

**ACTION INTER-PRC 10.2
GDR-PRC ISIS & CHM**

**“INTERACTION SYSTEME / ENVIRONNEMENT POUR
L’INTERPRETATION DES SIGNAUX ET DES IMAGES”**

RAPPORT DE SYNTHÈSE VERSION V

TABLE DES MATIERES

| | |
|--|----|
| INTRODUCTION..... | 1 |
| PARTIE A - CO-EXPLOITER DES RESSOURCES | 3 |
| I - Fusion | 3 |
| 1. Fusion Sensorielle | 4 |
| 2. Fusion Syntaxique | 4 |
| 3. Fusion Structurale | 5 |
| 4. Fusion de Décisions | 5 |
| 5. Fusion Temporelle..... | 6 |
| 6. Stratégies de Fusion | 8 |
| II - Compétition..... | 8 |
| 1. Compétition de compétences..... | 8 |
| 2. Compétition d'accès..... | 8 |
| PARTIE B - COORDONNER..... | 10 |
| I - Coopération Opportuniste | 10 |
| 1. Contrôle Centralisé..... | 10 |
| 2. Contrôle Décentralisé..... | 11 |
| II - Coopération Planifiée..... | 12 |
| 1. Planification de Tâches | 12 |
| 2. Construction de Concepts..... | 13 |
| 3. Pilotage de Programmes..... | 13 |
| PARTIE C - ADAPTER..... | 15 |
| I - Adaptation Numérique | 15 |
| 1. Techniques de Couplage | 15 |
| 2. Techniques d'Optimisation | 16 |
| 3. Co-Adaptation du Couple Système Environnement..... | 17 |
| II - Adaptation Symbolique..... | 18 |
| 1. Focalisation d'attention et navigation | 18 |
| 2. Estimation du contexte | 18 |
| 3. Sélection a priori | 19 |
| 4. Ajustement a posteriori | 21 |
| III - Adaptation à l'Usager | 22 |
| 1. Modélisation des Stratégies Perceptives de l'Usager | 22 |
| 2. Prise en Compte Dynamique des Attitudes Perceptives de l'Usager | 23 |

TABLE DES MATIERES (SUITE)

| | |
|--|----|
| PARTIE D - APPRENDRE | 25 |
| I - Apprentissage de Classes | 25 |
| 1. Apprentissage supervisé..... | 25 |
| 2. Apprentissage non supervisé..... | 25 |
| II - Apprentissage de Règles | 26 |
| III - La “Machine Perceptive” | 27 |
| PARTIE E - CONCEVOIR..... | 28 |
| I - Vers des Systèmes Intégrés Autonomes..... | 28 |
| II - Méthodologies pour la Conception de Systèmes de Vision..... | 28 |
| DISCUSSION | 30 |
| REFERENCES..... | 32 |

INTRODUCTION

L'objectif de ce rapport est de contribuer à une vision synthétique des problèmes liés à la conception de systèmes d'interprétation des signaux et des images et des solutions envisagées. L'accent y est mis sur le principe d'interaction, qui fait la spécificité de l'Action 10.2.

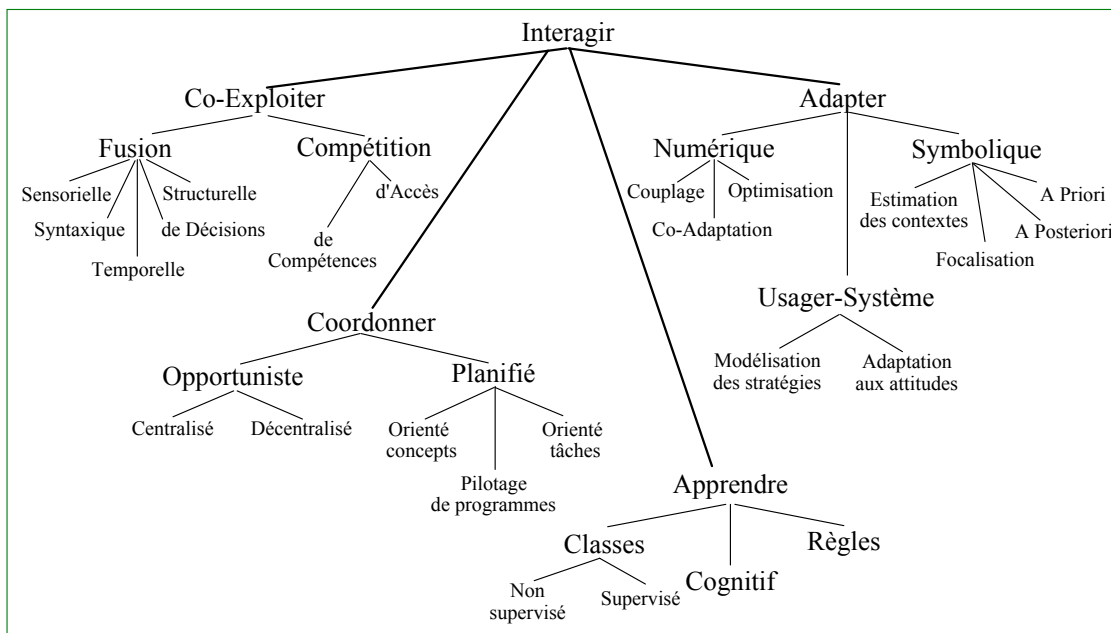


Fig.1: Typologie des modes d'interaction¹

Ce principe se décline selon différents points de vue, selon que l'on envisage l'interaction en termes de co-exploitation, de coordination, d'adaptation ou d'apprentissage :

- co-exploitation : il s'agit ici d'exploiter de manière simultanée plusieurs points de vue sur une donnée (fusion) ou sur la façon de la traiter (compétition) pour enrichir l'analyse et en exploiter les conflits éventuels;
- coordination : il s'agit de coordonner l'application d'actions, en tenant compte à la fois de la complexité des tâches à mettre en oeuvre (problème de la planification) et de la nécessité de préserver la réactivité du système aux situations imprévues (problème de l'opportunisme);
- adaptation : plusieurs approches sont également à distinguer, selon que l'adaptation est envisagée sous un angle numérique en termes de lois de

¹D'après une contribution d'Olivier Oppizzi, Laboratoire d'Informatique / Institut d'Informatique, Avignon.

dépendance ou d'optimisation, ou sous un angle symbolique, comme un problème de sélection-adaptation contextuelle : les questions d'estimation du contexte et de focalisation d'attention sont alors à considérer; outre l'adaptation système-environnement, on considère également les problèmes issus de l'adaptation à l'utilisateur du système;

- apprentissage : on distingue l'apprentissage de classes, supervisé ou non, de l'apprentissage de règles ; on évoque aussi la question de l'apprentissage en continu.

La hiérarchie illustrée en figure 1 reflète ces points de vue et l'organisation du présent rapport, qui en découle. Une réflexion sur les méthodologies de conception nécessaires à l'élaboration de systèmes intégrant ces fonctionnalités est abordée dans la dernière partie.

PARTIE A - CO-EXPLOITER DES RESSOURCES

Le terme de ressource est ici employé au sens large de données, modèles, connaissances, opérateurs, stratégies ou systèmes. L'interaction exploite alors la possibilité de mettre en jeu plusieurs points de vue sur une donnée ou sur la façon de la traiter : on parle de fusion dans le premier cas, et de compétition dans le second. La fusion est définie comme "un procédé d'intégration d'informations provenant de sources multiples en vue de produire les données unifiées les plus spécifiques et les plus pertinentes sur une entité, activité ou événement" [Bulletin de l'AFIA n°24, dossier fusion de données]. Au contraire, la compétition est définie comme la mise en oeuvre simultanée d'opérateurs susceptibles de produire des informations multiples et éventuellement conflictuelles sur une entité.

Les deux processus sont ainsi naturellement perçus comme complémentaires. Il s'agit dans les deux cas de multiplier les points de vue, qu'ils soient descriptifs ou opératoires, pour enrichir l'analyse et en exploiter les conflits éventuels.

I - FUSION

Les approches par fusion, dans leur sens le plus large, sont issues de l'utilisation conjointe de² :

- plusieurs capteurs;
- différentes résolutions, différents instants, différents modes ou niveaux d'analyse;
- données externes (cartes, atlas, schémas, images déjà interprétées).

Elles se posent en termes différents selon le niveau d'abstraction des données ; ainsi lors d'une fusion « numérique / numérique », des problèmes de recalage interviennent tout d'abord, avant même l'étape de fusion. Dans les tâches de fusion « symbolique / symbolique », ce problème de recalage est ramené à celui d'une mise en correspondance. Le problème est plus complexe lors d'une fusion « numérique / symbolique » car il est alors généralement nécessaire de définir un niveau commun de représentation des informations pour les associer. C'est très souvent une représentation symbolique qui est choisie (représentation des signaux par des graphes d'adjacence par exemple en traitement des images).

Dans ce qui suit, nous distinguerons successivement :

- la fusion sensorielle ou fusion multi-capteurs ;
- la fusion syntaxique ou fusion par projection ou appariement ;
- la fusion structurelle ou fusion de graphes ;
- la fusion de décisions ;
- la fusion temporelle, ou fusion d'événements survenant à des instants différents.

²D'après une contribution de Sylvie Philipp, ETIS, Cergy.

L'intérêt de ces approches est de gérer explicitement et de manière rationnelle les conflits de décision, et la multiplicité des représentations, au lieu de les ignorer [Oppizzi 96].

1. Fusion Sensorielle

Les techniques de fusion sensorielles sont souhaitables dans la plupart des systèmes traitant d'imagerie multi-modale (en télédétection par exemple ou en imagerie médicale). Rappelons cependant qu'elles ne sont pas indispensables ; dans la plupart des systèmes aujourd'hui opérationnels, les flots de données issus des divers capteurs sont traités séparément jusqu'à un niveau suffisamment élevé pour qu'ils puissent être livrés conjointement à l'expert qui décide (c'est-à-dire qui en fait la fusion). Ainsi en est-il des diverses modalités médicales, généralement analysées séparément par les divers experts (radiologues, neurologues, etc.), ou des images satellitaires. Dans la majorité des cas le diagnostic est essentiellement porté sur une seule source, la seconde n'intervenant qu'à posteriori pour confirmer la décision.

Les modèles de fusion utilisés dans ces approches font appel à diverses techniques dont la plus utilisée est la théorie bayésienne de la décision [Bloch 94]. Cette approche bénéficie d'un très long passé qui lui a permis de développer des outils puissants et adaptés. Son utilisation dans des problèmes précis demeure cependant souvent sous-optimale car elle prend assez mal en compte l'ignorance de l'utilisateur. Pour cela on lui préfère parfois une approche plus intuitive comme celle des ensembles flous, pour laquelle des outils spécifiques de combinaison des informations ont été développés, ou des approches prenant explicitement en compte l'incomplétude de nos connaissances (théorie de la croyance, théorie des possibilités).

Nos progrès dans le fonctionnement des méthodes de fusion sensorielle nous conduisent également aujourd'hui à proposer des techniques hybrides qui tirent parti des avantages particuliers de chacune des approches précitées. Ces techniques consistent à choisir de manière adaptative les informations nécessaires pour un problème donné en fonction des spécificités des images. C'est le cas en particulier des approches à base de règles, comme dans MESSIE [Clément 93] ou dans les travaux de Valéry Lefèvre [Lefèvre 94]. Dans ces travaux, différentes images sont disponibles (radar, IR, visible, etc), qui sont utilisées de manière combinée, selon les buts du système.

2. Fusion Syntaxique

Les techniques de "fusion syntaxique" exploitent images numériques et images "étiquetées" (e.g. images représentant des supports d'objets), à des fins de focalisation ou de fusion d'étiquetage :

- focalisation : des zones de focalisation peuvent être obtenues par projection d'une carte contour ou région, obtenue par analyse d'une modalité image, sur une image d'une autre modalité ;
- fusion d'étiquettes : il s'agit de raisonner sur des images étiquetées obtenues par

l'analyse de différentes modalités, en vue d'infirmer, confirmer ou enrichir les descriptions de ces images ; ces techniques, couramment utilisées en analyse multi-échelles sont aussi utilisées en analyse multi-modale, comme dans [Lefèvre 94].

Une méthode symbolique pour la détection de contours a été proposée³ qui se fonde sur l'emploi de représentations symboliques à caractère sémantique. Les symboles utilisés permettent d'obtenir des informations concernant les propriétés bidimensionnelles du contours et de guider un processus de suivi de contour.

3. Fusion Structurelle

Les techniques mises en jeu à ce niveau visent à comparer les structures même des images issues des divers capteurs⁴. Elles s'appuient donc sur des descriptions d'ensemble qui peuvent s'adapter aux déformations ou à la variabilité des phénomènes observés. Les représentations par graphes sont les plus fréquentes en traitement des images (graphes d'adjacence par exemple), mais aussi par arbres (arbres hiérarchiques multi-résolutions) et parfois par grammaires.

Les travaux de H. Moissinac [Moissinac 95] par exemple concernent l'interprétation conjointe de cartes et d'images aériennes en milieu urbain. Ils s'appuient sur une représentation de la ville par deux graphes duaux : l'un représentant le réseau routier, et l'autre la structure en pâtés de maisons. Un schéma global de gestion de l'incertain dans ces graphes a été construit, afin de permettre l'évaluation numérique des informations symboliques portées par les graphes et des résultats de la fusion avec les informations issues des images. Cette évaluation repose sur une collaboration étroite entre les connaissances symboliques (modèle de la ville, descriptions symboliques des objets) et des informations numériques (mesures numériques des propriétés de chaque objet, degré de confiance). La structure de graphes se révèle être un moyen efficace pour traduire symboliquement et numériquement les relations entre les objets complexes. Toute la force de ce système repose sur la capacité à fusionner et combiner compétitivement ou coopérativement de nombreux algorithmes d'analyse tout en tenant compte de la qualité des résultats obtenus par chacun.

Ces travaux se poursuivent actuellement par la thèse de D. Cassou-Ribehart⁵. Une procédure de fusion des graphes associés aux différentes sources d'information est étudiée afin d'obtenir une représentation partielle mais cohérente de la scène à interpréter. Les conflits apparaissant entre les différents graphes au cours de cette fusion ainsi que les lacunes propres au graphe résultant seront utilisés pour établir une stratégie de recherche d'informations cohérentes grâce à l'utilisation de détecteurs (agents) spécialisés travaillant de manière locale (focalisation sur les zones à problème).

Les travaux de J.-F. Mangin [Mangin 95] concernent la reconnaissance des sillons corticaux à partir d'IRM anatomique 3D. Les deux premières étapes de la méthode portent sur l'extraction automatique d'un graphe relationnel attribué représentant la

³D'après une contribution de M.C. Mraghni, Université de Tours

⁴D'après une contribution de Henri Maître, Département Images, ENST, Paris.

⁵D'après une contribution de Dominique Cassou-Ribehart, Département Images, ENST, Paris.

topographie corticale individuelle et sur la constitution d'une base de données de tels graphes étiquetés. Ensuite, un modèle structurel stochastique est inféré à partir de la base de données. Ce modèle est un prototype structurel dont les noeuds peuvent être divisés selon des contraintes syntaxiques et grammaticales. Il repose sur des interprétations originales de la variabilité structurelle inter-individus de la topographie corticale. Ce prototype est muni d'une structure de graphe stochastique prenant en compte cette variabilité. Le processus de reconnaissance est alors posé comme un problème d'étiquetage dont la solution, définie comme l'estimation du maximum a posteriori d'un champ markovien, est obtenue par recuit simulé.

4. Fusion de Décisions

Le principe est ici la mise en concurrence de décisions issues de classifieurs indépendants ou de règles, chaque "décideur" fournissant un point de vue différent sur le problème.

Cette technique a été appliquée par O. Oppizzi [Oppizzi 96], pour la reconnaissance de la parole, et par P. Loonis, pour le tri de poissons [Loonis 96].

Le mode de contrôle adopté dans [Oppizzi 96] est le suivant:

- Les informations produites par chaque règle sont associées à une mesure de confiance;
- Ces mesures de confiance sont fusionnées selon un double point de vue: obtention d'une valeur de confiance pour chaque phonème, toutes règles confondues, puis obtention d'une valeur agrégée pour le mot.

Dans l'approche proposée par [Loonis 96], des décisions indépendantes sont tout d'abord prises sur des jeux différents de paramètres, puis fusionnées par un perceptron multi-couches, ce dernier étant utilisé en auto-association, au terme d'un apprentissage spécifique (figure 2).

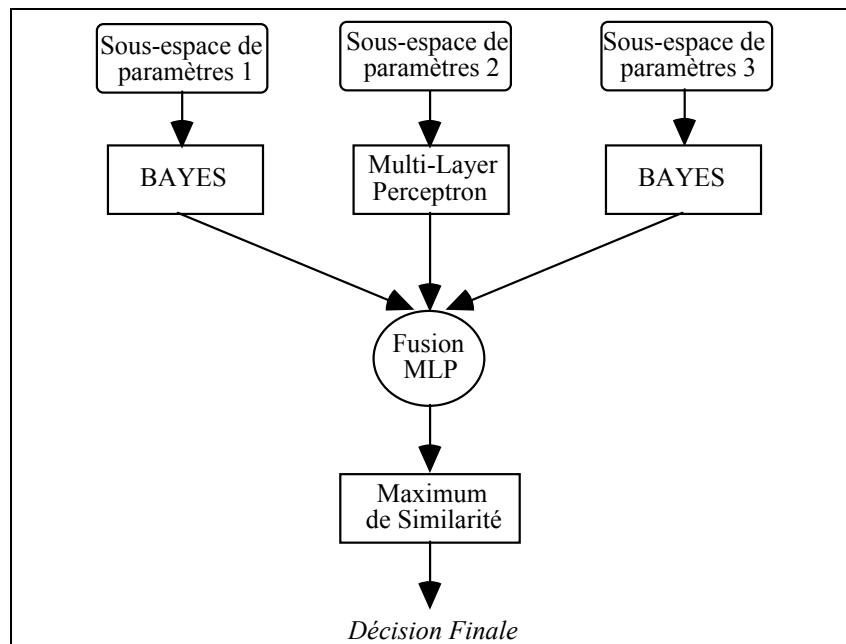


Fig. 2 : Un modèle pour la Fusion de Données

5. Fusion Temporelle

Les problèmes de fusion temporelle ont été évoqués dans le cadre de l'aide à la conduite automobile⁶. La dynamique du trafic routier est en effet génératrice d'événements survenant de manière imprévisible et issus de sources multiples et variées (obstacles, signalisation, etc).

Une information plus ou moins riche sur cette dynamique est obtenue à chaque instant, selon les événements perçus, qui peut conduire à une représentation incomplète ou erronée de la scène. Le processus de reconnaissance, dans ces conditions, s'appuie non seulement sur les mesures obtenues, mais également sur des critères de persistance des hypothèses : on procède dans ce but à un "suivi temporel" de ces hypothèses.

Ces travaux ont conduit au développement d'un système intégrant plusieurs niveaux de fusion⁷, depuis la fusion numérique jusqu'à la fusion symbolique (figure 3).

⁶D'après une contribution de Kamel Boucheffa, IEF, Orsay

⁷D'après une contribution de Roger Reynaud, IEF, Orsay

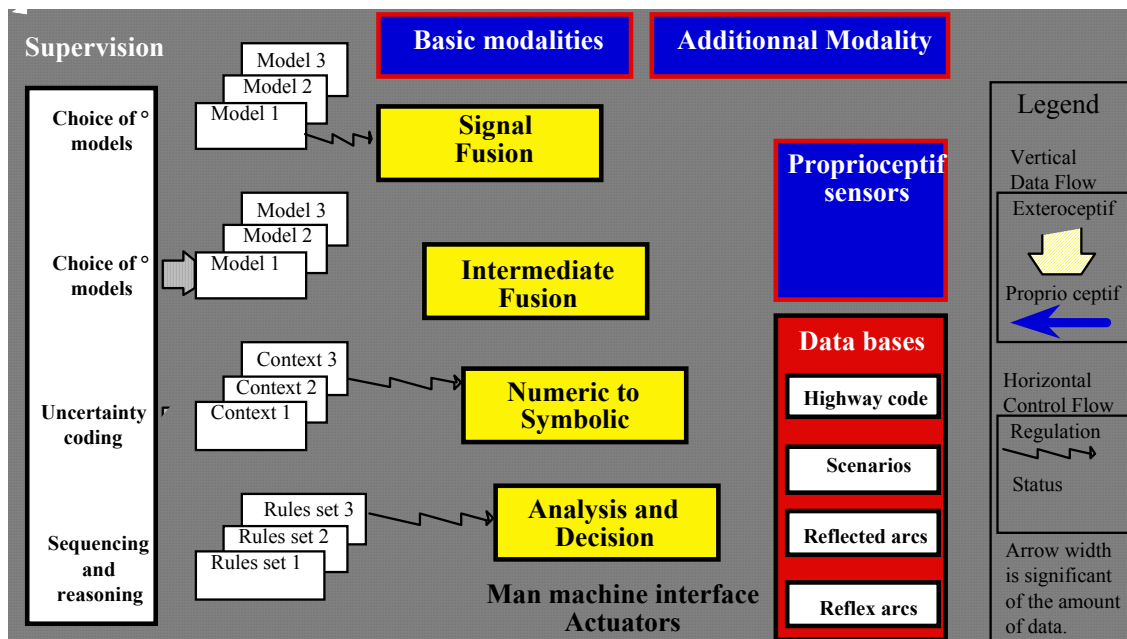


Fig. 3 : Un système intégrant différents niveaux de fusion (d'après [Reynaud 96]).

Les problèmes de fusion temporelle ont également été évoqués dans le cadre d'une application à l'interprétation de la cinétique artérielle 3D⁸. Le mouvement cardiaque est interprété à partir d'une séquence d'images angiographiques biplan, où les coronaires ont été reconstruites en 3D. Les coronaires recouvrant le myocarde, leur dynamique reflète en effet celle des régions myocardiques associées. L'interprétation implique deux étapes :

1. l'extraction de paramètres à la fois descriptifs du mouvement et cliniquement interprétables ;
2. l'élaboration de formalismes de représentation et de raisonnement spatio-temporels.

Plusieurs paramètres de description sont retenus et apportent une solution aux problèmes de transformations numériques/symboliques. Ils permettent la formation de segments homogènes et la mise en évidence des mouvements locaux de type contraction/expansion, et des mouvements de rotation. Une maquette de système d'interprétation a été développée, visant à interpréter les déformations par segment artériel, par région myocardique et de manière globale. La modélisation qualitative de l'espace est conduite à partir d'un système de référence intrinsèque au cœur et permet de décrire les territoires du myocarde, de structurer les artères en segments anatomiques et de décrire les relations spatiales entre les différents constituants.

La modélisation du temps se base sur des entités élémentaires de type instant et intervalle et des entités de plus haut niveau de type "moment" et "période". Une base de connaissance temporelle exprime les états et événements sur la scène en introduisant

⁸D'après une contribution de Mireille Garreau, John Puentes et Christian Roux, LTSI, Université de Rennes I, GBBA, Université Simon Bolivar, Caracas, et ENST Bretagne

par exemple les notions de phase (diastole/systole). Le raisonnement spatial opère par regroupement d'informations sur les différents segments et les régions associées, de même que les inférences temporelles procèdent par concaténation d'observations liées aux instants, aux moments et aux périodes. Les résultats obtenus sur une séquence d'images réelles permettent l'association de mouvements globaux pour chaque artère sur la séquence et la description du mouvement par région irriguée.

6. Stratégies de Fusion

La question des stratégies de fusion a été fréquemment évoquée. Deux stratégies différentes ont été étudiées dans le cadre du projet AMIBE⁹, pour l'intégration des modalités auditives et visuelles dans les systèmes de reconnaissance de la parole.

La première stratégie, fondée sur l'intégration directe, consiste à fusionner les données acoustiques et labiales à l'entrée du système de reconnaissance. L'inconvénient majeur de cette approche est qu'elle ne permet pas de traiter les phénomènes d'anticipation / rétention. La deuxième stratégie, dite à identification séparée et intégration asynchrone (ISIA), consiste à fusionner les résultats (N meilleures solutions) de deux sous-systèmes, l'un acoustique et l'autre labial, coopérants selon un schéma relativement simple. L'intérêt de cette seconde stratégie est de permettre le contrôle de la décision labiale par des contraintes syntaxiques et temporelles issues de la décision acoustique.

II - COMPETITION

La compétition peut s'exercer entre compétences, selon laquelle on va "mettre en compétition" plusieurs opérateurs afin d'en conserver le meilleur résultat. Elle peut aussi être considérée comme concurrence dans l'accès à une ressource que l'on cherche à s'approprier (cas de la segmentation, où l'on cherche à définir des "territoires" de points).

1. Compétition de compétences²

La compétition de compétences consiste à produire des solutions (descriptions ou décisions) conflictuelles: celles-ci seront ensuite analysées par des techniques de fusion. Le principe est d'appliquer différents opérateurs sur un même ensemble de données, puis de conserver le meilleur résultat ou un compromis entre les résultats obtenus.

Dans l'approche proposée par [Charroux 96], une même région est segmentée de manière concurrente par plusieurs opérateurs appliqués en parallèle. Pour chaque segmentation ainsi obtenue, les régions mal segmentées sont regroupées en composantes connexes afin d'être traitées à nouveau par les opérateurs restants. La procédure est appliquée récursivement tant qu'il reste des opérateurs et des zones mal segmentées. Un arbre de segmentation est donc progressivement construit, sous la forme d'un arbre et-ou.

⁹D'après une contribution de Mamoun Alissali, Paul Deléglise et Alexandrina Rogozan, Laboratoire d'Informatique de l'Université du Maine, Le Mans

Il faut ensuite comparer et fusionner tous les résultats pour garder dans chaque zone la meilleure segmentation. La segmentation est évaluée par la qualité de l'interprétation effectuée (degré d'appartenance aux classes).

2. Compétition d'accès^{10,11}

A défaut d'un parallélisme réel, l'accès concurrent à une ressource (l'image) peut être simulé par des processus interruptibles gérés par un séquenceur, comme dans [Salotti 94]. Les processus travaillent alors de manière incrémentale (e.g. croissance de région, suivi de contour, etc), pour un certain quantum de temps. Chaque processus a la capacité de s'interrompre, à la fin de chaque cycle d'analyse, et de demander le lancement d'un processus complémentaire, si besoin.

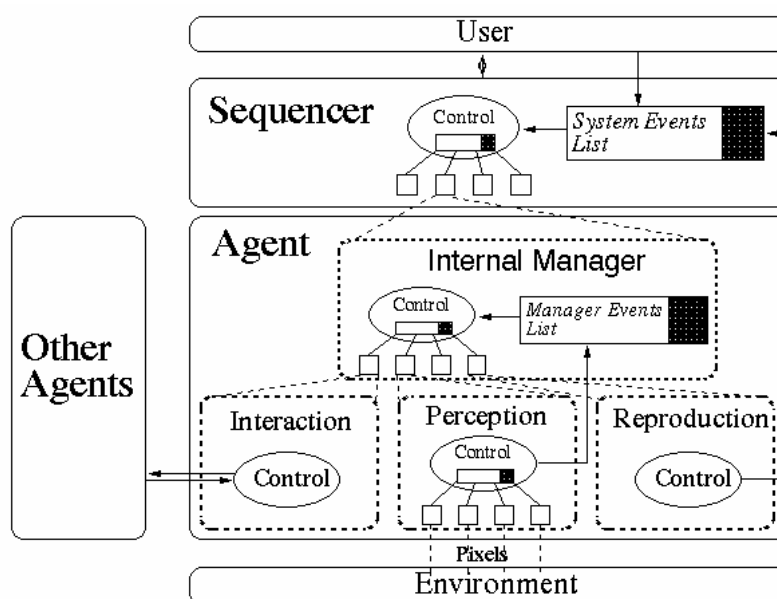


Fig. 4 : un modèle d'agent concurrent.

Selon cette approche, l'exécution des tâches est réellement entrelacée, et il devient possible de bénéficier d'une réelle interaction entre les traitements (notion de simultanéité des traitements).

Un modèle similaire a été développé par A. Boucher pour la segmentation de séquences d'images [Boucher 96]. L'approche se fonde sur le séquençage anonyme de tâches spécialisées exploitant un environnement commun. Le déclenchement d'un agent ou d'un comportement au sein d'un agent est décidé de manière autonome et décentralisée, au vu d'évènements rencontrés par l'agent dans l'environnement. Le modèle d'agent correspondant est donné dans la figure 4. Les agents sont spécialisés dans la recherche de telle ou telle information, par le biais du comportement de perception (segmentation de telle région par exemple). Ils sont dotés de comportements d'interaction qui permettent la négociation d'informations et de comportements de "reproduction" qui

¹⁰D'après une contribution de Marc Salotti, Groupe VISIA, CMCS, Université de Corte

¹¹D'après une contribution de Alain Boucher, TIMC-IMAG, Grenoble

déclenchent l'activation d'autres agents dans l'image.

Une telle concurrence peut conduire à des conflits d'accès, à partir desquels il peut être intéressant de raisonner.

PARTIE B - COORDONNER

Dans le domaine de l'interprétation d'images, la coordination d'actions pose un double problème, dû à la complexité des tâches à mettre en oeuvre et à la variabilité des situations à analyser. Ainsi, il convient que les traitements sélectionnés répondent à un double objectif : cohérence vis-à-vis du plan d'action en cours et adaptation fine aux situations rencontrées. Deux points de vue complémentaires - opportunisme et planification - sont souvent associés dans ce but, et intégrés dans le cadre de stratégies de contrôle qui en garantissent la robustesse. Ces stratégies sont évoquées en partie C.

I - COOPERATION OPPORTUNISTE

Deux modes différents de contrôle peuvent être adoptés pour régir la coopération entre tâches (qui peuvent être implantées comme des procédures, des processus ou des agents):

- le mode centralisé, selon lequel un superviseur décide de la tâche à entreprendre, au vu de l'état courant de la résolution (souvent représenté dans un tableau noir) ;
- le mode décentralisé selon lequel chaque tâche (ici conçue comme un agent) effectue cette décision de manière autonome.

Dans les 2 cas, l'accent est mis sur le caractère incrémental et constructif de la stratégie de résolution, chaque spécialiste apportant un complément d'information susceptible d'être exploité par un autre.

1. Contrôle Centralisé

La plupart des systèmes d'interprétation fonctionnent selon ce mode, qui permet une réelle souplesse dans le déroulement d'un plan d'action, grâce au partage d'informations réalisé par le blackboard.

Dans le système MESSIE [Clément 93], plusieurs spécialistes coopèrent sous le contrôle d'un superviseur ; ces spécialistes sont organisés de manière hiérarchique, selon les niveaux d'abstraction du problème.

Au Greyc [Clouard 95], une architecture de type BB1 a été développée, qui exploite un blackboard de contrôle.

Une approche fondée sur la coopération entre niveau global et niveau local a également été proposée¹² dans le cadre d'une application à lecture automatique de plans mécaniques. La méthode d'analyse comporte deux phases qui coopèrent, l'une utilisant les informations fournies par l'autre :

¹²D'après une contribution de Jean-Yves RAMEL, Equipe de Reconnaissance de Formes et Vision, INSA, Lyon

- vision globale : il y a ici extraction de l'ossature du document pour en obtenir une vision globale; ainsi un contexte, sur lequel peut se baser la suite de l'analyse, est constitué sous la forme d'un modèle global du document ;
- vision locale : des processus d'analyse locale travaillent à partir de ce contexte et le font évoluer jusqu'à la compréhension totale du document.

Au niveau local, des spécialistes coopèrent, grâce au modèle de représentation, pour extraire :

- les Formes Pleines
- le Texte
- les Courbes
- les Eléments Mécaniques

Les spécialistes travaillent en mode prédiction-vérification (génération d'hypothèses à l'aide du modèle sur la présence de textes, ou de courbes, puis vérification dans l'image initiale). Ils contribuent à l'évolution du modèle global, et sont activés selon un ordre pré-défini.

2. Contrôle Décentralisé

Dans le cas d'un contrôle décentralisé, tout agent doit pouvoir décider de l'opportunité de déclencher un autre agent. Dans ce but, il doit pouvoir s'interrompre dès que :

- son comportement n'est plus adapté;
- certains événements sont rencontrés;

et mettre alors en oeuvre le processus dont la compétence est adaptée à la nouvelle situation.

Différents modèles de contrôle peuvent être mis en oeuvre à cette fin. Le modèle proposé par [Salotti 94] est de type incrémental. Chaque processus a la capacité de s'interrompre, à la fin de chaque cycle d'analyse, et de demander le lancement d'un processus complémentaire, si besoin. Un exemple est donné dans la figure 5. Il s'agit d'une application à la segmentation d'images par coopération région-contour.

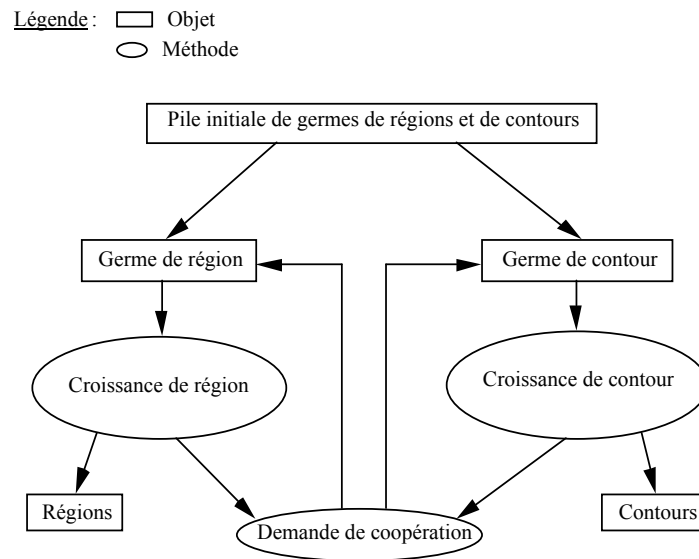


Fig. 5 : Contrôle incrémental de la coopération région-contour

L'intérêt d'un tel contrôle est de guider au mieux chaque étape en fonction des résultats qui viennent d'être obtenus. Par exemple, si une agrégation est problématique, le processus s'arrête momentanément (adaptation temporelle) et une détection des contours peut commencer exactement là où se situe le problème (adaptation spatiale), exploiter les résultats en cours de l'agrégation de points et déterminer si oui ou non un contour est présent. Pour qu'un contrôle incrémental soit efficace, il est nécessaire que chaque processus puisse s'interrompre de lui-même pour demander une coopération avec un autre processus (facilité de changement de contexte, liberté de l'information), de telle sorte que le lieu et la manière ne soient pas imposés et que des informations complémentaires soient obtenues.

II - COOPERATION PLANIFIEE

Deux points de vue co-existent dans le domaine de la coopération planifiée, selon que l'on pose le problème en termes de tâches à accomplir ou de concepts à construire. Dans tous les cas, on cherche à entrelacer génération de plans et exécution, afin de garantir la robustesse du système. Ces points de vue sont examinés successivement.

1. Planification de Tâches

Dans le cadre de l'Atelier logiciel d'Intégration des Connaissances en Traitement et Interprétation d'Images¹³, le GREYC s'intéresse aux problèmes liés à la construction de plans de traitement d'images et étudie en particulier deux modes de construction:

- construction manuelle sous forme de graphes "tache-méthode-outil",
- construction automatique à l'aide d'un générateur de plans.

Dans les deux cas, un plan consiste en un enchaînement d'opérateurs paramétrés et

¹³D'après une contribution de Marinette Revenu, GREYC - ISMRA - Equipe Image, Caen

L'exécution du plan permet de résoudre une requête de traitement d'images associée à un contexte donné.

La motivation est non seulement de résoudre des applications réelles, mais aussi de formaliser la connaissance des experts. Les opérateurs sont considérés comme des briques de base qu'il convient d'assembler sous forme d'un graphe, avant de les exécuter sous forme d'un programme C++.

En mode de construction manuelle, le problème à résoudre est décomposé sous forme de tâches, de méthodes pour accomplir ces tâches et d'outils (les opérateurs de traitement d'images). Le contrôle de l'application est alors entièrement à la charge de l'utilisateur qui doit disposer d'une interface conviviale, facilitant l'agencement des briques de base que sont les tâches, les méthodes et les outils.

En mode automatique, c'est le générateur de plans qui résout la requête. Une requête est exprimée sous la forme d'une liste attributs-valeurs. Un plan est représenté sous forme d'un arbre "et-puis" de tâches à 5 niveaux. Une tâche est caractérisée par un but à atteindre, des contraintes, des images d'entrée, des images de sortie, et des règles d'évaluation si elles existent. La décomposition des tâches en sous-tâches est effectuée par des sources de connaissances, dans le cadre d'un contrôle de type Tableau Noir. La difficulté essentielle réside dans l'écriture des sources de connaissances nécessaires à l'automatisation du processus, et dans l'évaluation des résultats intermédiaires (mécanismes d'adaptation symboliques).

Enfin, l'équipe étudie la manière de faire coopérer ces deux modes: construction automatique pour la résolution de tâches bien spécifiées, et construction manuelle quand on s'attaque à de nouveaux problèmes.

L'équipe du GREYC propose dans ce but de permettre la réutilisation des éléments quelque soit leur niveau dans le plan (i.e. les tâches comme les outils)¹⁴. Cependant, un simple "nommage" des tâches pour en présenter une liste alphabétique deviendrait vite insuffisante. Une étude des techniques de Raisonnement par Cas a conduit à essayer de définir une ontologie des tâches et des taxonomies qui permettent de classer une nouvelle tâche et de la retrouver pour la réutiliser.

Le premier pas dans cette direction a été un travail de recherche bibliographique et d'analyse du modèle de résolution du générateur de plan BORG [Clouard 94]. Ceci nous conduit à proposer les 4 axes de classification suivants :

- les 3 niveaux d'abstraction définis dans le planificateur automatique BORG (objectif, fonctionnalité, procédure)
- les différentes phases de traitement d'une image (pré-traitement, segmentation, caractérisation, resegmentation)
- les classes d'images (définies suivant des critères physiques, perceptuels et sémantiques)
- les objets du traitement (régions, contours, graphes, ...)

¹⁴D'après une contribution de Valérie Ficet, GREYC - ISMRA - Equipe Image, Caen

Une approche similaire est utilisée à l'ENSEA [Lefèvre 94] : à partir d'une requête émise dans le domaine de la scène, il y a élaboration dynamique d'enchaînements de traitements. L'ensemble des traitements est contrôlé par un système multi-agents hiérarchisé en niveaux allant de la scène (haut-niveau) à l'image (bas-niveau). Un accent particulier est mis sur la représentation des connaissances opératoires. On distingue en effet la notion de score d'aptitude et de score de réussite. La théorie de Dempster-Shafer est employée pour déterminer, à chaque niveau, l'opérateur (et son paramétrage) le plus adapté à résoudre le problème.

Le système MESSIE [Clément 93] présente également des caractéristiques similaires, chaque spécialiste possédant des capacités de planification propres.

2. Construction de Concepts

L'opérateur est ici considéré comme un "constructeur d'indices visuels", et la modélisation doit rendre compte des transformations d'information qu'il réalise, de façon à pouvoir propager ces transformations dans une chaîne de traitements¹⁵. Un enchaînement d'opérateurs est alors un processus de construction de concepts exprimés en termes d'indices visuels. Un langage de description de concepts a été défini, qui permet à l'utilisateur de formuler un problème en termes d'objectifs à atteindre. Un moteur de planification construit ensuite le graphe d'opérateurs correspondant ; il décrit l'information présente à chaque noeud du graphe [Dejean 96].

3. Pilotage de Programmes

Enfin, le pilotage automatique de programmes est abordé dans le cadre du projet ORION [Thonnat 93] : les auteurs insistent sur la nécessité d'entrelacer planification et contrôle d'exécution, afin de garantir la robustesse du système face à un environnement changeant, non prédictible, et à des actions dont la réussite est variable.

En effet, un objectif du projet ORION, à l'INRIA, est de faciliter la construction de systèmes intelligents automatiques et adaptatifs, c'est-à-dire des systèmes incorporant explicitement une expertise (intelligents), fonctionnant sans intervention humaine (automatiques) et ayant des capacités de réaction vis-à-vis des changements de leur environnement (adaptatifs) ; le domaine d'applications privilégié est l'interprétation d'images. Pour cela, le projet s'intéresse au pilotage automatique de programmes qui nécessite de planifier les traitements et de contrôler l'exécution de codes de calcul d'une bibliothèque ; le logiciel OCAPI est une implémentation d'un tel système sous forme d'un moteur de pilotage.

Dans ce logiciel, la connaissance est exprimée sous forme d'objets et de règles. La base de connaissance est structurée par les opérateurs élémentaires (programmes de traitement d'images), les opérateurs complexes (enchaînements d'opérateurs), les buts (fonctionnalités), les contextes d'utilisation. La décomposition d'un but est exprimée sous forme d'une séquence de sous-buts, à l'aide d'une structure proche de celle des squelettes utilisés en planification. Outre une facilité d'expression pour l'expert, cela

¹⁵D'après une contribution de Patrice Dalle, IRIT Toulouse

permet une réelle efficacité du système. Notons également que celui-ci est indépendant du(des) langage(s) de programmations de la(des)bibliothèque(s).

Le système a été conçu pour un fonctionnement automatique. Cependant il est possible de demander l'intervention de l'utilisateur à différents points du raisonnement, par exemple pour lui demander une information de contexte, une évaluation de résultats ou un ajustement de paramètres. De même celui-ci dispose d'une interface graphique d'exécution lui permettant de suivre le raisonnement et les résultats du système.

PARTIE C - ADAPTER

L'adaptation est définie comme la capacité à sélectionner et ajuster des opérateurs ou traitements de manière adaptée, étant donné un contexte (couple situation - but) de résolution. Elle peut viser également la co-adaptation du couple système environnement, où l'on cherche une adaptation entre la représentation de l'environnement et le système d'analyse associé. Cette adaptation peut être formulée de manière :

- numérique, en termes de lois de dépendance ou selon un principe d'optimisation;
- symbolique, en termes de règles de sélection, d'ajustement et d'évaluation.

Il est apparu nécessaire de traiter de manière conjointe l'adaptation à l'environnement et l'adaptation à l'usager : cet aspect est traité en section III.

I - ADAPTATION NUMERIQUE

On distingue ici les techniques de couplage qui permettent l'adaptation dynamique d'un niveau de modélisation à un autre, par la formalisation d'une relation d'interdépendance entre les 2 modèles, et les techniques d'optimisation, qui définissent l'adaptation en termes de maximisation d'une performance. D'autres auteurs cherchent un couplage "optimum" entre situations et opérateurs par co-adaptation de l'environnement modélisé et du système d'analyse.

1. Techniques de Couplage

Dans le cadre du projet Applications Multimodales pour Interfaces et Bornes Evoluées (projet AMIBE soutenu par le PRC Informatique 1993 - 1995), est étudiée la bimodalité naturelle auditive et visuelle de la communication orale. La reconnaissance automatique de la parole s'opère en synchronisant une "lecture labiale" avec un module de reconnaissance des formes acoustiques, par Modèles de Markov Cachés (MMC).

Pour fusionner les données acoustiques et articulatoires, plusieurs alternatives se présentent. Les informations peuvent être traitées sans discernement, par un MMC classique ; le vecteur d'observations est la concaténation des deux familles de paramètres labiaux et acoustiques ; ils sont considérés comme indépendants et l'utilisation de pondérations permet de réduire l'importance de l'une par rapport à l'autre (on parle dans ce cas d'approche globale et de MMC global).

Deux extensions de cette approche globale ont été proposées¹⁶ :

- un modèle hiérarchique, qui permet le couplage au niveau intra-modal entre variables au niveau acoustique et variables au niveau phonétique. Au niveau

¹⁶D'après une contribution de Régine André-Obrecht, IRIT, Toulouse.

phonétique, un premier réseau décrit le vocabulaire à l'aide d'entités de nature linguistiques. Une source markovienne acoustique est ensuite associée à chaque unité. De plus, les durées sont prises en compte au niveau phonétique, par des lois de probabilités continues. Cette approche permet de mieux découpler dimensions temporelles et dimensions acoustiques;

- un modèle maître-esclave [Jacob 96] qui permet le couplage au niveau inter-modal entre signaux labiaux et signaux acoustiques. Un automate esclave modélisant les phénomènes acoustiques est piloté par un automate maître décrivant les phénomènes labiaux. En fonction de l'état courant dans lequel se trouve le modèle maître, les probabilités de transition dans le modèle esclave sont modifiées.

Le principe des modèles dits "maître-esclave" repose sur la modélisation d'une application non plus par un MMC unique, mais par deux MMC mis en parallèle et corrélés. L'idée générale est de parvenir à une adaptation dynamique des lois de probabilités d'un des modèles de Markov cachés, en fonction du contexte courant modélisé par l'autre MMC. Un modèle maître-esclave est équivalent mathématiquement à un modèle classique de type MMC mais l'inconvénient de ce modèle réside dans son important nombre d'états et de lois.

2. Techniques d'Optimisation

Les techniques d'optimisation sont assez fréquemment utilisées en segmentation d'images, le principe étant de minimiser une erreur estimée de segmentation, ou de maximiser la cohérence de l'étiquetage obtenu.

Un modèle d'optimisation distribué a été présenté pour la détection de contours¹⁷. L'image à analyser est partitionnée en zones, d'après l'estimation du bruit et de la texture rencontrés. Chaque zone est alors analysée de manière concurrente. Des opérateurs de segmentation sont choisis de manière adaptée, au vu des caractéristiques des zones, puis ajustés par optimisation.

L'erreur de segmentation est exprimée comme une fonctionnelle complexe (1) intégrant les erreurs dites de sur-segmentation (détection de faux contours) et les erreurs de sous-segmentation (non-détection de vrais contours) [Spinu 96].

$$E[f](S, R) = f \supseteq E_{\text{sousseg}}(S, R) + (1-f) \supseteq E_{\text{surseg}}(S, R) \quad (1)$$

Les termes de l'erreur, au lieu d'une expression qualitative informelle, sont de plus exprimés en fonction des taux de bonne détection T_C et T_{NC} qui sont estimés par analyse de l'étiquetage obtenu, à l'aide d'heuristiques de décision : étant donné un prédicat "représentatif" de la qualité d'une primitive, la méthode consiste à comptabiliser les primitives vérifiant ce prédicat. Ainsi, on montre que minimiser l'erreur de segmentation revient à minimiser l'une des expressions (2) ou (3).

¹⁷D'après une contribution de Catherine Garbay, TIMC-IMAG, Grenoble

$$G[f](S, R) = (1 - f - T_C) \supseteq \text{card}(C(S)) \quad (2)$$

$$H[f](S, R) = (f - T_{NC}) \supseteq \text{card}(NC(S)) \quad (3)$$

Il est envisagé d'introduire des contraintes inter-zones (4), afin de garantir la consistance globale de l'étiquetage, tendant ainsi vers une optimisation distribuée.

$$E_{Z_{\text{total}}} = E^Z(S) + \sum_{Z' \text{ adjacent to } Z} E_{\text{incons}}(S, S') \quad (4)$$

3. Co-Adaptation du Couple Système Environnement

Dans une perspective de mise au point d'un système de reconnaissance audiovisuelle de la parole, il est difficile pour le concepteur d'un tel système de prévoir à l'avance d'une part les conditions dans lesquelles le système sera utilisé (les bruits dans un environnement de bureau sont très différents de ceux d'un atelier de travail) et d'autre part les caractéristiques du locuteur qui utilisera ce système (accent personnel, etc.)¹⁸.

Un modèle de reconnaissance audiovisuelle de la parole fondé sur un modèle d'environnement virtuel (EV) et sur un modèle de locuteur virtuel (LV) est proposé [Kabré 96]. Le modèle d'EV permet soit de générer des signaux synthétiques de longueur d'onde variées simulant d'une certaine manière le filtrage du signal de parole par les différents environnements d'utilisation, soit d'acquérir des signaux audiovisuels réels issus d'un microphone et d'une caméra. Le modèle LV permet de simuler les variations liées au locuteur (mouvement de la tête qui s'éloigne du microphone, voix plus basse, etc.).

Un couplage entre le modèle de l'environnement et celui du locuteur est réalisé avec les techniques connexionnistes qui sont utilisées à la fois pour l'extraction et l'intégration des informations acoustiques et visuelles dans la phase de reconnaissance. Un algorithme génétique est utilisé pour sélectionner le système qui présente la meilleure robustesse pour un environnement et un locuteur donnés. Ce couplage entre un système de reconnaissance audiovisuelle de la parole avec son environnement permet d'améliorer son adaptabilité et sa robustesse tout en lui conférant une certaine autonomie.

Une méthodologie similaire est proposée par [Loonis 96] dans une application à la reconnaissance de formes (tri de poissons)¹⁹. La méthode adoptée consiste à mettre à disposition du système un ensemble d'outils d'extraction d'attributs, un ensemble d'outils décisionnels et de fournir au système l'opportunité de se construire en s'adaptant à son environnement.

¹⁸D'après une contribution de Harouna KABRE, CLIPS-IMAG, Grenoble.

¹⁹D'après une intervention de Christophe DEMKO et Pierre LOONIS, Laboratoire d'Informatique et d'Imagerie Industrielles, Université de La Rochelle.

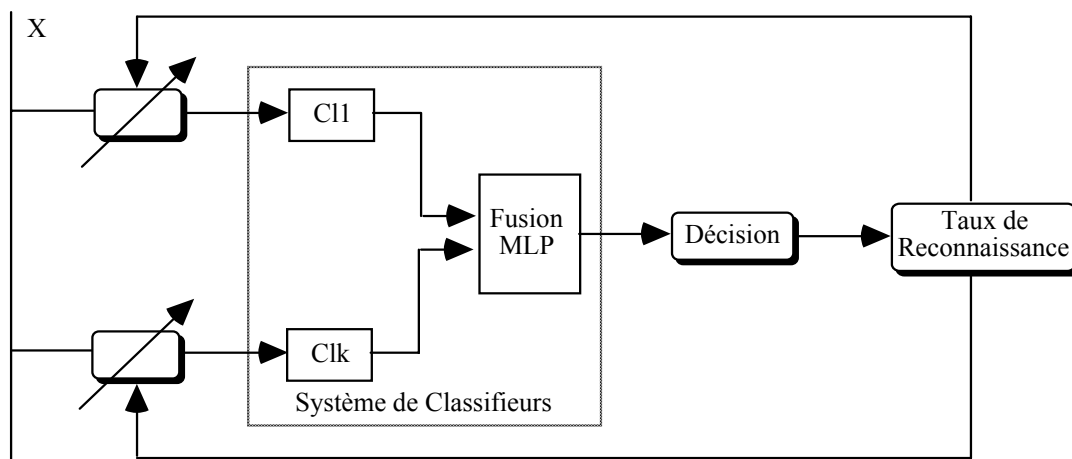


Fig. 6 : Une méthodologie pour l'adaptation du couple classifieur - environnement de description

L'objectif est la recherche d'une autonomie du système de Reconnaissance des Formes. L'expert du domaine doit apporter une information exogène complémentaire et non une connaissance nécessaire à l'élaboration du système. Afin de minimiser cet "a priori" souvent nécessaire dans la conception d'un système, une technique d'optimisation de type génétique est utilisée pour sélectionner le meilleur "appariement" entre un système de classifieurs fondé sur le principe de fusion et les données (figure 6).

II - ADAPTATION SYMBOLIQUE

On évoque ici les modèles heuristiques régissant le choix d'action ou la focalisation d'attention. Deux familles d'approche se complètent, selon que la sélection se fait sur la base d'une connaissance *a priori* (notion d'aptitude) ou selon qu'un ajustement est opéré dans le cadre de boucles de feedback (également appelées boucles de réparation) impliquant une évaluation *a posteriori* de la réussite de l'analyse. Les problèmes de focalisation d'attention sont évoqués tout d'abord, ainsi que les questions d'estimation du contexte.

1. Focalisation d'attention et navigation

Dans le système MESSIE [Clément 93], une information a priori est utilisée, mais ici plutôt à des fins de focalisation d'attention : il s'agit en effet de choisir la condition de prise de vue la plus adaptée, à partir d'une connaissance sur les aptitudes des capteurs (sensibilité aux matériaux et à la géométrie) et d'une description du but recherché.

La notion de contexte est aussi très utilisée pour l'aide à la navigation dans les bases de données hypermédia. Une application a été présentée dans le cadre du projet ATLAS²⁰, dans le domaine de l'imagerie cérébrale [Montabord 95]. Deux catégories de connaissance sont représentées dans la base de connaissance : les connaissances sources et les connaissances contextuelles. Les premières décrivent un modèle du cerveau, alors

²⁰D'après une contribution de Christian Barillot, Laboratoire SIM, Faculté de Médecine, Rennes

que les secondes décrivent d'une part les objectifs médicaux et d'autre part les connaissances relatives aux domaines concernés par ces tâches, guidant ainsi la navigation de l'utilisateur.

Une démarche d'interprétation de type descendante (génération d'hypothèses/vérification) a été utilisée à l'IGN pour faciliter et fiabiliser l'interprétation d'images aériennes, et plus particulièrement l'extraction automatique des routes²¹. Elle consiste à utiliser une Base de Données Cartographiques déjà existante mais de précision très médiocre par rapport à la résolution de l'image et au résultat souhaité.

La BD Cartographiques est utilisée pour générer des hypothèses de routes qui fournissent non seulement la localisation approximative des routes dans l'image mais également des informations sémantiques (type de route, largeur, contexte géographique...). La phase de "vérification" consiste alors à trouver, pour chaque hypothèse, la position exacte de la route dans l'image. La stratégie d'extraction des routes est locale et adaptative. C'est-à-dire que plusieurs algorithmes d'extraction de routes sont disponibles ; pour chaque hypothèse, l'algorithme le plus approprié est sélectionné et ses paramètres sont ajustés en fonction des caractéristiques et du contexte de la route.

2. Estimation du contexte

Les données contextuelles ne sont pas toujours fournies a priori, et il peut au contraire être intéressant de les calculer de manière dynamique.

Ainsi, dans les zones non texturées et non bruitées d'une image, un détecteur de contours haute-échelle donne des contours plutôt longs, isolés et bien localisés, alors que dans les zones texturées et/ou bruitées il fournit de nombreux contours courts ou longs et ramifiés, proches les uns des autres²².

Pour formuler ces caractéristiques, des attributs sur les contours et leurs voisinages ont été définis. Les voisinages sont définis par le diagramme de Voronoï construit à partir des contours. Les attributs sont géométriques, topologiques ou relatifs à la texture. Ces derniers consistent en des distances entre textures de part et d'autre des contours.

Une application de ces attributs concerne l'élaboration de cartes de texture/bruit, qui peuvent servir pour une coopération de méthodes de segmentation.

3. Sélection a priori

Ces approches s'appuient sur une description des contextes d'analyse et sur la connaissance a priori des aptitudes des opérateurs à traiter ces contextes.

Au Greyc, un contexte est modélisé selon plusieurs points de vue [Clouard 95] :

²¹D'après une contribution de Ghislaine Bordes, Gérard Giraudon et Olivier Jamet, Institut Géographique National, Saint-Mandé et Projet Pastis, INRIA Sophia

²²D'après une contribution de Florence Huet & Sylvie Philipp, ETIS-ENSEA, Cergy

- physique : type d'image et son origine
- perceptif : type des informations présentes
- sémantique : description des objets présents dans la scène

Selon ce contexte, il y a choix d'une stratégie adaptée, puis choix de techniques et choix d'opérateurs ; ces choix sont exprimés sous forme de règles, comme dans l'exemple donné dans la figure 7.

| | |
|----------------------|--|
| déclencheur : | \exists une tâche T avec le but B |
| condition : | la tâche T n'est pas encore résolue |
| action : | si <le contexte possède la valeur y> ou <T possède la contrainte z> |
| | alors <créer la décomposition de la tâche T en un enchaînement donné de sous-tâches ayant chacune un but, des contraintes, des règles d'évaluation et un chemin pour récupérer les images d'entrée> |
| | sinon <créer la même décomposition mais avec des valeurs de contraintes et des règles d'évaluation différentes> |

Fig. 7 : Un exemple de règle de planification

A l'ENSEA la théorie de Dempster-Shafer est employée pour calculer des scores d'aptitude et déterminer l'opérateur (et son paramétrage) le plus adapté pour résoudre un problème [Lefèvre 94]. Cette étude a été menée dans le cadre d'une application à la photo-interprétation.

Un score est calculé en fonction d'un contexte décrivant le but (type de problème et type d'objectif) et la situation (type de capteur, conditions géographiques, conditions météorologiques, etc). De plus, il dépend d'une description a priori des compétences de l'agent, donnée par l'expert et stipulant par exemple qu'un agent Véhicule fonctionne correctement lorsque la recherche concerne la détection de véhicules à basse altitude par un temps ensoleillé, chacun de ces critères étant pondéré par un degré d'importance. L'aptitude est ensuite calculée comme un intervalle, en termes de crédibilité (poids cumulé des critères de compétence effectivement présents dans le contexte) et de plausibilité (complément à 1 du poids cumulé des critères non présents). La présence d'un contexte moins précis augmente ainsi l'incertitude sur l'aptitude de l'agent à traiter le problème soumis. Un exemple est donné figure 8.

| | | |
|-------------|--------------------|------|
| Compétences | Temps (ensoleillé) | 0.10 |
| | Altitude (bas) | 0.15 |

| | | |
|------------|---|-------------|
| | Problème (véhicule) | 0.20 |
| | Objectif (détection) | 0.10 |
| | Nb capteurs (>1) | 0.10 |
| | Facteur d'incertitude | 0.10 |
| Contexte 1 | (Temps gris), (Altitude basse), (Problème véhicule), (Objectif détection), (Nb Capteurs 1) | |
| | Crédibilité = 0.15 (Altitude) + 0.20 (Problème) + 0.10 (Objective) = | 0.45 |
| | Plausibilité = 1 - [0.10 (Temps) + 0.10 (Nb Sensor)] = | 0.80 |
| Contexte 2 | (Problème véhicule), (Objectif détection), (Nb Capteurs 1) | |
| | Crédibilité = 0.20 (Problème) + 0.10 (Objective) = | 0.30 |
| | Plausibilité = 1 - 0.10 (Nb Sensor) = | 0.90 |

Fig. 8 : Exemple de calcul des scores d'aptitude selon deux contextes différents.

Dans le logiciel OCAPI, la sélection d'un opérateur a lieu en 4 phases : dans une première phase, les règles de choix sont déclenchées : elles permettent de conseiller, déconseiller avec d'éventuelles pondérations les opérateurs sélectionnables ; dans une seconde phase, un score global est calculé pour chacun des opérateurs ; dans une troisième phase, les opérateurs sont classés par ordre croissant avec élimination des scores nuls. L'opérateur de score maximum est alors choisi pour exécution immédiate, les autres restant possibles en cas de retour arrière.

Les règles de choix s'appuient d'une part sur la connaissance des différents opérateurs (leurs caractéristiques), d'autre part sur la base de faits actuelle (principalement les valeurs du contexte et les arguments d'entrée de la requête). Elles permettent de classer, puis de choisir le "meilleur" opérateur pour la requête donnée dans le contexte donné.

Le contexte (figure 9) est structuré en niveaux et permet de définir des informations globales à l'ensemble du traitement ou locales à une partie du traitement. Ces informations sont utilisées entre autres pour choisir entre les méthodes. Dans le cas d'images, le contexte décrira le domaine d'application (robotique, biologie, scènes routières...), les conditions d'acquisition des images (type de capteurs, longueurs d'onde...), les conditions d'observation (en ville, par mauvais temps...), les caractéristiques de l'image (présence de reflets, résolution...). Des informations plus spécifiques à une application peuvent également être exprimées, ainsi que des informations sémantiques sur le contenu de la scène (par exemple : étoiles [présentes absentes...]).

formation-image

instrument-observation
 type : schmidt, 6m
 resolution : [0,360]
 numériseur

| |
|--|
| type : microdensitometre, ccd resolution support physique : plaque photo, ccd longueur-onde : H-alpha, IR, V, B, R temps-exposition : court, normal, long |
| scene region-observee type : etoiles conditions-observation fond-du-ciel : absorption-atmospherique : pauvre, moyenne, bonne, excellente classe : 1,2,3,4,5 |
| calibrage type : recepteur, astronomique donnees : presentes, absentes |

Fig. 9 : Exemple de contexte pour le domaine "images astronomiques"

Une fois l'opérateur sélectionné *a priori*, il faut sélectionner les valeurs de ses paramètres. Cela s'implémente ici avec des règles d'initialisation.

4. Ajustement *a posteriori*

Dans le système MESSIE [Clément 93], un objet est caractérisé par 4 points de vue : Géométrie, Radiométrie, Contexte Spatial et Fonctionnalité. Le degré de certitude d'une interprétation est alors calculé comme une combinaison des coefficients de certitude attachés à ces points de vue, ce qui dans le cas du contexte spatial, fait intervenir la certitude des interprétations effectuées dans le contexte de l'objet étudié (une technique de relaxation est utilisée pour la propagation des contraintes). Le calcul de ces coefficients est fondé sur la théorie des possibilités et des ensembles flous.

Deux cas d'échecs de l'interprétation sont traités :

- cas de non-détection : lorsque l'objet recherché n'est pas trouvé, une focalisation d'attention est demandée et une nouvelle description sous contraintes effectuée;
- cas de conflits de localisation : il s'agit de résoudre des conflits d'occupation, qui peuvent correspondre soit à des erreurs effectives, soit à des conflits "usuels" (par exemple un téléphone sur une table).

Un système mixte de reconnaissance d'images aériennes a été développé à l'ENSEA [Cocquerez 92], qui exploite également une approche par relaxation pour propager les contraintes de voisinage, au niveau sémantique : il s'agit ici de lever les ambiguïtés résultant d'un étiquetage multiple.

Le graphe d'adjacence des régions obtenues avec l'opérateur Nagdif est tout d'abord calculé. L'interprétation procède ensuite en deux phases :

- application d'un réseau de neurones (ou système expert ou classifieur flou) pour

une première classification des régions : l'architecture du réseau est obtenue par similitude avec un système expert (règles ET, OU, seuillage des attributs en entrée du réseau de neurones);

- relaxation discrète, ou floue, ou application d'un réseau de neurones pour la propagation des contraintes de voisinage : un étiquetage multiple étant obtenu à l'issue de la première phase, il s'agit de le réduire en supprimant les étiquettes contradictoires entre régions adjacentes.

Le mécanisme d'ajustement *a posteriori* d'un système développé à l'aide d'OCAPI [Clément 93, Thonnat 93] permet de remettre en cause soit des valeurs de paramètres soit des choix d'opérateurs. Il est précédé par une phase de diagnostic de l'erreur appelée évaluation des résultats ; celle-ci peut avoir lieu après l'exécution d'un opérateur qu'il soit élémentaire ou complexe ; elle est soit automatique, soit manuelle. Elle permet d'émettre un certain nombre de jugements et éventuellement de provoquer un retour arrière. Les jugements peuvent être des jugements spécifiques à une donnée de sortie (*r1 est trop grand, is est floue*) ou des jugements globaux sur l'exécution (*résultats en cours d'amélioration, détection ambiguë*). Si un retour arrière est décidé, il peut s'implémenter de 2 façons : sur les valeurs des paramètres de l'opérateur évalué (ajustement des paramètres) ou sur le choix de l'opérateur (replanification).

Les mécanismes d'ajustement sont implémentés par des méthodes d'ajustement prédéfinies (par pourcentage, par dichotomie...) ou programmées dans la base de connaissances sous forme de règles de production. Elles donnent accès aux jugements portés par l'évaluation, aux informations sur les essais en cours, et aux informations provenant de la requête.

La replanification s'appuie sur le classement précédemment effectué en phase de planification. Un mécanisme plus performant, appelé réparation, a été présenté dans la thèse de Régis Vincent.

La plupart des systèmes combinent sélection a priori et ajustement, comme dans Prometheus, un système pour l'interprétation de scènes routières conçu en OCAPI [Thonnat 93].

III - ADAPTATION A L'USAGER

1. Modélisation des Stratégies Perceptives de l'Usager²³

Le principal objectif de ces travaux est de modéliser les processus perceptifs impliqués dans l'extraction d'informations pertinentes d'un document composite. Cette extraction est essentiellement envisagée comme le résultat d'un processus faisant coopérer trois approches : une approche technique de traitement d'images basée sur les caractéristiques perceptives du document étudié, une approche psychologique liée au projet de l'observateur, et une approche physiologique basée sur l'organisation du champ visuel et de la mosaïque rétinienne.

²³D'après une contribution de Véronique Eglin, Laboratoire RFV, INSA Lyon

Compte-tenu de la grande diversité de disciplines impliquées, il faut s'intéresser de près, à la fois aux mécanismes sélectifs de capture d'informations liés pour une grande part à l'organisation de la rétine, et aux indices perceptifs du document soulevant le problème de la succession des points de fixation.

Notre travail consiste, dans un premier temps, à proposer un mode de représentation de l'image rétinienne, la plus conforme possible aux réalités biologiques. Cette description constitue la base indispensable à la simulation du déplacement oculaire de l'observateur sur un document.

Simuler le parcours du regard sur le document revient ainsi à rechercher et à classer les indices perceptifs les plus discriminants, à définir des constantes perceptives dans le processus d'exploration et à traduire enfin en terme de déplacements les objectifs de l'observateur ; ce qui revient à faire coopérer deux acteurs dont les interventions sont de nature différente. Notre démarche va ainsi mettre en évidence le fait que l'enregistrement du monde visuel est un phénomène discontinu, fragmenté en échantillons et séparés par des mouvements oculaires commandés par l'observateur.

2. Prise en Compte Dynamique des Attitudes Perceptives de l'Usager²⁴

La Communication Homme-Machine s'intéresse à l'amélioration des échanges entre les hommes et les machines. Cette amélioration passe par l'introduction de nouveaux canaux de communication qui ne doivent pas augmenter la charge cognitive de l'utilisateur ni le gêner. En effet, pour plus d'efficacité, l'utilisateur est censé se focaliser sur l'accomplissement de sa tâche. Dans ce contexte, la prise en compte des informations visuelles sur l'utilisateur peut améliorer l'interaction, à condition de disposer d'outils de capture discrets pour que l'utilisateur puisse effectuer des mouvements spontanés.

Dans un premier temps, on s'intéresse à l'exploitation des mouvements oculaires en développant un système (CapRe) de capture du regard d'un utilisateur assis devant sa machine. CapRe peut être utilisé comme outil de mesure pour des expériences ergonomiques ou cognitives dans le cadre de l'utilisation d'une machine informatique, et pour l'évaluation de plates-formes matérielles et/ou logicielles dans le cadre des interfaces Homme-Machine. L'ordinateur "voit" si l'utilisateur veut interagir avec lui et dans quelle zone de l'écran il regarde. Ceci peut enlever certaines ambiguïtés et permet de contextualiser les commandes verbales dans les interfaces multimodales.

CapRe passe par plusieurs phases qui doivent satisfaire un certain nombre de contraintes liées à l'interaction homme-machine, en particulier le temps réel. Ceci est obtenu par le fait d'utiliser soit des calculs simples dans une grande région de l'image, soit des calculs complexes dans une petite région de l'image. Les trois premières phases réalisent la localisation et le suivi du visage, du nez et des yeux. Ces phases fonctionnent selon deux états : un état d'initialisation et un état d'adaptation. L'état d'initialisation consiste à s'assurer de localiser le visage et ses composantes (nez et yeux). Dans l'état d'adaptation, on suit le visage et ses composantes dans une région voisine de leurs localisations précédentes, ceci en adaptant les paramètres de

²⁴D'après une contribution de Rachid Gherbu, LIMSI-CNRS, Orsay

reconnaissance. Dans le restant des phases, le système évalue l'inclinaison du visage de l'utilisateur par rapport à l'écran et détermine le vecteur composé par le centre de la pupille et le barycentre du blanc de l'œil. Enfin, il calcule la direction du regard par composition des deux vecteurs précédents. Le système reste dans cet état d'adaptation jusqu'à ce qu'il perde le visage. Dans ce cas, il revient dans l'état d'initialisation.

PARTIE D - APPRENDRE

Le thème de l'apprentissage se décline ici selon deux points de vue différents : celui de l'apprentissage de classes et celui de l'apprentissage de règles. Notons une approche de l'apprentissage en continu dans le premier cas, en termes de réorganisation de l'espace des classes. La section III est dédiée à la description d'une "machine perceptive" dans le domaine de l'olfaction.

I - APPRENTISSAGE DE CLASSES

1. Apprentissage supervisé

La prétopologie²⁵ est un modèle mathématique développé, principalement à Lyon, dans un contexte de mathématiques appliquées aux sciences économiques et sociales. L'outil de base de ce modèle est une fonction d'ensemble appelée adhérence (notée ad) qui à toute partie A d'un ensemble permet d'associer la partie $ad(A)$ qui le contient ; à la différence de ce qui se passe en topologie classique on n'impose pas à la fonction adhérence d'être idempotente.

En reconnaissance des formes et analyse d'images, cet outil permet de faire de la modélisation-perception et aussi peut engendrer des développements algorithmiques ; en effet, on peut par exemple considérer que $ad\{x\}$ est l'ensemble des objets (ou des représentations) que l'on perçoit quand on veut percevoir (ou s'intéresser à) x ; la non idempotence permet de générer, par itération, des groupements de points autour de certains marqueurs (en classification automatique.. ou en segmentation d'images).

Jusqu'à présent, l'adhérence a toujours été évaluée de façon statique, par exemple en prenant pour $ad\{x\}$ l'ensemble de ses plus proches voisins. Une méthode de décision fondée sur l'utilisation d'adhérences adaptatives empruntées au formalisme de la prétopologie est ici utilisée [Henry 96], dans le cadre d'une application à la reconnaissance de caractères. Cette approche permet un ajustement dynamique des classes, exprimées au niveau morphologique, selon les informations produites par un étage d'analyse contextuelle travaillant au niveau lexical. Un système de pondération des modèles et de correction continue des poids permet de réorganiser l'espace des classes, selon les erreurs détectées au niveau contextuel, plaçant ainsi ce système dans un contexte d'apprentissage en continu.

2. Apprentissage non supervisé

Deux aspects de l'apprentissage non supervisé sont évoqués ci-dessous : le premier traite de la classification alors que le second aborde plutôt le domaine de la "fouille de données" (data mining).

²⁵D'après une contribution d'Hubert Emptoz, Laboratoire Reconnaissance de Formes et Vision, INSA, Lyon.

Une nouvelle architecture neuro-mimétique pour la détection et la classification non supervisée de mixtures d'événements acoustiques a été proposée²⁶. Cette méthode fondée sur un minimum de connaissances a priori utilise l'analyse en composantes indépendantes et les réseaux projectifs: un réseau à deux couches est construit et met à jour dynamiquement une classification des stimuli par détection de rupture, création et élimination de classes.

Un second processus réorganise ensuite les classes obtenues de façon à minimiser leur dépendance statistique. Cette dernière opération permet de retrouver les formes originelles de classes à partir de leurs mixtures linéaires. Notre classifieur a été testé sur des données multidimensionnelles puis sur des mixtures de signaux acoustiques. Les premiers résultats obtenus sont encourageants, malgré la mauvaise modélisation des sons temporellement structurés.

Une autre approche consiste à utiliser les réseaux de neurones multicouches pour réaliser des fonctions complexes de recherche d'information à partir de corpus d'apprentissage étiquetés. L'idée présentée ici²⁷ est de rechercher dans les poids du réseau les règles de décision simulées. Dans un premier temps, nous nous intéressons à la sélection des paramètres utiles, ce qui nous permettra d'éliminer les données superflues, mais aussi d'orienter la recherche d'une représentation plus efficace des objets.

Deux applications dans le domaine de la parole, pour la reconnaissance de phonèmes, ont montré des résultats très intéressants :

- la pondération automatique de règles de logique floue dans un système multi-agents, avec des performances améliorées par rapport aux tests statistiques;
- la reconnaissance de voyelles phonétiques avec une mesure de l'influence des contextes coarticulatoires gauche et droit.

Nous avons retrouvé dans les poids des réseaux l'utilisation des formants. Cette technique nous permet en outre un contrôle efficace de l'apprentissage des réseaux et de leur robustesse par la génération automatique d'exemples, et d'accroître nos connaissances vers des règles expertes de décision.

II - APPRENTISSAGE DE REGLES

L'objectif du projet présenté par L. Mascarilla est la conception d'un système neuro-flou pour l'apprentissage automatique de règles²⁸. Le projet s'insère dans le cadre plus général de la conception d'un système expert en interprétation d'images satellite pour la cartographie automatique, où données multi-spectrales et connaissances sont combinées pour produire une classification finale.

Deux niveaux de réseaux de neurones sont mis en oeuvre pour induire, puis optimiser,

²⁶D'après une contribution de Georges Linares, Laboratoire d'Informatique d'Avignon

²⁷D'après une contribution de Régis Quelavoine, Laboratoire d'Informatique d'Avignon

²⁸D'après une contribution de Laurent MASCARILLA, I.R.I.T. , Toulouse

des règles floues à partir d'exemples de classes de végétation désignées sur l'image par un expert.

Dans une première phase, un réseau neuro-flou de type compétitif entraîné sur les échantillons fournit, à partir des données exogènes, une base de connaissances et à partir des données image un classifieur. À ce stade, les notions d'information relative et de mesure de fréquence floue sont utilisées pour produire des règles compactes affectées d'un facteur de certitude. Une approximation linguistique permet ensuite de les présenter en langage naturel à l'expert qui peut alors les valider ou les modifier. À l'issue de cette phase, une première carte peut être obtenue en fusionnant, pour chaque pixel, les degrés d'appartenance issus de la partie règle et de la partie image.

Dans une seconde phase, un perceptron neuro-flou affine les résultats en ajustant les paramètres de la base de connaissances sur la base d'échantillons. En particulier, les formes des fonctions d'appartenance et les connecteurs logiques sont optimisés. De la même façon qu'à la fin de la première phase, une carte finale est produite.

III - LA "MACHINE PERCEPTIVE"

Il s'agit d'une recherche en cours qui porte sur la mise au point d'une architecture pour une machine "perceptive intelligente", la machine MaPe²⁹. Des travaux en neuropsychologie cognitive ont servi de support pour dégager une architecture fonctionnelle capable d'accomplir des tâches d'apprentissage, de reconnaissance et de manipulation de concepts symboliques [Zighed 96]. Cette machine perceptive a été testée sur la modalité visuelle (reconnaissance de chiffres manuscrits) et une expérience est en cours pour un test en olfaction.

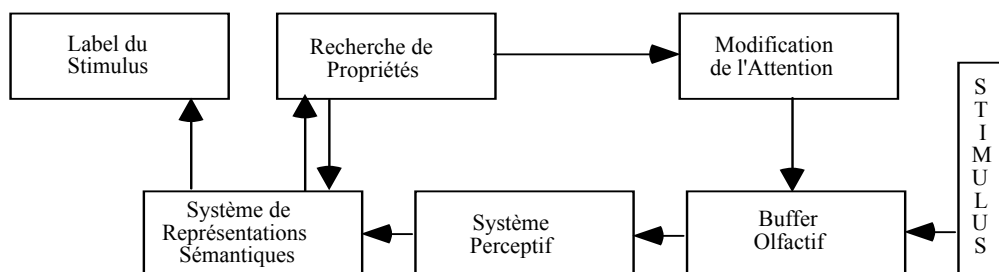


Fig. 10 : Une architecture fonctionnelle modélisant les processus cognitifs

Le principe de base est que la facilité à identifier un stimulus dépend de la capacité à le décrire à un niveau sémantique. Plus précisément, un stimulus qui a déjà été perçu, même sous une forme différente, sera plus facile à identifier (priming effect). Une stratégie de reconnaissance (figure 10) fondée sur ce principe est proposée, qui implique la coopération entre 2 modules principaux :

- un "Module Perceptif" est tout d'abord activé, qui permet l'"évocation" des odeurs les plus proches de l'odeur perçue, au sens d'un critère de voisinage donné ; un "réseau de cooptation" est mis en oeuvre dans ce but, qui définit une structure de graphe dont les

²⁹D'après une contribution de Abdelkader ZIGHED, E.R.I.C., Lyon.

arêtes définissent des relations de voisinage entre stimuli ;

- une "Mémoire Associative" est ensuite mise en oeuvre, qui est dédiée aux traitements "profonds". Elle contient une base de faits (les propriétés associées aux odeurs dans le réseau) et une base de règles permettant leur interprétation symbolique ; au signe perçu sont associées par héritage les propriétés des signes "évoqués", pondérées par importance dépendant de leur distance dans le graphe de cooptation. Un mécanisme de vote permet de contrôler le processus d'héritage.
- Les règles sont activées en cas d'indétermination (le stimulus s'apparente simultanément à plusieurs éléments de la base de fait) pour extraire de nouvelles propriétés (modification de l'attention).

PARTIE E - CONCEVOIR

La question se pose finalement d'intégrer les fonctionnalités que nous venons d'évoquer au sein de systèmes d'interprétation intégrés. Les deux contributions suivantes apportent une réflexion dans ce sens.

I - VERS DES SYSTEMES INTEGRES AUTONOMES

L'objectif du projet Ravi est de progresser à la fois dans la conception des "grandes facultés", notamment l'apprentissage et le contrôle, et dans la construction concrète d'un système informatique capable d'intégrer ces concepts³⁰.

Cette approche fait suite au constat qu'en Vision par Ordinateur (et dans d'autres domaines liés à l'Intelligence Artificielle), dès lors qu'on aborde le niveau dit symbolique, il y a un trop grand fossé entre le discours qui présente ce que l'on veut faire, et les programmes que l'on sait écrire.

Deux postulats importants sont à considérer dans ce cadre :

- co-existence de différents langages de programmation : dans un système évolué plusieurs langages, de haut niveau et de bas niveau, sont nécessaires. Cela implique la liberté du choix du langage de programmation: le système doit être capable de faire marcher ensemble des modules écrits dans différents langages;
- généralisation de l'apprentissage : celui-ci peut atteindre toutes les parties d'un système. En particulier, il peut donner lieu à la création dynamique de nouvelles fonctions dans des langages de bas niveau. Sans cela, le rendement de l'apprentissage sera maigre.

II - METHODOLOGIES POUR LA CONCEPTION DE SYSTEMES DE VISION

L'objectif de cette contribution a été de faire le point sur les différentes méthodologies du domaine de la vision par ordinateur (reconstructionnisme, vision passive, vision active, idéalisme, Gestalt, systémique, smart sensing, purposive vision, active perception, animate vision, purposive reconstructionism, ...) ³¹.

Leur étude et leur comparaison conduit à s'interroger sur ce que l'on cherche, pourquoi l'on cherche et comment l'on cherche. L'objectif plus précisément est de mettre en avant les contraintes qui s'exercent sur la conception, en tentant de les exploiter de manière explicite et intégrante. Une approche systémique est proposée, afin de bâtir une méthodologie de conception non réductionniste. L'hypothèse sous jacente est que la puissance d'un système de vision ne réside pas dans son architecture mais dans sa

³⁰D'après une contribution d'Augustin Lux, Laboratoire Gravir, Grenoble

³¹D'après une contribution de Jean-Michel Jolion, Laboratoire Reconnaissance de Forme et Vision, INSA, Lyon

complexité [Jolion 94].

Un problème important dans ce cadre est de parvenir à intégrer au sein d'un système de conception cohérente les multiples éléments d'information nécessaires à une vision bas niveau puissante³² : position, couleur et caractéristiques de voisinage des pixels, contours et régions à différentes étapes de leur construction.... Il est également de gérer ces informations de manière adéquate (quelles informations à quel moment) et de prendre des décisions robustes. Le problème est que nous ne disposons pas d'un modèle de raisonnement à partir duquel pourraient être déduites des règles de décision. La solution la plus simple est d'appliquer des critères définis de manière empirique et de procéder par tâtonnement. Cette approche est intrinsèquement limitée car l'évaluation des résultats est très subjective. Il est donc nécessaire d'apprendre à utiliser correctement les informations disponibles dans la prise de décision. L'apprentissage peut mettre en jeu un expert humain sélectionnant par exemple des configurations locales de contour estimées "à la limite de la décidabilité", à partir desquels des fonctions de décision peuvent être estimées. Une autre approche consiste à demander à l'expert de construire des cartes de référence "bas niveau", à partir d'un découpage de l'image en vues locales partielles.

Quelle que soit l'approche adoptée, la Reconnaissance des Formes revient toujours à construire et utiliser une partition de l'ensemble des données possibles³³. Ce paradigme réducteur - plusieurs signaux dans une même classe exclusive - porte en lui-même la notion de variabilité, pierre d'achoppement des systèmes de reconnaissance. La solution adoptée consiste bien souvent à restreindre le problème par l'application de contraintes à divers niveaux, et à prendre en compte toutes les variantes possibles de l'information utile.

Pourtant la simple observation de l'activité perceptive humaine montre que celle-ci est loin d'être aussi réductrice. La variabilité n'est pas une notion perceptive. Pour résoudre le paradoxe nous proposons l'idée que la perception - naturelle ou artificielle - ne se réduit pas a priori à une simple classification de signaux, mais qu'elle met en oeuvre à divers niveaux de structuration une "description complète" de ces signaux. Par "description complète" nous entendons un ensemble de descripteurs adaptés au niveau d'abstraction considéré et tels que leur spécification permette de reconstruire un signal perceptivement identique à l'original. Il peut subsister une "variabilité résiduelle", qui représente l'ensemble des variations non perceptibles dans les données.

Ce paradigme appelé Traitement des Formes englobe celui de la Reconnaissance des Formes. Il revient à considérer autant de partitions différentes de l'ensemble des données qu'il y a d'aspects perceptibles dans le signal. Il se traduit pratiquement par l'idée qu'il faut chercher à reconnaître simultanément tous les aspects perceptibles du signal. Ainsi la reconnaissance de la parole implique la reconnaissance simultanée des caractéristiques du locuteur, de sa voix et de la situation d'élocution ; la reconnaissance de scènes visuelles implique la reconnaissance simultanée des conditions d'éclairage, de la position de l'observateur et des caractéristiques du système optique. Loin de compliquer le problème nous pensons que c'est le seul moyen de démêler les différentes

³²D'après une contribution de Marc Salotti, Groupe VISIA, CMCS, Université de Corte

³³D'après une contribution de Jean-Sylvain Lienard, LIMSI-CNRS, Orsay

informations présentes dans le signal et finalement de reconnaître celle à laquelle on s'intéresse plus précisément.

Les descripteurs peuvent être de deux sortes : ils représentent soit des éléments du niveau considéré ("quoi"), soit la position de ces éléments dans un espace propre à ce niveau ("ou"). Le fait que tout élément puisse se trouver en toute position permet de mettre en oeuvre un processus d'apprentissage des classes reposant sur la convergence de multiples analogies. Ce processus d'"induction analogique" permet, contrairement aux méthodes de RF évoquées plus haut, de ne pas avoir à apprendre toutes les variantes possibles de l'information que l'on cherche à caractériser. Il constitue la base même du Traitement des Formes.

DISCUSSION

Divers modèles pour la co-exploitation, la coordination et l'adaptation de ressources ont été présentés. L'intérêt et l'apport de ces travaux a été démontré par des applications très variées à l'imagerie médicale et satellitaire, au traitement des documents, et au traitement des signaux de parole.

Différentes approches ont été étudiées (et souvent utilisées en association) pour "minimiser l'a priori" dans la conception de ces systèmes :

- les approches concurrentes permettent d'éviter le choix a priori d'opérateurs, en impliquant plutôt une analyse a posteriori des résultats obtenus. Les décisions sont ainsi reportées, et l'accent plutôt mis sur les méthodes de fusion ultérieures des résultats ; l'approche semble s'opposer à une approche par coordination et/ou adaptation ;
- les approches incrémentales (opportunistes) vont également dans ce sens, en procédant par construction progressive et contrôlée de la solution. Le couplage environnement / système se fait ici de manière assez naturelle, et permet l'adaptation effective des méthodes utilisées ; il convient néanmoins d'être capable de le formaliser ;
- les approches de type essai / erreur permettent la "réparation" des erreurs d'analyse par la mise en oeuvre de boucles de feedback. La difficulté est ici d'évaluer la qualité du résultat obtenu, et de formaliser le couplage erreur / réparation ;
- l'emploi d'une modélisation floue permet également d'éviter l'emploi prématuré de seuils de décision.

Dans l'ensemble, il convient d'employer des méthodes robustes et/ou d'être capable de les évaluer. Selon les cas, l'accent est plutôt mis sur les problèmes de fusion, d'adaptation, ou de correction.

Les modèles d'interaction utilisés demeurent encore assez simples, peut-être du fait d'une difficulté à modéliser de manière riche les contextes et relations d'interdépendance entre opérateurs, méthodes ou processus. Le paradigme actuel est moins en effet de multiplier les opérateurs mis en jeu (qu'on les appelle méthodes, processus, agents ou spécialistes), ce qui par ailleurs correspond à une nécessité effective. Il est plutôt de créer des relations de dépendances permettant de faire évoluer les méthodes d'analyse et de reconnaissance en cours d'exécution. Le besoin a également été souligné de progresser dans le domaine de la planification, où la difficulté à expliciter les notions de contexte et d'enchaînement d'opérateurs a été discutée ; le manque d'outils appropriés de modélisation se fait sentir ici.

On peut également noter un sorte de fracture entre modèles purement numériques (MMC en particulier) et modèles plutôt symboliques, malgré une intervention sur le

couplage réseau de neurones / système expert. Cet effet de fracture numérique / symbolique a été discuté, certains l'analysant comme le reflet d'une vie scientifique dynamique (développement de tendance de recherche concurrentes), d'autres attestant plutôt d'une évolution visant à en atténuer les frontières et à en rapprocher les concepts (rapprochement entre graphes et modèles de Markov par exemple).

Il est clair qu'une des limitations avancée par beaucoup de participants est l'effet "usine à gaz" obtenu dès lors qu'on essaie de modéliser "en contexte" les opérations à mettre en oeuvre, ou que l'on cherche à intégrer des mécanismes trop différents d'analyse. Il convient certainement de poursuivre le débat engagé sur les questions de méthodologies (comment s'écarter d'une vision "idéaliste" de la conception) et de modélisation (quels outils, quels langages, comment résoudre le compromis entre puissance et généralité ?).

La dépendance entre stratégies de résolution et représentations a été souligné. A cet égard, il a été noté que l'actuel rapport est plutôt centré sur la présentation de "modèles opératoires", au détriment peut-être d'une approche plus "représentationnelle" de la connaissance et des informations manipulées. La question de l'introduction de connaissances a priori, liées ou non au domaine d'application, a également été débattue.

Enfin, plusieurs exposés et discussions ont montré l'importance de "garder l'homme dans la boucle", insistant donc bien sur les deux pôles d'interaction : système / environnement et système / usager. Dans ce but, l'utilisateur peut être simplement vu comme un "outil" nécessaire à la mise en oeuvre d'une tâche, s'intégrant ainsi au système informatique de manière naturelle ; il peut être vu à l'opposé comme un système à part entière, fonctionnant selon une logique propre, et interagissant de manière complexe avec le système informatique. Les stratégies perceptives mises en oeuvre par l'utilisateur peuvent par exemple être exploitées en vue d'améliorer le mode de fonctionnement et d'interaction des systèmes à concevoir. Une typologie des objectifs attendus par l'utilisateur / réalisables par le système s'avère dans tous les cas nécessaire.

REFERENCES

- [Barillot 94] Barillot, C., Gibaud, B., Montabord, E., Garlatti, S., Gauthier, N. and Kanellos, I. : An information system to manage anatomical knowledge and image data about brain, SPIE vol. 2359, Visualization in Biomedical computing, USA, pp. 424-434, 1994.
- [Bloch 95] Bloch, I. & Maitre, H. : Fusion de données en traitement d'images : modèles d'information et décisions, revue TS, numéro spécial "Fusion de Données", 11(6):435-446, 1994.
- [Bordes 97] Bordes, G., Giraudon, G. and Jamet, O. : Automatic road extraction from grey-level images based on object database, AEROSENSE, SPIE, Orlando, avril 1997.
- [Boucher 96] Boucher, A. & Garbay, C. : Segmentation de séquences d'images cytologiques par un système multi-agents, Actes des 4ièmes Journées Francophones IAD & Systèmes Multi-Agents, Hermès, pp. 125-135, 1996.
- [Charroux 96] B. Charroux, "Analyse d'images : coopération d'opérateurs de segmentation guidée par l'interprétation", Thèse de l'Université Paris XI Orsay, janv. 1996
- [Clément 93] Clément, V., Giraudon, G., Houzelle, S. and Sandakly, F. : Interpretation of remotely sensed images in a context of multi-sensor fusion using a multispecialist architecture, *IEEE trans. on Geoscience and Remote Sensing*, 31 (4) : 779-791, 1993.
- [Clément 93] Clement, V. and Thonnat, M. : A Knowledge-Based Approach to Integration of Image Procedures Processing, CVGIP: Image Understanding, Vol.57, No.2, March 1993
- [Clouard 95] Clouard, R., Revenu, M., Elmoataz, A. and Porquet, C. : Un Atelier Logiciel pour la Conception d'Applications de Segmentation d'Images, Actes du Colloque GRETSI, 1995
- [Cocquerez 92] J.P. Cocquerez, P. Gaussier, S. Philipp, Système d'interprétation mixte : réseaux de neurones/système-expert appliqué aux images aériennes, *Traitement du Signal*, 9(5): 421-439.
- [Dejean 96] Dejean, Ph. & Dalle, P. : Modèle symbolique de la donnée de traitement d'images, Actes du 10ième Congrès RFIA, pp.746-755, Rennes, 1996.
- [Henry 96] Henry, J.L. : Reconnaissance et contexte : une approche coopérative pour la lecture de textes imprimés, Thèse de Doctorat, INSA, Lyon, 1996.
- [Jacob 96] Jacob, B. and Sénac, C. : Un modèle Maître-Esclave pour la fusion des données acoustiques et articulatoires en Reconnaissance Automatique de la Parole JEP 96.
- [Jolion 94] Jolion, J-M. : Computer Vision Methodologies, in CVGIP: Image Understanding, 59(1):53-71, 1994.
- [Kabré 96] H. Kabre, Robustness of Automatic Speech Recognition Systems in an Artificial Life Framework, *J. of Computer Speech and Language*, janv. 1996
- [Lefèvre 94] Lefèvre, V. : Système multi-agents d'aide à la photo-interprétation sur architecture multi-processeur, Thèse de Doctorat, Université de Paris-Sud, 1994.
- [Le Peutrec 94] Le Peutrec, S. and Courant, M. : Living system dynamics : a new approach for knowledge representation, in "Parallelism and Artificial intelligence", (B. Hirsbrunner, M. Courant and M. Aguilar, eds), Univ. of Fribourg Series in Computer Sciences, vol.

3, 1994.

- [Liénard 93] Liénard, J.S. : Apprentissage de transformations point à point par induction analogique, Rapport interne, LIMSI : 93-07, 1993
- [Linarès 97] Linarès, G., Nocera, P., Meloni, H. : Mixed acoustic events classification using ICA and subspace classifier, IEEE ICASSP'97, avril 97, Munich, Germany.
- [Loonis 96] Loonis, P. : Contribution à la minimisation de l'a priori en Reconnaissance des Formes, Thèse de Doctorat, Université de la Rochelle, 1996..
- [Mangin 95] Mangin, J.F., Regis, J., Bloch, I., Frouin, V., Samson, Y. and Lopez-Krahe, J. : A MRF-based random graph modelling the human cortical topography, CVR-MED'95, Nice, 177-183, April 1995.
- [Moissinac 95] Moissinac, H., Maitre, H. and Bloch, I. : Markov random fields and graphs for the uncertainty management and symbolic data fusion in a urban scene interpretation, in "SPIE Image and Signal Processing for Remote Sensing II", vol. 2579, pp. 298-309, 1995.
- [Montabord 95] Montabord, E., Gibaud, B. & Barillot, Ch. : Hyper-Yaka: hypermédia et base de connaissances sensible au contexte, Actes de la Conférence "Langage et Modélisation à Objets", Nancy, 1995, pp. 153-171.
- [Oppizzi 96] Oppizzi, O., Fournier, D., Gilles, P. and Méloni, H. : Décodage acoustico-phonétique flou, JEP96.
- [Opizzi 97] Opizzi, O. and Quélavoine, R. : Rescoring under fuzzy degrees with a multilayer neural network in a rule-based speech recognition system, IEEE ICASSP'97, avril 97, Munich, Germany.
- [Puentes 96] Puentes, J. : Analyse symbolique du mouvement cardiaque en angiographie vasculaire, Thèse de l'Université de Rennes I, 19 décembre 1996.
- [Reynaud 96] Reynaud, R. & Maurin, T. : Control and supervision of a fusion processing line by competition among models, CESA'96.
- [Salotti 94] Salotti, M. : Gestion des informations dans les premières étapes de la vision par ordinateur. Thèse de Doctorat en Informatique, Institut National Polytechnique de Grenoble, 25 janvier 1994.
- [Spinu 96] Spinu, C., Chassery, J.M. & Garbay, C. : Détection de contours par Estimation et Minimisation des Erreurs, Actes du 10ième Congrès "Reconnaissance des Formes et Intelligence Artificielle", AFCET / AFIA, Vol. 2, pp. 1052-1062, 1996.
- [Thonnat 93] Thonnat, M., Clément, V. and Van den Elst, J. : Supervision of perception tasks for autonomous systems : the OCAPI approach, Rapport de Recherche INRIA N° 2000, Juin 1993.
- [Zighed 96] Zighed, D.A. and Sebban, M. : Cognitive model for the identification of odors, IPMU'96, Grenade, juillet 1996.