie, de modélisation fonctionnelle riche comme la balance énergétique d'une réaction chimique ou le bilan thermique d'un écosystème, la fusion d'images n'a que peu puisé dans ce fonds commun d'outils et recherche au contraire ses propres méthodes, moins dans la théorie des systèmes ou dans le contrôle adaptatif que dans les méthodes de la reconnaissance des formes ou de l'intelligence artificielle. Ces méthodes seront largement décrites dans les lignes ci-dessous.

Comme tout domaine en expansion, la fusion d'images souffre de définitions incertaines et de limites floues. L'un des objectifs de cet article est d'apporter quelques éclairages sur les notions de base qui fondent la fusion d'images ainsi que sur les étapes qui jalonnent sa mise en œuvre. Cette présentation est faite avec le parti pris du traicteur d'image qui laisse dans l'ombre des aspects fondamentaux du schéma global d'exploitation de l'image : la conception du système d'ensemble incluant capteurs et centres de décision, les aspects de contrôle et de commande avec les

éventuels retours de la décision vers la mesure, les implications matérielles des choix tant sur les organes de calcul que sur les réseaux de communication, les conséquences de ces choix sur un bilan global du système dans une logique mettant en relation une « qualité de service » (optimalité des décisions, taux d'erreurs, retard à la décision, sensibilité aux défaillances, sensibilité aux leures, etc.) et un coût global de la chaîne.

1.1. QU'EST-CE QUE LA FUSION D'IMAGES ?

Aux diverses définitions déjà proposées, nous préférons la suivante : *la fusion d'images est l'utilisation conjointe d'images hétérogènes pour l'aide à la prise de décision.*

Il n'y a fusion que s'il y a utilisation conjointe d'images. Leur traitement, séquentiel ou parallèle, conduisant à des décisions séparées, ne relève pas de la fusion.

Il y a fusion entre images de natures différentes seulement, c'est-à-dire pour lesquelles on ne sait définir une métrique pertinente qui permettrait naturellement de ramener le problème d'analyse dans celui classique des espaces vectoriels. Ainsi traiter des images de même source comme on le pratique en stéréovision, en analyse de séquences ou en suivi multitemporel ne relève-t-il pas de la fusion. Traiter des images multispectrales ou a fortiori des images colorées apparaît ainsi à la limite du champ de la fusion chaque fois qu'un changement de base judicieux permet de ramener les variables scalaires dans un même espace vectoriel doté d'une métrique pertinente (par exemple une distance colorimétrique pour l'observateur standard CIE).

Enfin la fusion doit conduire à une décision. Si dans le cas le plus simple la décision est prise par un expert humain (photo-interprète ou praticien hospitalier), alors le rôle de la fusion est de présenter à celui-ci les seules informations pertinentes issues des diverses images et de les combiner de façon judicieuse (c'est de façon très classique des combinaisons colorées RVB ou IHS qui jouent alors ce rôle élémentaire). Mais nous pensons que l'objectif de la fusion est clairement d'apporter à l'avenir une aide moins « cosmétique » même à un expert humain.

1.2. POURQUOI LA FUSION D'IMAGES ?

La fusion d'images n'est pas un avatar des modes qui secouent le traitement de l'information depuis 20 ans, c'est la réponse naturelle des utilisateurs à la profusion des images produites par les systèmes contemporains d'investigation, que ce soit en imagerie aérienne et satellitaire ou en génie biologique et médical. Prenons deux exemples.

L'observation de la terre par satellites croît selon deux tendances : la première nous donne des images dont la résolution est de plus en plus fine et la cadence de plus en plus élevée; la seconde propose des signaux de plus en plus variés : à la profusion des canaux visibles spectrométriques répond la variété des bandes

radar (L, C, X, P avec leurs diverses polarisations) et des bandes infra-rouges. S'ajoutent, lors de l'exploitation de ces données, les informations collectées au sol et disponibles à travers les systèmes d'information géographique (cartes, cadastres, plans d'occupation des sols, relevés géologiques ou agronomiques, modèles altimétriques, etc.). Les décisions du photo-interprète relèvent de l'ensemble de ces sources. Son expertise, acquise par de longues études relayées par de multiples expériences, permet d'associer de façon très efficace l'indice détecté dans une image et sa confirmation dans une seconde s'il sait « lire » cette seconde et « rapprocher » les deux événements. La prolifération des images rend ces deux tâches de plus en plus difficiles par la création d'une combinatoire galopante, et par l'impossibilité d'acquérir une expertise réelle dans toutes les modalités disponibles.

Une situation assez semblable apparaît en imagerie médicale. Les techniques d'imagerie se diversifient : rayons X, résonance magnétique, imagerie nucléaire, imagerie ultrasonore, chacune pouvant se décliner selon des modalités diverses en fonction des protocoles d'acquisition. Selon une pratique ancienne, le milieu médical confie chaque type d'image à un expert qui porte un diagnostic partiel sur la modalité de sa spécialité, puis les spécialistes échangent leurs expériences et de cette confrontation naît le diagnostic final. Le souci de regrouper toutes les sources d'images sur une même console conduit à l'introduction progressive des systèmes d'archivage et de consultation intégrés à l'hôpital (les PACS des anglo-saxons¹). Mais sans les outils d'aide à la décision associés à ces systèmes, ces autoroutes informatiques et ces bases de données spécialisées demeurent encore d'une pratique modeste, tant l'exploitation de nombreuses images variées est une tâche difficile pour un seul expert.

Nous voyons sur ces deux exemples les raisons profondes de la nécessité de la fusion de données : apporter des outils d'aide à la décision capables d'intégrer, au fur et à mesure de leur mise au point, des modes d'imagerie nouveaux, en préparant et en raccourcissant l'étape d'élaboration de l'expertise humaine, gérer une profusion grandissante d'images sans sacrifier les apports potentiels d'une combinatoire complexe, c'est-à-dire extraire de chaque image toute l'information qu'elle peut apporter au-delà même de ce que les experts humains savent aujourd'hui en tirer².

1.3. RECALAGE

La fusion d'images apparaît comme une tâche complexe constituée de nombreuses étapes dont plusieurs sont aujourd'hui mal maîtrisées. Nous les verrons dans les parties suivantes.

1. Picture Archiving and Communication Systems (systèmes d'archivage et d'échange d'images)

2. Éventuellement, et s'il apparaissait que le diagnostic humain aujourd'hui fondé sur un petit nombre (deux ou au maximum trois images bien choisies) était un optimum non franchissable, la fusion de données devrait permettre d'établir clairement cette optimalité et proposer de façon raisonnée et transmissible les règles du choix des modalités optimales et de leurs associations.

Mais classiquement, elle commence par une étape de recalage des images dont l'objectif est de superposer exactement les pixels correspondant à un même objet observé dans les diverses modalités. Cette étape peut être facilitée s'il existe un référentiel absolu reconnu pour décrire la scène (par exemple, un repère géocodé en imagerie aérienne). Les techniques de recalage sont nombreuses et utilisent des principes variés (désignation interactive de points-amers, corrélation, recherche de points d'appui, programmation dynamique, isomorphismes de graphes, minimisation de distances entre primitives de plus haut niveau, etc.) (on pourra consulter [MAÏT-91] et [MANG-94] pour une synthèse des méthodes en imagerie aérienne et satellitaire et en imagerie médicale respectivement). Elles sont bien maîtrisées dans de nombreux cas, mais il demeure des domaines reconnus où le recalage automatique est peu fiable. C'est le cas lorsque les capteurs sont très hétérogènes et que les déformations géométriques qui affectent les images sont complexes.

Si le recalage géométrique des images est universellement pratiqué aujourd'hui avant toute étape de fusion, il n'est pas clair que ce soit une étape nécessaire et même souhaitable. En fait, l'utilisateur n'a jamais besoin que de connaître l'application d'une image dans l'autre; mettre en œuvre cette application au moyen d'un échantillonnage et de filtrage dégrade généralement les signaux sans autre bénéfice que d'autoriser un contrôle visuel de la superposition des images. Il est probable que cette transformation sera de moins en moins pratiquée dans les années à venir.

Ce recalage, qu'on l'effectue ou non, nous fait pénétrer immédiatement dans la problématique de la fusion d'images car il s'accompagne toujours d'incertitude et d'imprécision (sur les paramètres de la calibration, sur la localisation des points de contrôle). Cette incertitude et cette imprécision doivent être intégrées dans le raisonnement qui conduira ensuite à la décision.

Si l'étape de recalage semble donc généralement échapper à la fusion d'images (puisqu'il peut être fait auparavant de façon totalement indépendante), il demeure des circonstances où ce recalage doit être conçu dans la phase de fusion, c'est en particulier le cas du recalage symbolique où l'on associe à une image des descriptions de haut niveau des objets qui la composent : une ville dans l'association carte - image satellitaire, ou une aire fonctionnelle du cerveau dans l'association atlas - image de résonance magnétique. Dans cette situation, on ne conçoit pas aujourd'hui de mise en correspondance pixel par pixel, mais une association entre représentations de niveau élevé qui passe alors par les étapes de modélisation, puis de décision qui sont les composants de la fusion.

Dans la suite de cet article, nous supposerons que le recalage a été mené à son terme et qu'il existe donc l images de la scène observée, dans un même référentiel.

2. Modèles de fusion

Considérons un problème général de fusion d'images, où l'on dispose de l images I_j représentant des données hétérogènes. Soit x l'élément sur lequel porte la décision. Ce peut être un pixel de l'espace couvert par les images ou tout autre objet plus complexe extrait des images. La décision consiste à lui affecter un élément C_i d'un espace de décision $D = \{C_1, \dots, C_n\}$. Typiquement, les C_i , représentant les décisions possibles, peuvent être des classes recouvrant les images et auxquelles on souhaite affecter les pixels, ou encore des objets sémantiques auxquels on veut attribuer les points pour les reconnaître. La décision sur x est prise à partir d'informations $f_j(x)$ données par chacune des images I_j (par exemple le niveau de gris de x dans I_j ou des primitives issues de traitements plus complexes) et sur lesquelles s'appuie la fusion, souvent par l'intermédiaire de « mesures » $M_i^j(x)$, reliant des informations sur x extraites de I_j à une décision potentielle C_i sur x (voir table 1). Ces mesures ont des interprétations différentes et prennent des formes mathématiques variées selon les cadres théoriques employés pour cette fusion (voir partie 4).

Tableau 1. – Informations à fusionner : mesures issues des l images relatives aux n classes.

	I_1	I_2	...	I_l
C_1	M_1^1	M_1^2	...	M_1^l
C_2	M_2^1	M_2^2	...	M_2^l
...
C_n	M_n^1	M_n^2	...	M_n^l

Dans le cas idéal, la décision est prise à partir de la matrice des $M_i^j(x)$ complète, pour toutes les images I_j et toutes les classes C_i . 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 [MAER-93].

Un deuxième modèle consiste ainsi à prendre d'abord des décisions locales, sur chaque image séparément. Dans ce cas, une décision $d(j) \in D$ est prise pour I_j à partir de toutes les mesures faites sur cette image seulement (sur les colonnes du tableau 1). Cette décision qui peut être prise localement par chaque capteur est dite décentralisée. Puis, dans une deuxième étape, ces décisions sont fusionnées en une décision globale sur x . Ce modèle s'impose lorsque les images ne sont pas disponibles simultanément. Il permet des réponses rapides, grâce à des procédures spécifiques à chaque capteur, et peut être très facilement adapté à l'introduction de capteurs supplémentaires. Ce type de modèle s'appuie avec profit sur des techniques de contrôle adaptatif, et utilise souvent des architectures distribuées. Son principal inconvénient provient de ce qu'il tient mal compte des relations entre capteurs et des corrélations ou dépendances possibles entre images. 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 toute l'information sur les mesures de départ est perdue.

Un troisième modèle, « orthogonal » au précédent³, consiste à combiner par une opération F toutes les mesures relatives à la même décision C_j issues de toutes les images (les lignes du tableau 1), pour obtenir une mesure fusionnée $M_i(x) = F(M_i^j(x), 1 \leq j \leq l)$. Puis une décision est prise sur le résultat de cette combinaison. Ici, aucune décision intermédiaire n'est prise, et l'information est manipulée sous la forme de mesures jusqu'à la dernière étape. Ce modèle, comme le modèle global, est un modèle centralisé qui nécessite de disposer simultanément de toutes les images. Plus simple que le modèle global, il est moins souple que le précédent rendant plus difficile l'ajout éventuel de capteurs.

Enfin, un modèle intermédiaire, hybride, consiste à choisir de manière adaptative les informations nécessaires pour un problème donné en fonction des spécificités des images. Ce type de modèle copie souvent l'expert humain et fait intervenir des connaissances symboliques sur les images et sur les objets. Il est donc très utilisé dans les systèmes de fusion à base de règles : ainsi, dans le système MESSIE [CLEM-91], [HOUS-91], l'affectation d'un point à la classe « pont » est décidée par une forte réponse dans l'image radar, conditionnellement à une très basse réponse, caractéristique de l'eau, dans le canal XS3 de SPOT. Des architectures de type multi-agents sont bien adaptées à ce modèle.

Dans cet article, nous étudions essentiellement les modèles du troisième type, en nous restreignant aux méthodes de fusion de type numérique, qui leur sont le mieux adaptées. Les différentes étapes impliquées dans ce modèle, définition des M_i^j , combinaison par un opérateur F et décision, seront décrites dans la partie 4. Auparavant, une étude fine des informations disponibles dans les images et de leurs relations inter-images s'impose souvent pour mieux définir le processus de fusion et vérifier les hypothèses et les conditions d'application. Nous proposons dans la partie 3 des méthodes et outils pour mener à bien de manière rigoureuse cette étape, prenant en compte l'imprécision et l'incertitude des données et des connaissances. Nous nous focalisons sur les méthodes probabilistes et bayésiennes, sur la théorie des ensembles flous et des possibilités et sur la théorie des croyances de Dempster et Shafer, qui sont les méthodes numériques actuellement les plus exploitées pour la fusion d'images, et laissons délibérément de côté les autres [SOMB-89].

3. Le terme « orthogonal » est illustré par le tableau 1 et s'explique de la manière suivante : dans le deuxième modèle, toutes les classes sont prises en compte sur chaque image d'abord, alors que dans le troisième, on combine d'abord toutes les images pour chaque classe.

3. Redondance et complémentarité en termes de théorie de l'information

Lorsqu'on dispose d'un ensemble de l images 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 est souvent employée, qui projette chaque image sur les vecteurs propres de la matrice de corrélation, et permet ainsi d'obtenir l nouvelles images décorrélées, classées par ordre décroissant d'énergie. Une troncature aux l' ($l > l'$) premières images 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 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 image 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 [KULL-59], [MAÏT-93], [MAÏT-94]. À partir de la probabilité jointe des l premières images $p(I_1, \dots, I_l)$ (estimée le plus souvent par dénombrement, par exemple à partir de l'histogramme multidimensionnel des niveaux de gris des images), on définit l'entropie (ou information moyenne par pixel) des l premières images par :

$$H(I_1, \dots, I_l) = - \sum p(I_1, \dots, I_l) \log p(I_1, \dots, I_l), \quad (1)$$

et l'entropie apportée par la $(l + 1)$ ème image s'exprime, soit en fonction des entropies, soit en fonction des probabilités par :

$$\begin{aligned} H(I_{l+1} | I_1, \dots, I_l) &= H(I_1, \dots, I_{l+1}) - H(I_1, \dots, I_l) \\ &= - \sum p(I_1, \dots, I_{l+1}) \log p(I_{l+1} | I_1, \dots, I_l). \end{aligned} \quad (2)$$

Pour deux images, on définit ainsi la redondance⁴ entre elles par :

$$R(I_1, I_2) = H(I_1) + H(I_2) - H(I_1, I_2), \quad (3)$$

et la complémentarité de l'image I_2 par rapport à l'image I_1 , c'est-à-dire la quantité moyenne d'information qu'il faut ajouter à I_2 pour retrouver I_1 :

$$C(I_1 | I_2) = H(I_1 | I_2), \quad (4)$$

ce qui conduit à la relation :

$$H(I_1) = R(I_1, I_2) + C(I_1 | I_2). \quad (5)$$

4. Cette redondance ne peut malheureusement pas être étendue à plus de deux images sans perdre potentiellement sa propriété de positivité.

Des approches analogues peuvent être envisagées dans un cadre non probabiliste, en s'appuyant par exemple sur l'entropie floue [LUCA-72]. Le formalisme est pour l'instant moins développé dans cette direction.

En fusion d'images, on utilisera des images fortement redondantes pour confirmer une décision incertaine et des images complémentaires pour élargir le champ des décisions. Des images 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 mesures en prenant en compte des probabilités d'occurrence de configurations particulières de pixels, soit dans le cadre de classifications [MAÏT-94], soit dans celui de champs markoviens [BERT-94].

4. Imprécision et incertitude en fusion d'images

Imprécision et incertitude sont deux termes souvent abusivement confondus. L'imprécision concerne le contenu de l'information et porte donc sur un défaut quantitatif de connaissance, sur une mesure, tandis que l'incertitude est relative à la vérité d'une information, caractérisant sa conformité à la réalité [DUBO-88]. Une proposition peut être imprécise (« cet homme est grand »), incertaine (« cette lettre arrivera demain ») ou à la fois imprécise et incertaine (« il pleuvra beaucoup demain »).

La nécessité de prendre en compte dans la fusion des images les imprécisions et incertitudes des informations à combiner apparaît clairement à l'expérience.

Tout d'abord, ces caractéristiques de l'information sont toujours présentes dans les images. Elles trouvent leur origine à différents niveaux, depuis les phénomènes observés jusqu'aux traitements. 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 l'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 SAR 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 [SIMO-89].

De cet antagonisme naissent souvent des contradictions en fusion de données 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. Il conviendra donc qu'un système de décision en fusion gère explicitement incertitude et imprécision pour éviter les incohérences.

Dans cette partie, nous récapitulons les différentes méthodes permettant de les modéliser (partie 4.1.), de combiner des informations imprécises et incertaines dans un processus de fusion (partie 4.2.), pour prendre une décision, dernière étape du processus (partie 4.3.). Les méthodes étudiées couvrent la plupart des approches numériques. Les méthodes de type logique ne seront pas abordées ici. On pourra se reporter à [SOMB-89] pour une revue comparative des divers types de raisonnement adaptés aux données imprécises, incertaines et incomplètes.

4.1. MODÉLISATION

De nombreuses théories ont été proposées pour représenter et gérer l'incertain et l'imprécis dans les informations, en particulier en intelligence artificielle pour raisonner sur ce type d'informations. Nous nous attachons à décrire ici comment elles sont utilisées en traitement d'images en nous restreignant aux approches numériques. L'étape de modélisation concerne le passage d'informations extraites des images à une représentation mathématique liée à une théorie particulière.

4.1.1. Modèles probabilistes et bayésiens

La théorie la plus exploitée dans la littérature est de loin la théorie des probabilités, associée à la théorie bayésienne de la décision [DUDA-73]. 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 partie 2 s'écrit-elle sous la forme :

$$M_i^j(x) = p(x \in C_i | I_j). \quad (6)$$

Cette probabilité est calculée à partir de caractéristiques $f_j(x)$ de l'information extraites des images. 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 6 ne dépend alors plus de toute l'image I_j mais s'écrit sous la forme simplifiée :

$$M_i^j(x) = p(x \in C_i | f_j(x)). \quad (7)$$

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 l'image I_j), sont apprises par dénombrement sur des zones de test et on en déduit les probabilités des équations 6 et 7 par application de la règle de Bayes.

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, [LEE-87], [LUO-89], [CHAU-95]). 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.

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 (comment caractériser des zones qui ne soient pas du blé en imagerie aérienne?), 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 (cas des boucles logiques [PEAR-86b]).

4.1.2. Modèles flous

Parmi les techniques non probabilistes qui ont fait leur apparition depuis une dizaine d'années en traitement d'images (surtout en imagerie satellitaire, puis, plus récemment, en imagerie médicale), la théorie des ensembles flous fournit un très bon outil pour représenter explicitement des informations imprécises, sous la forme de fonctions d'appartenance [ZADE-65], [KAUF-77], [BANO-78]. La mesure $M_i^j(x)$ prend alors la forme :

$$M_i^j(x) = \mu_i^j(x), \quad (8)$$

où $\mu_i^j(x)$ désigne par exemple le degré d'appartenance de x à la classe C_i selon l'image I_j , ou la traduction d'une information symbolique exprimée par une variable linguistique (par exemple [DELL-92]), etc.

On trouve dans la littérature essentiellement deux approches pour l'utilisation des ensembles flous en traitement d'image [BLOC-94a] : 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 deuxième utilise les ensembles flous pour représenter directement les classes ou structures dans l'image, recouvrant spatialement les objets d'une mesure floue. 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 une mesure floue. 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 démuné pour définir ces fonctions. Dans la plupart des applications, cette définition 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 [BEZD-81]. 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 [ZADE-78], [DUBO-88], 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 (qui sont des fonctions d'appartenance à des ensembles flous) et de deux fonctions caractérisant les événements : la possibilité Π et la nécessité N , définies à partir de la distribution de possibilité pour un événement $A \subset S$ par :

$$\Pi(A) = \sup \{ \pi(s), s \in A \}, \quad (9)$$

$$N(A) = \inf \{ (1 - \pi(s)), s \notin A \} = 1 - \Pi(A^C), \quad (10)$$

où A^C désigne le complémentaire de A (l'événement contraire).

Des contraintes de normalisation sont introduites dans cette théorie, plus faiblement que dans les probabilités :

$$\sup \{ \pi(x), x \in S \} = 1, \quad (11)$$

$$\Pi(A) > 0 \Rightarrow N(A) = 0, \quad (12)$$

$$N(A) > 0 \Rightarrow \Pi(A) = 1, \quad (13)$$

$$\max[\Pi(A), \Pi(A^C)] = 1, \quad (14)$$

$$\min[N(A), N(A^C)] = 0. \quad (15)$$

Une distribution de possibilité s'interprète comme une fonction donnant le degré de possibilité pour qu'une variable prenne la valeur s , S étant l'ensemble des valeurs que peut prendre la variable. Dans le cadre de la fusion numérique d'images, une application possible de cette théorie consiste à prendre $S = D$ (l'ensemble des classes) et à définir la mesure M_i^j par :

$$M_i^j(x) = \pi_j(C_i)(x), \quad (16)$$

c'est-à-dire comme le degré de possibilité pour que la classe à laquelle appartient x prenne la valeur C_i , selon l'image I_j . La possibilité et la nécessité pour une classe C_i s'écrivent alors :

$$\Pi_j(C_i) = \pi_j(C_i), \quad N_j(C_i) = \inf \{(1 - \pi_j(C_k)), C_k \neq C_i\}. \quad (17)$$

Pour un sous-ensemble quelconque A de D , la possibilité et la nécessité sont calculées d'après les formules 9 et 10.

4.1.3. Modélisation dans la théorie des croyances de Dempster et Shafer

La théorie des croyances permet, de manière analogue à la théorie des possibilités, de représenter à la fois l'imprécision et l'incertitude, à l'aide de fonctions de masse m , de plausibilité Pls et de crédibilité Cr [SHAF-76], [GUAN-91]. 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.

Cette fois, la normalisation ne porte que sur les fonctions de masse :

$$\sum_{ACD} m(A) = 1. \quad (18)$$

La difficulté de la définition des fonctions de masse est augmentée ici si l'on veut affecter des masses aux hypothèses composées [GARV-86], [LOWR-91]. Jusqu'à présent, il existe essentiellement des méthodes heuristiques, qui calculent d'abord les masses sur les singletons dans une image I_j par :

$$m_j(\{C_i\})(x) = M_i^j(x), \quad (19)$$

où $M_i^j(x)$ est estimée comme une mesure d'une des autres théories exposées ci-dessus (probabilité, degré d'appartenance ou de possibilité), puis répartissent les masses sur l'ensemble des hypothèses composées, ou n'utilisent que certaines hypothèses composées, dans une démarche simplificatrice [LEE-87], [RASO-90], [CLEY-91], [ZAHZ-92].

Cependant, 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. De plus, cette théorie fournit explicitement une mesure de l'ignorance que l'on a sur un événement A et son complémentaire, comme longueur de l'intervalle de confiance $[Cr(A), Pls(A)]$. Si l'on affecte des masses uniquement aux hypothèses simples ($m(A) = 0$ pour $Card(A) > 1$), alors les trois fonctions m , Cr 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 croyances.

4.1.4. Postulats de base pour modéliser des fonctions de confiance

La popularité des probabilités est essentiellement due à leur interprétation fréquentiste⁵, ou encore « objective », d'où découlent directement les méthodes d'apprentissage, où l'on « compte » le nombre d'occurrences d'un phénomène. Cependant, cette interprétation, assez récente puisqu'elle est directement liée à l'essor des sciences physiques [SHAF-78], [SHAF-86], n'est pas toujours acceptable quand il s'agit de modéliser un phénomène unique ou quand des aspects plus subjectifs interviennent. De cette constatation est née l'approche « subjectiviste » des probabilités. Il s'agit alors de manière plus générale d'attribuer un degré de confiance à un événement ou une opinion sans référence à sa répétition. L'approche « subjectiviste » des probabilités suscite encore des polémiques, et cela depuis de nombreuses années [KEMB-42], [GOOD-59], [DUBO-93], la question clé étant : « les probabilités subjectives doivent-elles être nécessairement additives ? ». Ainsi, ce sont les axiomes des probabilités qui sont ici remis en cause. Sans entrer dans ce débat, rappelons que ces « axiomes » peuvent en réalité être déduits d'un ensemble de postulats de base, dictant intuitivement ce que l'on attend de degrés de confiance et d'une logique inductive. Cette approche a donné lieu à une école des probabilités subjectives, à la suite des travaux de Cox [COX-46].

Le premier de ces postulats fondamentaux est celui de la cohérence, ou encore de la non-contradiction. À des propositions qui ont toutes la même valeur de vérité, on doit attribuer des confiances égales. De plus, si un problème peut être résolu, à partir des mêmes données, de plusieurs manières différentes, elles doivent toutes conduire au même résultat et leurs conclusions ne doivent pas être contradictoires.

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

Le deuxième postulat impose une continuité de la méthode; on ne souhaite pas de changement brutal dans un résultat si une des données change un peu.

Le troisième postulat est celui de l'universalité ou complétude : à toute proposition bien définie on doit pouvoir affecter un degré de confiance. De plus, les degrés de confiance doivent pouvoir être comparés.

Le quatrième postulat exprime que les énoncés doivent être 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.

Enfin, le dernier postulat interdit le refus d'information. Cela signifie qu'il ne faut pas tirer de conclusion à partir d'informations partielles et qu'en particulier il est important de prendre en 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 Q n'est connu que conditionnellement à un état de connaissance e qui représente des informations reliées à la confiance dans Q et qui sont supposées (ou crues) vraies. Un tel degré de confiance est noté $[Q|e]$.

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 $[\bar{A}|e]$. On en déduit alors ([COX-46], [JAYN-57], [TRIB-72], [DEMO-93]), à l'aide des règles de la logique classique imposées par le postulat 4, la forme générale de T et S ⁶. Les deux résultats fondamentaux qui en découlent sont les relations bien connues des probabilités composées (règle du produit) et du tiers exclu (règle de la somme) :

$$p(AB|e) = p(A|Be)p(B|e), \quad (20)$$

$$p(A|e) + p(\bar{A}|e) = 1. \quad (21)$$

Tout le formalisme des probabilités trouve ainsi une base axiomatique indépendante des notions usuelles d'occurrence.

6. Ces développements supposent que T et S sont deux fois différentiables. En réalité, des résultats sur les équations fonctionnelles [ACZE-66] permettent de réduire les hypothèses sur T .

Les propriétés différentes sur lesquelles reposent les autres modélisations présentées plus haut découlent de postulats qui peuvent être différents de ces cinq postulats, soit par généralisation, soit par spécialisation, soit en acceptant une incohérence interne, soit encore par substitution d'une propriété par une autre [HORV-86].

Ainsi, la théorie des ensembles flous peut-elle traiter des propositions non claires, par généralisation du postulat 4. La règle du produit (équation 20) déduite de la dépendance du contexte est substituée par d'autres expressions de combinaison, du type $\mu(AB) = \min[\mu(A), \mu(B)]$ (voir partie 4.2.).

Le postulat 2 (continuité) est violé dans la théorie des possibilités et dans la théorie des croyances de Demspter et Shafer, puisqu'on affecte deux mesures aux propositions (Π et N ou Pls et Cr) et non une seule. La propriété de complétude est généralisée dans la théorie des croyances puisqu'il est possible d'avoir des propositions bien définies auxquelles on n'affecte pas de degré de confiance. De plus, la propriété de complémentarité (équation 21) n'est pas vérifiée, puisqu'on n'a que des relations du type $Pls(A) = 1 - Cr(\bar{A})$.

Des exemples de spécialisation et d'incohérence interne se trouvent par exemple dans le système MYCIN [SHOR-75].

4.2. COMBINAISON

Une fois les informations issues de chaque capteur représentées selon un des modèles décrits précédemment, elles peuvent être combinées, selon des règles propres au cadre théorique choisi. Cette étape constitue à proprement parler la fusion des images. Dans la suite, nous prenons l'exemple de la classification à partir d'images multi-sources. Les C_i sont donc ici des classes.

4.2.1. Combinaison probabiliste et bayésienne

Dans le modèle probabiliste et 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 :

$$p(x \in C_i | I_1, \dots, I_l), \quad (22)$$

à l'aide de la règle de Bayes :

$$p(x \in C_i | I_1, \dots, I_l) = \frac{p(I_1, \dots, I_l | x \in C_i) p(x \in C_i)}{p(I_1, \dots, I_l)}, \quad (23)$$

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) :

$$p(x \in C_i | I_1, \dots, I_l) = \frac{p(I_1 | x \in C_i) \dots p(I_l | x \in C_i, I_1, \dots, I_{l-1}) p(x \in C_i)}{p(I_1) p(I_2 | I_1) \dots p(I_l | I_1, \dots, I_{l-1})} \quad (24)$$

Très souvent, étant donné 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.

L'avantage de cette théorie, du point de vue de la combinaison, est qu'elle repose sur de solides bases mathématiques, et peut être utilisée pour la mise à jour de réseaux complexes de connaissances [PEAR-86a], [PEAR-86b].

Cependant, 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 $p(x \in C_i)$, qui est souvent délicate et est 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. La théorie probabiliste et bayésienne combine les informations de façon conjonctive (voir partie 4.2.3.), ce qui 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.

4.2.2. Combinaison dans la théorie des croyances

De même que l'équation de Bayes impose la règle de fusion en probabilités, la règle de combinaison est imposée dans la théorie des croyances (règle de combinaison orthogonale de Dempster [SHAF-76]) et s'écrit (m_j étant la fonction de masse définie pour l'image j), pour tout sous-ensemble A de D non vide :

$$(m_1 \oplus m_2 \oplus \dots \oplus m_l)(A) = \frac{\sum_{B_1 \cap \dots \cap B_l = A} m_1(B_1) m_2(B_2) \dots m_l(B_l)}{1 - \sum_{B_1 \cap \dots \cap B_l = \emptyset} m_1(B_1) m_2(B_2) \dots m_l(B_l)},$$

et $(m_1 \oplus m_2 \oplus \dots \oplus m_l)(\emptyset) = 0$ (25)

si le dénominateur de l'équation 25 est non nul, c'est-à-dire si :

$$k = \sum_{B_1 \cap \dots \cap B_l = \emptyset} m_1(B_1) m_2(B_2) \dots m_l(B_l) < 1. \quad (26)$$

Cette quantité mesure le conflit entre les sources et est donc directement prise en compte dans la combinaison sous forme de facteur de normalisation, ce qui constitue un avantage par rapport aux méthodes probabilistes. Elle représente la masse qui serait affectée à l'ensemble vide si l'on n'avait pas cette normalisation. 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. 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).

À l'origine, cette règle de combinaison était réputée applicable seulement sous l'hypothèse d'indépendance des sources. Il a été montré [QUIN-89], [QUIN-91] que la règle est encore applicable sans cette hypothèse, en s'appuyant sur l'analogie avec les ensembles fermés aléatoires.

Lorsque les fonctions m , Cr , et Pls sont des probabilités (c'est-à-dire lorsque les seuls éléments focaux sont des singletons), 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 et que seul l'incertain des données doit être pris en compte.

4.2.3. Combinaisons floues et possibilistes

Dans la théorie des ensembles flous et des possibilités, de multiples règles de combinaison sont possibles [DUBO-85], [YAGE-91], et c'est là un de leurs avantages principaux. On y trouve en particulier les T-normes, les T-conormes [MENG-42], [SCHW-83], les moyennes [YAGE-88], [GRAB-93], les sommes symétriques, et des opérateurs prenant en compte des mesures de conflit ou encore de fiabilité des sources [DUBO-92], [DEVE-93]. Le choix d'un tel opérateur peut se faire selon plusieurs critères pour la fusion d'images [BLOC-94b].

Un premier critère est le comportement de l'opérateur. On distingue les opérateurs autonomes à comportement constant (toujours conjonctif comme les T-normes, toujours disjonctif comme les T-conormes, toujours de compromis comme les moyennes⁷), les opérateurs autonomes à comportement variable (comme les sommes symétriques), les opérateurs dépendant du contexte (c'est-à-dire fonction d'une mesure globale comme le conflit ou la fiabilité).

À l'intérieur de chacune de ces trois classes d'opérateurs, un deuxième critère de choix repose sur les propriétés que l'on souhaite attribuer à l'opérateur : idempotence, ou au contraire renforcement ou affaiblissement, etc.

7. Un opérateur F de combinaison floue est conjonctif si $F(a, b) \leq \min(a, b)$, disjonctif si $F(a, b) \geq \max(a, b)$ et de compromis si $\min(a, b) \leq F(a, b) \leq \max(a, b)$, correspondant respectivement à un comportement sévère, indulgent, ou prudent.

Un troisième critère est son comportement par rapport à des situations conflictuelles.

Enfin, ses capacités à discriminer des situations différentes constituent un dernier critère. Un opérateur peu discriminant est un opérateur qui fournit le même résultat pour la combinaison de données très différentes. Par exemple, la T-conorme $F(a, b) = \min(a + b, 1)$ sature très vite à 1, et on a ainsi $F(0.5, 0.5) = 1$, $F(0.1, 0.9) = 1$, ou encore $F(0.8, 0.8) = 1$, alors que ces trois situations ont des interprétations bien différentes.

Dans le cas de la théorie des possibilités, la combinaison de distributions de possibilités n'est pas toujours normalisée (par exemple si l'opérateur est une T-norme). La normalisation fait intervenir un terme qui peut s'apparenter au conflit de la théorie des croyances mais qui a une nature beaucoup plus locale, puisqu'il est du type :

$$\sup \{TN[\pi_1(s), \pi_2(s)], s \in S\}, \quad (27)$$

où TN est une T-norme par exemple (une faible valeur de ce terme de normalisation correspond à un fort conflit). L'interprétation de cette quantité en termes de conflit correspond bien à l'intuition pour des distributions de possibilités en triangle ou en trapèze, 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.

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.

4.3. 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

$$x \in C_i \text{ si } p(x \in C_i | I_1, \dots, I_l) = \max \{p(x \in C_k | I_1, \dots, I_l), 1 \leq k \leq n\}, \quad (28)$$

mais de très nombreux autres critères ont été développés par les probabilistes et des 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 démun

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

Les autres théories sont notamment plus pauvres sur ces critères; on trouve essentiellement :

– pour les ensembles flous (et de manière analogue pour la théorie des possibilités), le maximum des degrés d'appartenance :

$$x \in C_i \text{ si } \mu_i(x) = \max \{\mu_k(x), 1 \leq k \leq n\}, \quad (29)$$

où μ_k désigne la fonction d'appartenance à la classe k résultant de la combinaison;

– pour la théorie des croyances, plusieurs règles sont possibles [APPR-91] : le maximum de plausibilité :

$$x \in C_i \text{ si } Pls(C_i)(x) = \max \{Pls(C_k)(x), 1 \leq k \leq n\}, \quad (30)$$

le maximum de crédibilité :

$$x \in C_i \text{ si } Cr(C_i)(x) = \max \{Cr(C_k)(x), 1 \leq k \leq n\}, \quad (31)$$

le maximum de crédibilité sans recouvrement des intervalles de confiance (sans risque d'erreur) :

$$x \in C_i \text{ si } Cr(C_i)(x) \geq \max \{Pls(C_k)(x), 1 \leq k \leq n, k \neq i\}, \quad (32)$$

cette dernière condition étant particulièrement stricte.

La qualité de la décision est mesurée essentiellement selon deux critères :

- le premier porte sur la « netteté » de la décision : la mesure (probabilité, degré d'appartenance, etc.) associée à la décision est comparée à un seuil, choisi selon les applications;
- le deuxième porte sur le caractère « discriminant » de la décision, évalué par comparaison des deux valeurs les plus fortes.

La décision peut également être remise en cause si la combinaison elle-même est critiquable. C'est le cas par exemple si toutes les hypothèses faites se révèlent a posteriori non vérifiées, ou encore si le conflit entre les sources est trop fort, ce qui peut invalider toute la fusion dans la théorie des croyances.

Dans les cas où les mesures de qualité de la décision ne sont pas satisfaisantes, plusieurs attitudes sont possibles :

- simplement signaler que la décision prise n'est pas fiable,
- refuser la décision en créant une classe de rejet,
- examiner les points concernés à la lumière d'autres informations (de type contextuel par exemple sur les images, comme le font les techniques markoviennes ou de relaxation dans les approches probabilistes).

Le choix d'une de ces attitudes est très lié au problème traité et repose beaucoup sur des heuristiques dans la plupart des applications.

5. Conclusion

Nous avons montré dans cet article l'importance de l'étape de modélisation des connaissances dans la résolution du problème de fusion pour le traitement d'images. Nous avons insisté sur les éléments importants qui permettent de choisir une modélisation adaptée. Nous avons vu qu'ils relevaient d'une part de la scène que l'on traite par le biais de l'imprécision et de l'incertitude attachées aux mesures que l'on en fera, d'autre part de l'application que l'on vise par l'intermédiaire de grandeurs abstraites mais fondamentales qui président à l'association des mesures : hypothèses du tiers exclu, complétude et complémentarité des connaissances, sévérité, tolérance ou prudence des combinaisons, etc. On a ainsi pu constater que le formalisme théoriquement le mieux fondé, celui de la modélisation probabiliste, était aussi le plus contraignant (puisque'il ne laisse aucune place à des adaptations à un problème particulier) mais aussi le plus exigeant (il impose non seulement que de nombreuses lois soient connues, mais également qu'elles vérifient entre elles de strictes contraintes de cohérence). En échange, il fournit un cadre conceptuel riche, mais auquel les notions d'ignorance et d'imprécision font défaut. Au contraire, les théories des croyances, des ensembles flous et des possibilités permettent de modéliser à la fois l'imprécision et l'incertitude. La théorie des croyances laisse une place centrale à l'ignorance et permet une modélisation plus riche, mais elle demeure rigide quant à la combinaison des informations. La théorie des ensembles flous et des possibilités propose un cadre beaucoup plus souple sur ce dernier point, permettant de manipuler des concepts subjectivement définis et de les associer d'une façon très adaptée à une connaissance que nous avons sur l'application. Bien que nous nous soyons focalisés sur les problèmes spécifiques du traitement d'images pour la fusion de données, la plupart des méthodes décrites et des arguments invoqués sont exploitables pour des problèmes de fusion de données d'autres types, même hétérogènes.

Il n'est pas évident aujourd'hui que dans les conditions pratiques que l'on rencontre en imagerie, il soit toujours possible de tirer profit des bénéfices des approches probabilistes, les autres méthodes pouvant se montrer soit plus souples soit plus robustes. Il est cependant assez partagé que les méthodes probabilistes conduisent souvent aux solutions optimales lorsqu'elles sont appliquées dans des conditions idéales de connaissance des phénomènes observés et de leurs perturbations. Elles constitueront donc très souvent une sorte de « cas limite » pour adapter les autres méthodes à des cas imparfaits, en particulier pour l'étape de modélisation des informations, cas limites vers lesquels on s'efforce de tendre, pour des problèmes où des combinaisons de type conjonctif s'imposent, lorsque, par exemple, notre connaissance des lois a priori ou conditionnelles s'améliore, ou que la notion d'ignorance, souvent nécessaire, s'estompe.

BIBLIOGRAPHIE

- [ACZE-66] J. Aczél : *Lectures on Functional Equations and Their Applications*, Academic Press, New York, 1966.
- [APPR-91] A. Appriou : *Probabilités et incertitude en fusion de données multi-senseurs*, Revue Scientifique et Technique de la Défense, No-11, 27-40, 1991.
- [BANO-78] G. Banon : *Distinction entre plusieurs sous-ensembles de mesures floues*, Colloque International sur la Théorie et les Applications des Sous-Ensembles Flous, Marseille, 20-22 sept. 1978.
- [BERT-94] M. Berthod, G. Giraudon, E. Volden : *Vers une modélisation de la redondance et de la complémentarité d'images*, Séminaire sur la fusion de données en imagerie aérienne et satellitaire, CNRS, GdR 134, Sophia Antipolis, février 1994.
- [BEZD-81] J. C. Bezdek : *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, New-York, 1981.
- [BLOC-94a] I. Bloch : *Fuzzy Sets in Image Processing*, ACM Symposium on Applied Computing, Phoenix, Arizona, 175-179, March 6-8, 1994.
- [BLOC-94b] I. Bloch : *Information Combination Operators for Data Fusion : A Comparative Review with Classification*, Rapport de Recherche Télécom Paris 94D013, 1994. To appear in IEEE SMC, 1995.
- [CHAU-95] S. Chauvin : *Fusion probabiliste et bayésienne en imagerie satellitaire : modèles, performances et limites*, Rapport de Recherche Télécom Paris, 1995.
- [CLEM-91] V. Clément, G. Giraudon, S. Houzelle : *Interpretation of Remotely Sensed Images in a Context of Multisensor Fusion*, ECCV 92, Santa Margherita Ligure, Italy, 815-819, May 1992.
- [CLEY-91] J. Van Cleynenbreugel, S. A. Osinga, F. Fierens, P. Suetens, A. Oosterlinck : *Road Extraction from Multi-temporal Satellite Images by an Evidential Reasoning Approach*, Pattern Recognition Letters 12, 371-380, 1991.
- [COX-46] R. T. Cox : *Probability, Frequency and Reasonable Expectation*, American Journal of Physics, Vol. 14, No-1, 1-14, 1946.
- [DELL-92] S. Dellepiane, G. Venturi, G. Vernazza : *Model Generation and Model Matching of Real Images by a Fuzzy Approach*, Pattern Recognition, Vol. 25, No-2, 115-137, 1992.
- [DEMO-93] G. Demoment : *Probabilités, modélisation des incertitudes, inférence logique, et traitement des données expérimentales*, Cours de l'Université de Paris-Sud, Centre d'Orsay, septembre 1993.
- [DEVE-93] S. Deveughele, B. Dubuisson : *Using Possibility Theory in Perception : An Application in Artificial Vision*, Second IEEE Int. Conf. on Fuzzy Systems, San Francisco, California, 821-826, March 28 - April 1, 1993.
- [DUBO-85] D. Dubois, H. Prade : *A Review of Fuzzy Set Aggregation Connectives*, Information Sciences 36, 85-121, 1985.
- [DUBO-88] D. Dubois, H. Prade : *Théorie des possibilités, applications à la représentation des connaissances en informatique*, Masson, Paris, 1988.
- [DUBO-92] D. Dubois, H. Prade : *Combination of Information in the Framework of Possibility Theory*, in Data Fusion in Robotics and Machine Intelligence, M. Al Abidi et al. eds, Academic Press, 1992.
- [DUBO-93] D. Dubois, H. Prade : *Fuzzy Sets and Probability : Misunderstandings, Bridges and Gaps*, Second IEEE Int. Conf. on Fuzzy Systems, San Francisco, California, 1059-1068, March 28 - April 1, 1993.
- [DUDA-73] R. Duda, P. Hart : *Pattern Classification and Scene Analysis*, New-York, Wiley, 1973.
- [GARV-86] T. D. Garvey : *Evidential Reasoning for Land-Use Classification*, International Association of Pattern Recognition, Technical Committee 7 Workshop "Analytical Methods in Remote Sensing for Geographic Information Systems", Paris, oct. 23-24, 1986.
- [GOOD-59] I. J. Good : *Kinds of Probability*, Science, Vol. 129, No-3347, 443-447, February 1959.

- [GRAB-93] M. Grabisch : *On the Use of Fuzzy Integrals as a Fuzzy Connective*, Second IEEE Int. Conf. on Fuzzy Systems, San Francisco, California, 213-218, March 28 - April 1, 1993.
- [GUAN-91] J. Guan, D. A. Bell : *Evidence Theory and its Applications*, North-Holland, Amsterdam, 1991.
- [HORV-86] E. J. Horvitz, D. E. Heckerman, C. P. Langlotz : *A Framework for Comparing Alternative Formalisms for Plausible Reasoning*, Proc. of the National Conference on Artificial Intelligence, 210-214, 1986.
- [HOUZ-91] S. Houzelle, G. Giraudon : *Automatic Feature Extraction Using Data Fusion in Remote Sensing*, SPIE on Sensor Fusion, Boston, nov. 1991.
- [JAYN-57] E. T. Jaynes : *Information Theory and Statistical Mechanics*, Physical Review, Vol. 106, No-4, 620-630, 1957.
- [KAUF-77] A. Kaufmann : *Introduction à la théorie des sous-ensembles flous*, Masson, 1977.
- [KEMB-42] E. C. Kemble : *Is the Frequency Theory of Probability Adequate for All Scientific Purposes?* Am. J. Physics, vol. 10, 6-16, 1942.
- [KULL-59] S. Kullback : *Information Theory and Statistics*, Wiley, New York, 1959.
- [LEE-87] T. Lee, J. A. Richards, P. H. Swain : *Probabilistic and Evidential Approaches for Multisource Data Analysis*, IEEE Transactions on Geoscience and Remote Sensing, Vol. GE-25, No-3, 283-293, 1987.
- [LOWR-91] J. D. Lowrance, T. M. Strat, L. P. Wesley, T. D. Garvey, E. H. Ruspini, D. E. Wilkins : *The Theory, Implementation, and Practice of Evidential Reasoning*, SRI Project 5701 Final Report, Palo Alto, June 1991.
- [LUCA-72] A. De Luca, S. Termini : *A Definition of Non-Probabilistic Entropy in the Setting of Fuzzy Set Theory*, Information and Control 20, 301-312, 1972.
- [LUO-89] R. C. Luo, M. G. Kay : *Multisensor Integration and Fusion in Intelligent Systems*, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 19, No-5, 901-931, 1989.
- [MAER-93] P. De Maertelaere, P. Ravazzola, P. Ghesquière, A. Beltrando : *Architectures et méthodes de fusion pour la classification multisource*, Actes du Quatorzième Colloque GRETSI, Juan-les-Pins, 983-986, 1993.
- [MAÎT-91] 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.
- [MAÎT-93] H. Maître : *Entropie, information (et image)*, Partie I, Rapport Télécom Paris 93D009, mai 1993.
- [MAÎT-94] H. Maître : *Entropie, information et image*, Partie II, Rapport Télécom Paris 94D006, janvier 1994.
- [MANG-94] J. F. Mangin, V. Frouin, I. Bloch, J. Lopez-Krahe, B. Bendriem : *Fast Nonsupervised 3D Registration of PET and MR Images of the Brain*, Journal of Cerebral Blood Flow and Metabolism, Vol. 14, 749-762, 1994.
- [MENG-42] K. Menger : *Statistical Metrics*, Proc. Nat. Acad. Sci. USA, 28, 535-537, 1942.
- [PEAR-86a] J. Pearl : *On Evidential Reasoning in a Hierarchy of Hypotheses*, Artificial Intelligence 28, 9-15, 1986.
- [PEAR-86b] J. Pearl : *Fusion, Propagation, and Structuring in Belief Networks*, Artificial Intelligence 29, 241-288, 1986.
- [QUIN-89] P. Quinio : *Representation and Accumulation of Uncertain Informations : a Theoretical Comparison of Probabilistic and Some Non-Probabilistic Formalisms*, Ito Lab., Tohoku University, 1989.
- [QUIN-91] P. Quinio, T. Matsuyama : *Random Closed Sets : a Unified Approach to the Representation of Imprecision and Uncertainty*, Symbolic and Quantitative Approaches to Uncertainty, ECSQAU, Marseille, France, R. Kruse, P. Siegel (Eds.), Springer Verlag, 282-286, 1991.
- [RASO-90] H. Rasouljan, W. E. Thompson, L. F. Kazda, R. Parra-Loera : *Application of the Mathematical Theory of Evidence to the Image Cueing and Image Segmentation Problem*, SPIE Vol. 1310 Signal and Image Processing Systems Performance Evaluation, 199-206, 1990.
- [SCHW-83] B. Schweizer, A. Sklar : *Probabilistic Metric Spaces*, North Holland, Amsterdam, 1983.
- [SHAF-76] G. Shafer : *A Mathematical Theory of Evidence*, Princeton University Press, 1976.
- [SHAF-78] G. Shafer : *Non-Additive Probabilities in the Work of Bernoulli and Lambert*, Archive for History of Exact Sciences, Vol. 19, 309-370, 1978.
- [SHAF-86] G. Shafer : *The Combination of Evidence*, International Journal of Intelligent Systems, Vol. 1, 155-179, 1986.
- [SHOR-75] E. H. Shortliffe, B. G. Buchanan : *A model of Inexact Reasoning in Medicine*, Mathematical Biosciences 23, 351-379, 1975.
- [SIMO-89] J. C. Simon : *Introduction in "From Pixels to Features"*, V-X, North Holland, Amsterdam, 1989.
- [SOMB-89] Léa Sombé : *Raisonnements sur des informations incomplètes en intelligence artificielle*, Teknea, Marseille, 1989.
- [TRIB-72] M. Tribus : *Décisions rationnelles dans l'incertain*, Masson, Paris, 1972.
- [YAGE-88] R. R. Yager : *On Ordered Weighted Averaging Aggregation Operators in Multi Criteria Decision Making*, IEEE Trans. SMC, Vol. 18, No-1, 183-190, 1988.
- [YAGE-91] R. R. Yager : *Connectives and Quantifiers in Fuzzy Sets*, Fuzzy Sets and Systems 40, 39-75, 1991.
- [ZADE-65] L. A. Zadeh : *Fuzzy Sets*, Inform. and Control 8, 338-353, 1965.
- [ZADE-78] L. A. Zadeh : *Fuzzy Sets as a Basis for a Theory of Possibility*, Fuzzy Sets and Systems 1, 3-28, 1978.
- [ZAHZ-92] E. Zahzah : *Contribution à la représentation des connaissances et à leur utilisation pour l'interprétation automatique des images satellites*, Thèse de Docteur de l'Université Paul Sabatier, Toulouse, 1992.

Manuscrit reçu le 19 mai 1994.

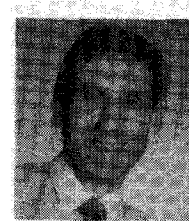
LES AUTEURS

I. BLOCH



Isabelle Bloch est née en 1964. Ingénieur Civil des Mines de Paris en 1986, elle soutient sa thèse de doctorat en 1990 à Télécom Paris. Elle est maître de conférences à Télécom Paris. Son activité de recherche est consacrée au traitement d'images et d'objets 3D, aux ensembles flous, à la morphologie mathématique 3D et floue, à la fusion de données en traitement d'images, à la théorie de la décision, à l'imagerie médicale et satellitaire, à la géométrie discrète et algorithmique 3D.

H. MAITRE



Henri Maître est né en 1948. Il est diplômé de l'École Centrale de Lyon en 1971, Docteur es Sciences Physiques de Paris VI en 1982. Il enseigne le traitement d'images numériques depuis 1973 à Télécom Paris. Chef du Département Images de Télécom Paris, il mène des recherches en traitement d'images, interprétation d'images et vision par ordinateur.