

MESSIE : un système multi spécialistes en vision.

Application à l'interprétation en imagerie aérienne

*MESSIE : a Multi Expert System in Computer Vision.
Application for Aerial Imagery Interpretation*



Gérard GIRAUDON

INRIA,
2004 route des Lucioles, B.P. 93
06902 Sophia Antipolis Cedex

Gérard Giraudon est né le 14 mars 1954. Il a soutenu sa thèse de doctorat en Mathématiques Appliquées en 1979 à l'Université de Nice. Il a travaillé chez IBM (Laboratoire Recherche et Développement

de La Gaude) de 1976-1979, puis chez Sintra-Alcatel de 1980-1984 (Groupe d'Études Avancées), avant de rejoindre en 1984 le projet Analyse de Scène de l'INRIA, U.R. de Sophia Antipolis, où il est aujourd'hui Directeur de Recherche (projet Pastis). Ses centres d'intérêt, dans le domaine de la Vision par ordinateur, sont assez larges et comprennent entre autre les domaines liés à la Segmentation d'Image, à la modélisation et à la représentation des connaissances en Vision ainsi que les domaines liés au parallélisme en Vision.



Philippe GARNESSON

INRIA,
2004 route des Lucioles, B.P. 93
06902 Sophia Antipolis Cedex

Philippe Garnesson a présenté une thèse en Informatique à l'université de Nice-Sophia Antipolis en 1991. Domaines d'intérêt : interprétation en imagerie aérienne et plus récemment participe à un projet CNES concernant un véhicule planétaire (VAP).



Philippe MONTÉSINOS

LERI/EERIE
École pour les Études et la Recherche
en Informatique et Électronique
Parc Scientifique G. Besse
F-30000 Nîmes
Téléphone (33) 66 38 70 21
email : montesin@eerie.eerie.fr

Philippe Montésinos est responsable de l'équipe Vision par Ordinateur du Laboratoire d'Études et de Recherche en Informatique de Nîmes (LERI). Il est diplômé de l'Université de Nice Sophia-Antipolis (Doctorat en Sciences de l'Ingénieur, effectué à l'INRIA Sophia-Antipolis en 90). Il est Maître de Conférences à l'École pour les Études et la Recherche en Informatique et Électronique à Nîmes (EERIE) où il assure actuellement des cours de Vision par Ordinateur. Ses travaux de recherche portent essentiellement sur l'Analyse de Scènes et le Groupement Perceptif.

RÉSUMÉ

Ce travail s'inscrit par le cadre du développement méthodologique d'une architecture multi spécialistes, appelée MESSIE, pour l'interprétation d'images. Nous présentons, dans cet article, les spécifications du système pour l'interprétation d'objets, en particulier ceux faits de main d'homme dans l'imagerie aérienne. La principale difficulté de ce problème est

l'expression des connaissances nécessaires à l'interprétation : les connaissances liées aux Objets, à la Stratégie et à la Scène. Nous montrons comment une modélisation des objets physiques de la scène, exprimée sous quatre points de vue (géométrie-radiométrie (forme-aspect), relation spatiale entre objets (contexte) et fonction (sémantique)) permet d'aboutir

à la spécification d'une architecture. L'architecture MESSIE⁽¹⁾, de type tableau noir, est organisée sur ces quatre types de bases de connaissances, qui schématiquement se décomposent en deux niveaux. Le premier niveau correspond à la connaissance liée à la Scène et à la Stratégie et le deuxième correspond aux spécialistes (un par type d'objet). Chaque niveau n'accède qu'à certains des points de vue. Nous montrons la modularité d'une telle approche et la facilité de mise en place de diverses

stratégies et nous illustrons l'article avec les résultats obtenus par quatre spécialistes d'objets artificiels (routes, bâtiments, ombres et voitures) et par un spécialiste d'objet naturel (cours d'eau).

MOTS CLÉS

Vision par Ordinateur, Interprétation d'images, images aériennes.

ABSTRACT

This work is in the context of the methodological development of a multispecialist architecture called MESSIE. This paper presents the system specifications for the interpretation of man-made structures in the field of aerial imagery. The main difficulty of such a system is the knowledge expression necessary for the interpretation: among them, strategy, scene and objects knowledges. We show how the choosen modeling of physical objects described by four points of view allows to lead to the specification of an architecture. These points of view are geometry and radiometry modeling (shape-aspect), spatial relation of the objet with the others (context) and the fonctionnality (semantic). MESSIE is a blackboard architecture organized with four types of knowledge bases that schematically are grouped in two hierarchical levels. The first

level corresponds to the Scene and Strategy level and the second corresponds to the specialists (one for each objet). Each level works only with certain points of view. We show the modularity of a such approach and the facility to use different strategies. We illustrate this paper with results obtained with four specialists: Road, Building, Shadow and Vehicule for man-made objects and one specialist for natural object (river).

KEY WORDS

Computer Vision, Scene interpretation, aerial imagery.

1. Introduction

Cet article présente un système à base de connaissances, appelé MESSIE (Multi Expert System for Scene Interpretation and Evaluation), dans le cadre de l'interprétation d'images aériennes. Notre but est de s'intéresser aux aspects méthodologiques de l'analyse de scènes et aux problèmes liés à la modélisation de l'information, ainsi qu'au développement d'une architecture flexible et évolutive constituant une plateforme d'intégration de connaissances pour l'analyse de scènes.

Notre propos concerne donc simultanément la problématique de la Vision par Ordinateur et celle de l'Intelligence Artificielle, mais il se concentre surtout sur le lien entre ces deux techniques, lien réalisé par l'intermédiaire de la modélisation des connaissances.

On peut considérer que la réalisation d'un tel système nécessite l'étude des trois problèmes suivants :

- la modélisation des objets qu'il faut manipuler (objets du discours), en particulier celle liée aux objets physiques, à la scène et à la stratégie d'interprétation,
- la capacité des algorithmes de Vision utilisés à extraire de l'image les caractéristiques recherchées,
- l'architecture du système comme élément fédérateur des deux domaines ci-dessus pour la réalisation de l'application.

Notre choix est de considérer d'abord la modélisation des objets du discours, et plus particulièrement celle des objets physiques, afin d'obtenir les contraintes qui nous permettront de construire l'architecture la mieux adaptée, cette modélisation devant elle-même tenir compte du type et de la qualité des informations fournies par les algorithmes de Vision.

MESSIE, initialisé par le contrat Esprit n° 1588, a pour but de valider certains concepts de modélisation en développant un démonstrateur. Les concepts concernés par cette validation sont ceux qui nous semblent nécessaires en vue de la réalisation d'un système opérationnel de photo-interprétation automatique. La démarche, que nous avons entreprise, est illustrée pour une application particulière, la détection de bâtiments et de routes sur les images aériennes, qui est couramment utilisée par la communauté scientifique pour illustrer le lien entre la vision par ordinateur et l'intelligence artificielle.

Une analyse détaillée de la littérature sur ce sujet (voir ci-après et [18]), nous permet de définir les contraintes qu'il est nécessaire de prendre en compte, tant sur l'aspect modélisation que sur l'aspect architectural, pour la réalisation d'un tel démonstrateur :

- L'utilisation de modèle contraint des objets physiques n'est pas souhaitable ; il est donc nécessaire de définir des modèles génériques issus du sens commun : qu'est-ce qu'une route, un bâtiment ?
- Par contre, il est nécessaire d'explicitier une connaissance générique basée à la fois sur la connaissance géométrique et la connaissance radiométrique des objets à interpréter.
- La grande diversité et la complexité des objets imposent une organisation hiérarchique, particulièrement pour la modélisation géométrique, c'est-à-dire une décomposition

⁽¹⁾ This work is supported by the Esprit Project 1588.

arborescente par règles d'assemblage de primitives symboliques.

- Il est nécessaire d'avoir plusieurs bases de connaissances, une par type d'objet à interpréter.
- Le contrôle doit permettre l'émission d'hypothèses devant être vérifiées par des retours éventuels sur l'image originale.
- Il faut absolument éviter l'explosion combinatoire des hypothèses émises.
- Il faut définir une base de faits commune permettant la mise en place d'une stratégie globale au niveau de la scène, afin de permettre une exploitation du contexte.
- Le rôle de la stratégie est fondamental, mais celle-ci dépend des objets eux-mêmes. Commencer par des buts simples, afin d'avoir un comportement opportuniste en fonction des résultats, semble être une approche raisonnable.

MESSIE, développé à partir de ces contraintes, est un système multi spécialistes, bâti autour d'une architecture de type tableau noir (« blackboard »), dont l'objectif est de réaliser une interprétation basée uniquement sur une connaissance générique des objets sans utiliser de connaissances exogènes à la donnée image. Même si la réalisation effective du système est importante, puisqu'elle prouve la validité de la modélisation, les contributions principales, que nous présentons dans cet article, concernent d'une part la modélisation des objets physiques (notion de point de vue) permettant le découplage entre l'expression de la connaissance et la façon dont elle est utilisée, et d'autre part le développement de primitives spatiales permettant la création d'opérateurs généraux de localisation utilisant complètement la connaissance sur les relations spatiales des objets. La partie algorithmique, très liée au traitement de l'image, ne sera pas développée dans cet article et le lecteur est invité à se référer aux articles cités.

Cet article décrit tout d'abord les principaux travaux concernant l'interprétation d'images aériennes. La section 3 présente ensuite le problème de l'expression de la connaissance nécessaire pour l'interprétation d'une scène. La section 4 concerne la description de l'architecture MESSIE, qui permet la mise en œuvre de cette connaissance. Enfin notre travail est illustré par les résultats obtenus lors des différentes phases d'exécution pour le problème de l'extraction des routes, des bâtiments et des voitures.

2. Travaux sur le domaine

Il existe une large littérature depuis le début des années 80 sur le problème du lien entre la vision par ordinateur et l'intelligence artificielle dans le cadre applicatif de l'analyse d'images aériennes. Sans vouloir être exhaustif, nous donnons ci-après les principales références qui ont servi à notre analyse et nous ont permis de définir les contraintes présentées dans la section précédente.

Tout d'abord en ce qui concerne la présentation des systèmes pour la représentation des connaissances, on peut citer l'article de synthèse de Binford [1] sur les systèmes d'analyse d'images utilisant de la connaissance, les travaux de Levine et Nazif [44] sur l'utilisation d'un système expert pour la segmentation en régions, les travaux de l'équipe de Riseman sur le système d'analyse de scènes VISIONS [23] et en particulier l'application hiérarchique de la coopération contours-régions [47] pour résoudre le problème de l'interprétation des images aériennes d'aéroports. Pour ces mêmes images, les travaux de McKeown [39] sont intéressants, car ils mettent en œuvre une base de connaissances avec contrôle hiérarchique construisant des objets de plus en plus élaborés en faisant intervenir une connaissance explicite à chaque niveau. McKeown utilise plusieurs sources d'information, images multispectrales, images stéréo, et il se donne un modèle théorique d'aéroport sur lequel il vient projeter (cohérence spatiale et mesure de distance) les hypothèses d'objets. Il y a encore les travaux de Nagao et Matsuyama [43] sur l'interprétation de scènes suburbaines dans une imagerie multispectrale, repris et développés dans le système SIGMA [38], où Matsuyama *et al.* proposent une architecture multi experts (géométrie, sélection de modèles, traitement d'images), où le chemin de communication entre experts est prédéfini. La représentation des objets physiques est basée sur un langage orienté objet : une instance d'objet étant considérée comme un agent actif qui met en œuvre des règles de production. On peut également citer l'intéressant travail de Nicolin et Gabler [45] qui proposent un système d'interprétation à contrôle mixte (guidé par les données et par les buts) et une stratégie par raffinement itératif. La connaissance est décrite au moyen de règles et de méta règles et un réseau sémantique est utilisé pour la modélisation des objets et des relations entre ces objets. La description des objets (ou nœuds du réseau) se fait de manière hiérarchique (part de, composé par, ...).

Plus liées aux problèmes de Vision pour l'extraction des objets, on peut citer les recherches de Lowe sur le groupement perceptuel [35], ou encore les travaux portant sur la détection des bâtiments et leur aspect 3D [33] [41] [36] [29] [16] ou ceux portant sur la détection des routes [30] [40] [15].

La communauté française s'est également intéressée au problème de l'interprétation d'images aériennes et on peut citer la contribution de Cocquerez *et al.* [7] [49] [19], qui présentent, dans leur dernier article, un système d'interprétation d'images aériennes avec un mélange efficace de techniques neuronales et d'un système à base de règles. Dans le cadre de l'interprétation d'images satellites, Desachy [11] propose un système expert à base de règles, qui prend en compte des informations exogènes à l'image, du type informations géocodées ou bien encore informations sur le relief, pour réaliser une classification complète de l'occupation des sols.

Enfin le lecteur pourra également consulter avec intérêt l'article de Henri Maitre [37] pour son étude synthétique de ces domaines d'applications et par sa présentation bibliographique.

3. Problèmes de modélisation

Comme nous l'avons souligné plus haut, une des grandes difficultés de l'interprétation de scènes est l'expression de la diversité de la connaissance liée aux objets du discours. En fait les connaissances a priori que l'on doit posséder appartiennent à deux groupes :

- connaissances sur le monde des objets réels : capteurs et objets physiques,
- connaissances sur le processus d'interprétation : stratégie et contrôle, scène interprétée.

3.1. MODÉLISATION DES OBJETS PHYSIQUES

Dans le cadre de notre application, nous supposons que nous disposons uniquement d'une information générique sur les objets présents dans l'image : nous ne connaissons précisément ni leur taille ni leur forme. Dans ces conditions, le problème de la modélisation est un problème particulièrement complexe et pour lequel nous ne présentons que des éléments de réponse. Une hypothèse simplificatrice est de considérer les objets de la scène comme des objets à deux dimensions (seule leur projection au sol est prise en considération). Dans le cadre de notre application cette hypothèse est suffisante : l'aspect 3D des objets a peu d'influence dans le cas de vue aérienne.

Nous avons choisi, en étudiant le problème de la modélisation des objets, de rendre celle-ci indépendante des capteurs utilisés, en vue d'une application future dans le cadre de la fusion multisensorielle. Bien qu'il y ait une énorme différence d'aspect, de forme et de taille, ..., entre les objets de la scène, nous souhaitons un formalisme suffisamment général et homogène, qui permette d'incorporer des connaissances a priori, même si ces connaissances peuvent varier fortement d'un objet à l'autre. Dans la littérature, la modélisation des objets est généralement peu détaillée et sa description se limite souvent à une représentation hiérarchique (de type composé de, part de, ...) [45] [39] [47] éventuellement associée à des règles de production [38].

Une structuration selon quatre points de vue

Nous proposons, pour répondre à ces considérations, un formalisme par points de vue, implémentable avec un langage orienté-objet. Les points de vue sont au nombre de quatre, chacun est indépendant des autres et peut être vide si aucune connaissance a priori n'existe :

1. **La forme** : Ce point de vue caractérise la forme **géométrique** de l'objet. Cette forme est « générique » dans la mesure où elle modélise des objets qui dans la réalité peuvent sembler différents (variation de forme, dimension, orientation, ...). Dans le cas d'objets complexes, le modèle est décomposé en parties (ou encore objets) plus élémentaires qui facilitent la caractérisation.

Pour notre application, bien que nous nous limitons à des modèles 2D, la caractérisation de la forme géométrique

reste relativement difficile étant donné la variété des objets que nous avons à représenter : certains sont artificiels (bâtiments, routes, ...) alors que d'autres sont « naturels » (rivières, forêts, ...). Pour prendre en compte cette diversité, nous avons essentiellement utilisé trois modes de représentation :

- une représentation basée sur la notion de ruban (objet linéique : les routes et les rivières)
- une représentation polygonale pour les objets compacts et bien définis géométriquement comme par exemple les bâtiments, les champs et les ombres,
- une représentation « surfacique » (régions) pour les objets dont la forme est difficilement modélisable par des segments, comme par exemple les arbres, les forêts, ...

Néanmoins, il faut noter que cette classification est purement formelle : aucune contrainte n'est imposée pour définir un objet.

2. **L'aspect** : Sous la terminologie « aspect », nous voulons caractériser la connaissance sur l'apparence visuelle des objets, liée à l'aspect **radiométrique** dans l'image. Cela se traduit par un adjectif (clair, foncé...) et par une mesure d'homogénéité (écart-type autour de la moyenne). Cette caractérisation de la radiométrie peut être globale ou dépendante des différentes parties dont l'objet est composé. Pour être plus générique et s'affranchir de la dépendance par rapport aux capteurs, ce point de vue doit être étendu par un point de vue matériau, que nous avons commencé à réaliser [3].

3. **Le contexte** : Ce point de vue est relatif à la connaissance que l'on a a priori des **relations spatiales** qu'un objet peut avoir avec d'autres objets : une ombre est à côté d'un bâtiment, une voiture est sur une route, un pont est sur une rivière. Il se traduit par une liste « d'heuristiques de localisation », définies par deux objets et une relation spatiale. Cette connaissance, utilisée pour guider l'analyse de la scène, n'est pas liée au point de vue géométrie ou radiométrie. Elle reste vraie, quelle que soit la modélisation de la route ou de la voiture, et quel que soit le capteur utilisé. Dans notre cas, le contexte, mis en œuvre par un contrôleur de localisation (voir section suivante), permet soit de valider un objet par l'existence d'autres objets déjà détectés dans son contexte réel soit d'inférer une nouvelle hypothèse d'objet.

4. **La fonction** : Ce point de vue correspond à la fonctionnalité de l'objet (par exemple, une route est une « voie d'accès »). Ce point de vue, non utilisé dans notre application, doit être cependant pris en compte dans le cadre d'une proposition générale de modélisation de part l'importance sémantique d'un tel concept. Son utilisation complète est cependant non triviale et mérite des efforts soutenus.

Pour l'implémentation, la modélisation des objets repose sur l'utilisation d'un langage orienté objet créé pour cette application à partir de MicroCeyx [32], lui-même basé sur le langage Le-Lisp [2]. Actuellement cinq types d'objets sont modélisés et décrits dans la section suivante. Il s'agit des routes, des bâtiments, des rivières, des ombres et des voitures.

3.2. CONNAISSANCES SUR LE PROCESSUS D'INTERPRÉTATION

3.2.1. Représentation de la Scène interprétée

Le but d'un système d'interprétation est de construire un modèle sémantique de la scène observée. Ce modèle doit déterminer la position exacte des objets détectés et les caractéristiques liées à ces objets. Il faut par exemple que le système puisse répondre à une question du type : « quels sont les objets qui se trouvent « le long » de la route N° 28 ? ». Nous avons choisi, comme dans [45], de modéliser la scène interprétée par une matrice de localisation, décrite ultérieurement, dont chaque élément est un pointeur vers un objet validé.

3.2.2. Connaissances liées à la stratégie et au contrôle

La densité d'information contenue dans une image aérienne est trop importante pour une exploitation non hiérarchique. Selon l'approche pyramidale, utilisée classiquement en Vision, on considère que les pixels de l'image n'ont pas tous la même qualité informationnelle. Cette idée doit être transposée par MESSIE au niveau des objets qui composent la scène. Il semble en effet illusoire de considérer que tous les objets apportent la même quantité d'information vis-à-vis de l'interprétation ultérieure. Si l'objectif est de « chercher les routes et les bâtiments », il semble plus important de détecter en priorité une autoroute plutôt qu'un chemin de terre. Derrière la sémantique des objets, c'est le caractère « exceptionnel » et/ou « remarquable » que nous voulons utiliser. La plupart du temps, ces caractéristiques se traduisent au niveau de l'interprétation humaine sous le vocable « objet évident ». Un objet évident est un objet que l'œil humain voit immédiatement dans la scène. Il est en général macroscopique, à géométrie simple, homogène dans sa radiométrie, et bien contrasté par rapport à son environnement. La connaissance d'objets identifiés dans la scène permet alors de guider plus finement la recherche en émettant des hypothèses précises avec des contraintes fortes. Cela peut être illustré facilement par les exemples suivants :

- si deux bâtiments semblables et topologiquement proches sont déjà trouvés, alors rechercher en priorité, dans la même zone, un bâtiment ayant des caractéristiques semblables (orientation, taille, ...)
- si une hypothèse route est émise alors rechercher des véhicules sur la route.

On peut trouver une idée similaire dans [45], où les auteurs implémentent la détection des objets « remarquables » par une analyse pyramidale de l'image.

La stratégie et le contrôle doivent en fait gérer trois tâches principales :

- Définir et ordonner les buts à atteindre. Ceci est mis en œuvre dans les phases itératives d'interprétation selon des connaissances a priori (exemple : critères d'ordre) exprimées sous forme de règles de production. Ces connaissances définissent en fait la stratégie.
- Valider les hypothèses d'objets trouvés en recherchant des objets du contexte qui, soit ont déjà été trouvés dans la

scène interprétée, soit sont considérés comme nouveaux buts à atteindre.

- Gérer les conflits d'interprétation qui surviennent lorsque un élément matriciel de la scène interprétée pointe sur plusieurs objets. Le conflit peut être virtuel (un objet est dans le contexte d'un autre avec une relation « sur ») ou réel (les deux objets sont incompatibles).

Nous voyons ici deux caractéristiques fondamentales du système :

- Le caractère opportuniste de la stratégie.
- La nécessité de maîtriser la connaissance entre objets (type relation topologique) et la mise en place de mécanismes permettant la mise en œuvre de cette connaissance (opérateurs) au niveau Scène interprétée et au niveau image à interpréter (pixel).

Compte tenu des spécifications fournies dans cette section, nous nous sommes orientés vers une architecture de type tableau noir, que nous justifions ci-après.

4. Description de l'architecture

L'architecture de MESSIE repose sur le modèle du tableau noir [25] [26] [46] [8] [27] [34] dont l'intérêt, dans le domaine de la reconnaissance de formes, a été démontré à travers de nombreux travaux [23] [43] [24] [13] [14] [20] [12]. Cette approche est, en effet, particulièrement adaptée au problème de l'interprétation qui peut être considérée comme le résultat de la coopération de différents spécialistes. Cette approche est également essentielle du point de vue génie logiciel : elle facilite l'adjonction et la modification des bases de connaissance.

Néanmoins, notre propos ne porte pas sur les mécanismes propres aux tableaux noirs (qui doivent être considérés ici comme des outils), mais sur leur utilisation dans le cadre de l'analyse de scène. Dans MESSIE, les spécialistes sont hiérarchisés en deux niveaux (voir fig. 1) :

- **le niveau scène** qui permet de guider l'interprétation de la scène : c'est lui qui décide quels objets chercher et dans quelles zones de l'image. Il a une vision globale de l'interprétation et est capable de focaliser la recherche en fonction d'îlots de confiance ;
- **le niveau objet** : il se traduit par un ensemble de spécialistes, chacun d'entre eux étant spécialisé dans la détection d'une classe d'objets.

Étant donné l'ampleur du problème, une séparation claire entre ces deux niveaux de compétence est indispensable. En effet, alors que le niveau objet est proche du domaine d'application et donc concerné par les techniques de traitement d'images, le niveau scène est quant à lui, plus lié au domaine de l'Intelligence Artificielle. Il fait intervenir des concepts liés à la stratégie d'interprétation, indépendamment des techniques de reconnaissance utilisées. Par ailleurs, du point de vue complexité, cette décomposition en deux niveaux permet de rester efficace : la stratégie mise en œuvre au niveau scène ne porte que sur les objets ayant une signification d'un point de vue interprétation.

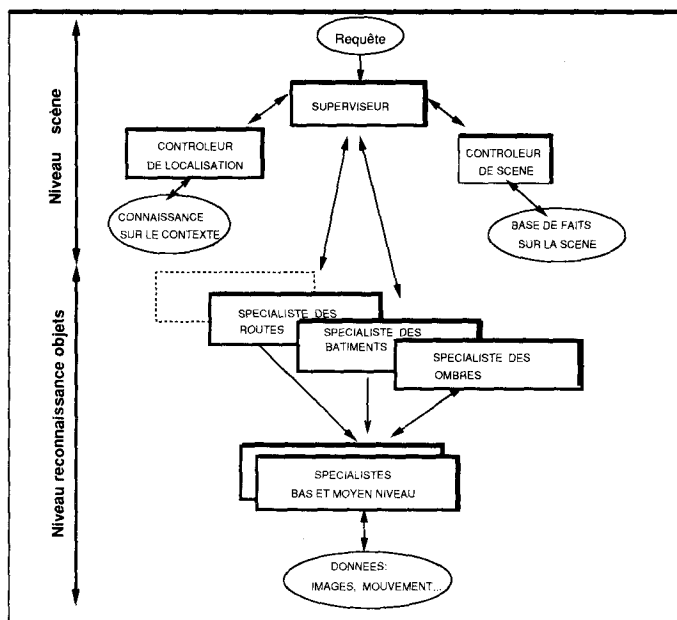


Figure 1. — MESSIE, Architecture fonctionnelle, deux niveaux de spécialistes : le niveau scène et le niveau objet.

Dans la suite de l'article, nous décrivons de façon plus précise les différents modules intervenant à chacun des deux niveaux. Une description plus fonctionnelle de l'ensemble des différents modules est présentée en dernière partie à travers un exemple concret.

4.1. LE NIVEAU OBJET

Le niveau objet correspond à un ensemble de spécialistes, chacun d'entre eux étant associé à la reconnaissance d'une classe d'objets (routes, bâtiments, ...).

Il faut noter que dans MESSIE ces spécialistes n'ont aucune autonomie vis-à-vis du processus d'interprétation de la scène. L'activation d'un spécialiste est conditionnée par la réception d'une requête émise par le niveau scène. Le spécialiste se contente uniquement d'accéder au « tableau noir » pour y placer les hypothèses d'objet correspondant à la requête traitée.

La solution qui consiste à laisser chaque spécialiste participer librement à l'interprétation ne nous semble pas viable : la complexité du problème conduirait rapidement à une explosion du nombre d'hypothèses. L'activité du spécialiste doit s'insérer dans une stratégie globale liée à l'analyse de la scène (voir exemple section 5).

Une requête de recherche d'objet spécifie la zone d'activation du spécialiste (focalisation sur une partie de la scène), ainsi que les valeurs des paramètres, qui lui permettent de s'adapter à l'état ponctuel de l'interprétation. Ce type de spécialiste n'utilise donc pas directement le contexte de compréhension des objets. Cette approche permet d'assurer la modularité du système.

Ces spécialistes ont une organisation hiérarchique : on distingue les spécialistes qui émettent des hypothèses d'objets au niveau de la scène et les spécialistes qui

correspondent aux traitements de bas et moyen niveaux. Ces derniers peuvent être assimilés à des « procédures » alors que les premiers réagissent à des requêtes émises par le superviseur (voir partie suivante).

Ces requêtes ont pour but de faire émettre, par ces mêmes spécialistes, des hypothèses d'objets qui sont cohérentes avec la modélisation effectuée des objets (points de vue « forme et aspect »). Seules ces hypothèses sont transmises au niveau du tableau noir. Les informations intermédiaires, qui ont permis l'élaboration de ces hypothèses, restent locales aux spécialistes.

Dans MESSIE, les différents spécialistes peuvent correspondre à des sources de connaissance exprimées dans un formalisme hétérogène (certains sont basés sur des règles, d'autres sur un programme C).

Un « encapsulage » dans un formalisme commun rend possible la coopération des spécialistes. Pour une description des techniques de traitement d'images que nous avons utilisées, nous renvoyons le lecteur à [42] [18].

4.2. LE NIVEAU SCÈNE

Le niveau scène (voir fig. 1) regroupe la connaissance sur la stratégie d'interprétation. Cette connaissance est répartie dans trois modules dont la terminologie (superviseur et contrôleurs, en opposition aux spécialistes objets) est empruntée au système HASP [26].

Alors que deux d'entre eux sont spécifiques à l'analyse de scène (le contrôleur de scène et le contrôleur de localisation), le troisième (le superviseur), situé au sommet de notre pyramide, est plus spécifique à l'architecture de tableau noir : son rôle est de sélectionner la base de connaissances (contrôleurs ou spécialistes) qui doit être activée.

Le superviseur

Dans MESSIE, le superviseur est un mécanisme qui centralise toutes les requêtes (voir fig. 2) émises par les

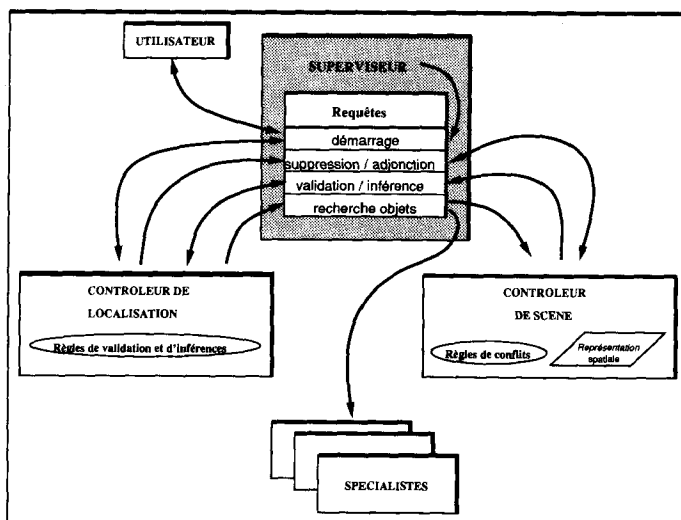


Figure 2. — Le superviseur : ordonnancement et répartition des requêtes.

autres spécialistes. En fonction de ces requêtes, la tâche du superviseur consiste à sélectionner le ou les spécialistes capables d'y répondre. En fait, le superviseur a un rôle décisionnel limité : le type des requêtes détermine implicitement si c'est le contrôleur de scène, le contrôleur de localisation ou un des spécialistes de reconnaissance d'objets qui doit répondre.

Dans ce dernier cas, un problème de choix peut se poser si plusieurs spécialistes sont associés à une classe d'objets. Le superviseur se contente alors d'utiliser des informations de contrôle qui sont transmises dans la requête (voir description ci-après du contrôleur de localisation et fig. 3).

<p>Déclenchement</p> <p>Objet de déclenchement (OD)</p> <p>préconditions</p> <p>Objectif : validation ou inférence</p> <p>Probabilité de succès</p>
<p>Contrôle</p> <p>Choix des spécialistes</p> <p>Objet Recherché (OR)</p> <p>consultation de la scène: autorisée, exclusive, interdite</p> <p>mode d'activation</p> <p>paramètres des spécialistes</p> <p>zone de recherche</p> <p>paramètres de fonctionnement</p> <p>contraintes sur les résultats</p>
<p>Actions</p>

Figure 3. — Structuration d'une règle de contexte spatial en trois parties, elles expriment des relations topologiques entre différents objets (ou parties d'objet).

L'importance du superviseur réside dans l'ordre du traitement des requêtes. Cet ordre est étroitement lié à la stratégie d'interprétation retenue, à savoir détecter les objets les plus remarquables de l'image puis s'appuyer sur cette connaissance afin d'étendre l'interprétation au reste de l'image.

En pratique, le superviseur traite en priorité les requêtes ayant trait à la recherche d'objets remarquables (requête de démarrage), puis les requêtes visant à étendre les îlots de confiance (requête d'inférence), et pour finir les requêtes visant à confirmer des hypothèses peu fiables (requête de validation). Le choix des zones sur lequel le système focalise son activité reste à la charge du contrôleur de localisation.

Le contrôleur de localisation

L'activité du contrôleur de localisation repose sur l'utilisation de règles (voir fig. 3), appelées « heuristiques de localisation », qui expriment des relations topologiques entre différents objets (ou parties d'objet). Le contrôleur de localisation s'appuie donc sur la connaissance décrite par le

point de vue contexte dans la modélisation des objets. Par exemple, lors de la modélisation des routes, le point de vue concernant le contexte contient l'information suivante : « il y a souvent des voitures sur une route ».

Lors de la détection d'un objet par le système, l'instance de l'objet créé hérite automatiquement d'un ou plusieurs champs correspondant aux règles associées à la classe de cet objet.

Ces informations (objets détectés et règles associées) sont exploitées par le contrôleur de localisation pour générer une nouvelle requête de recherche d'objet qui permet de répondre aux questions suivantes :

1. « **Quel objet chercher ?** » : En fonction de la requête (démarrage, validation ou inférence) qui a activé le contrôleur de localisation, celui-ci adopte une stratégie différente pour choisir l'objet à rechercher.

Dans le cas d'une requête d'inférence, un îlot de confiance est recherché. En pratique, le choix est conditionné par l'existence d'un objet détecté (**objet de déclenchement**, voir fig. 3) ayant un coefficient de vraisemblance élevé et pour lequel il existe une règle non utilisée dont l'**objectif** (voir fig. 3) est l'inférence.

Dans le cas d'une requête de validation, on s'oriente cette fois vers une règle dont l'objectif est la validation.

Dans le cas particulier d'une requête de démarrage, on considère des règles qui sont associées à un objet particulier : l'image (par exemple « chercher des routes dans l'image »). Lorsque plusieurs règles sont potentiellement activables, c'est celle dont la **probabilité de succès** est la meilleure (coefficient fixé a priori par l'utilisateur) qui est choisie.

2. « **Dans quelle zone de l'image ?** » : La requête créée hérite d'informations de contrôle qui sont spécifiées lors de la création de la règle. Parmi celles-ci, l'information **zone de recherche** permet de focaliser l'activité des spécialistes dans une zone réduite de l'image. Dans ce cadre, nous avons développé des outils facilitant l'expression de cette zone de recherche qui s'exprime par des concepts tels que « sur », « le long de », « entre », « autour ». Cette approche n'a été possible que moyennant un effort important pour caractériser de façon uniforme la représentation spatiale des objets [17].

Cela signifie que le système est capable de définir automatiquement une zone de recherche en fonction de la nature des objets mis en cause et de la relation (par exemple « entre ») spécifiée par l'opérateur.

3. « **Comment chercher ?** » : Les informations de contrôle permettent également de guider les méthodes de construction des objets. Par exemple, si plusieurs bâtiments de formes et de directions similaires ont été détectés, ces caractéristiques peuvent être utilisées pour extraire des bâtiments semblables.

En conclusion, ces règles permettent d'unifier dans un même formalisme les principaux éléments de stratégie qui interviennent lors de l'interprétation. Il faut également noter que l'expression de la connaissance représentée dans ces règles est totalement dissociée de leur utilisation par les spécialistes : elles correspondent à une connaissance de

haut niveau qui peut être fournie au système sans avoir à se soucier des méthodes qui seront mises en œuvre pour les exploiter.

Le contrôleur de scène

Le rôle du contrôleur de scène est de mettre à jour (adjonction/suppression) une base de données spatiales qui reflète la localisation dans la scène des objets identifiés ainsi que les relations qui peuvent exister entre ces derniers (voir fig. 4). Cette base de donnée est construite au fur et à mesure de l'interprétation, néanmoins rien n'exclut de l'initialiser avec, par exemple, un système d'information géographique (SIG) [48].

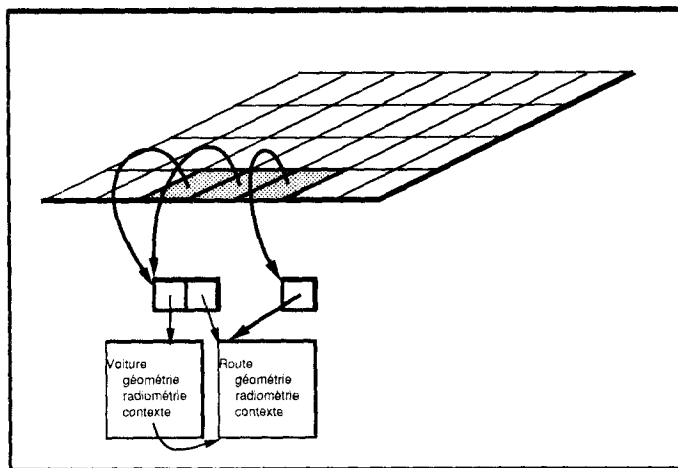


Figure 4. — L'accès aux objets de la scène : une représentation matricielle qui facilite la vérification de la cohérence spatiale.

Le concept de contrôleur est justifié par deux activités complémentaires :

1. Le contrôleur de scène est habilité à répondre, au même titre qu'un spécialiste de détection d'objet, à une requête de recherche d'objet. En effet, si le module de stratégie émet une requête visant à augmenter la vraisemblance d'une hypothèse, par exemple chercher une ombre près d'un bâtiment, il est souhaitable de vérifier parmi les objets détectés si l'un d'entre eux ne satisfait pas la requête. Il est en effet moins coûteux d'accéder à la base de données plutôt que faire appel à un processus vision souvent coûteux.

Cette possibilité de consultation présente également un intérêt pour l'utilisateur qui peut de façon analogue sélectionner certains objets sous contraintes : par exemple chercher les objets qui sont situés le long d'une route donnée.

2. Le contrôleur de scène intervient également pour vérifier, à chaque adjonction d'objet, la cohérence spatiale de la scène.

Cette activité est essentielle pour deux raisons : d'une part, elle permet d'intégrer au système un mécanisme de détection des conflits, d'autre part, elle présente un intérêt

certain pour permettre au système d'assurer la fusion d'hypothèses issues de méthodes différentes. Ce principe est fondamental s'il existe plusieurs spécialistes chargés de détecter la même classe d'objets, spécialistes qui s'appuient sur des méthodes différentes ou sur des données différentes dans un contexte multi-capteurs. En pratique, une fusion se traduit par la création d'une relation entre les deux objets et d'une mise à jour du coefficient de confiance associé à ce groupement.

La représentation matricielle (voir fig. 4), que nous avons retenue facilite la vérification de la cohérence spatiale : il faut tester si un élément matriciel pointe sur plusieurs objets. La détection d'une situation conflictuelle reste néanmoins un problème difficile qui ne se traduit pas nécessairement par un recouvrement d'hypothèses (par exemple si on détecte une ombre le long d'une hypothèse de champ). Réciproquement certains recouvrements correspondent à des conflits virtuels (par exemple, une voiture sur une route).

Pour permettre l'identification et la résolution des conflits, nous nous sommes orientés vers une solution qui consiste à fournir au système une liste des situations anormales : on évite ainsi certains conflits virtuels. D'un point de vue formalisme, nous utilisons des règles dont la partie précondition permet de caractériser le conflit et dont la partie action correspond à un plan de résolution du conflit qui peut conduire à l'élimination d'une hypothèse ou la gestion d'alternatives.

Il faut noter que ce problème difficile reste un axe de recherche pour lequel une solution générale reste difficile :

- Nous nous sommes limités à la détection de situations conflictuelles mettant en cause le recouvrement de deux objets. Cette vision, bien que simplifiée, n'est pas sans poser des difficultés : il faut en effet prendre en compte que le recouvrement n'est pas une information binaire. Il peut y avoir un chevauchement faible de deux hypothèses qui n'implique pas forcément l'existence d'un conflit.
- Si dans certains cas, l'utilisation d'une méthode spécifique peut être envisagée pour résoudre les conflits, dans d'autres cas, seules les informations contextuelles peuvent être utilisées. Dans ce cas, il faut noter que la recherche de ces relations contextuelles peut conduire au développement d'alternatives dont le coût en temps de calcul est difficilement prévisible.

Dans ce contexte, il nous semble souhaitable de laisser en suspens les conflits et de continuer l'interprétation à partir des situations les plus sûres. De cette façon, on peut espérer, dans une phase finale, résoudre les conflits à partir d'objets détectés par ailleurs.

5. Résultats sur différents cycles d'interprétation

Dans cette section, nous illustrons les résultats obtenus par MESSIE sur plusieurs exemples. Le but applicatif est l'interprétation dans une image aérienne, des routes, des bâtiments et des véhicules sur les routes en n'ayant qu'une

connaissance générique des objets. A titre d'exemple, nous étendons cette interprétation à celle des rivières.

Nous présentons les résultats obtenus par MESSIE sur des images dont la quantification est de 8 bits par pixel (dans le spectre visible). Le système est utilisé pour des images de taille 1 000 × 1 000 pixels avec deux résolutions possibles : 80 et 40 cm par pixel.

Tout d'abord, nous donnons à titre d'exemple l'expression des modèles des routes et des bâtiments. Puis nous présentons les résultats en deux étapes : la première correspond à la recherche des objets évidents, la deuxième correspond à l'exploitation du contexte.

5.1. MODÉLISATION DES OBJETS

Modélisation des routes

La modélisation des routes, que nous avons utilisée, suppose que les deux bords de la route sont visibles (pour plus de détails voir [18] [42]).

Les caractéristiques principales utilisées sont les suivantes :

- Pour la forme :
 - longueur : longue, moyenne, courte, (en fait, cette contrainte se traduit par une mesure en mètres)
 - forme : ruban (voir explications complémentaires dans le paragraphe suivant lors de la mise en œuvre de cette modélisation),
 - * ruban : composition de tronçons non connexes,
 - * tronçon : succession de boîtes connexes,
 - * boîte : créée à partir de deux segments parallèles.
- Pour l'aspect (radiométrie) :
 - Un tronçon est supposé être de radiométrie homogène. Cela se traduit par :
 - * un intervalle des niveaux de gris admissibles sur l'image,
 - * un seuil sur l'écart-type entre ces niveaux de gris pour le choix initial des boîtes, qui sont les germes des tronçons.
 - Le contraste entre la radiométrie de la route et son environnement doit être élevé. Cela se traduit par l'existence de forts gradients pour au moins une des boîtes du tronçon et donc par la possibilité de création de boîtes calculées à partir de segments issus de l'approximation polygonale de chaînes de contour [21].
- Pour le contexte :

Les informations précédentes sont généralement suffisantes pour détecter les routes qui apparaissent comme évidentes à l'œil humain. Mais cela ne suffit pas pour l'ensemble des routes présentes dans l'image. Aussi traduisons-nous dans le contexte potentiel, un ensemble de connaissances souvent disparates, comme par exemple le fait que les routes sont souvent situées près des bâtiments, que les voitures se trouvent sur les routes et que les routes sont connexes aux bords de l'image.

Modélisation des bâtiments

Pour les bâtiments, la modélisation utilisée s'appuie sur les informations suivantes :

- Pour la forme :
 - Forme : polygonale,

- polygone simple : rectangle,
- polygone complexe : composition de rectangles,
- rectangle : créé à partir d'un U et d'un segment,
- U : créé à partir d'un angle droit et d'un segment,
- angle droit : créé à partir de deux segments.

- Pour l'aspect (radiométrie) : clair ou foncé. Cette contrainte se traduit par un intervalle de niveau de gris admissible.

- Pour le contexte :

Nous avons établi dans le contexte d'un bâtiment essentiellement deux faits. Premièrement, que l'existence d'un bâtiment est accompagnée de l'existence d'une ombre. Deuxièmement, que l'existence d'un bâtiment particulier signifie souvent l'existence d'autres bâtiments de mêmes caractéristiques dans son voisinage.

5.2. RECHERCHE D'OBJETS ÉVIDENTS

Lors de l'activation du système, le processus d'interprétation ne dispose d'aucune information sur la localisation ou la forme des objets présents dans la scène. Les seules informations connues portent sur les caractéristiques de l'image et sur la direction du soleil. Le superviseur envoie donc une requête au contrôleur de localisation lui demandant d'inférer des objets évidents. Pour cela, le contrôleur de localisation utilise les règles de « démarrage » (voir paragraphe décrivant le contrôleur de localisation). Actuellement ces règles sont les suivantes :

- *Chercher des routes dans l'image.*

Pour réduire l'espace de recherche (dans l'optique de la recherche d'objets remarquables), nous spécifions des contraintes sur la longueur des routes recherchées (« des routes longues »).

- *Chercher des bâtiments dans l'image.*

D'une façon analogue à la règle précédente, l'espace de recherche est cette fois limité en recherchant des objets « clairs ».

- *Chercher des rivières dans l'image.*

D'une façon analogue aux routes, l'espace de recherche est limité en imposant des contraintes sur la longueur.

Les routes

La modélisation des routes précédemment évoquée (connaissances géométrique et radiométrique) est exploitée par un spécialiste utilisant un processus de groupement perceptuel. Ce processus est initialisé par la recherche de boîtes (segments parallèles) qui délimitent une région « homogène ». De nouvelles boîtes sont alors construites dynamiquement pour étendre la route et former ainsi des tronçons. Ces tronçons sont ensuite connectés pour reconstituer le réseau routier.

Les primitives de bases, utilisées par ce spécialiste, sont des segments de droite obtenus à partir d'un algorithme de détection de contours [10].

La figure 5 présente une image initiale (résolution 80 cm par pixel) sur laquelle le spécialiste « route » est activé. Son activité repose sur les segments de la figure 6. Il faut

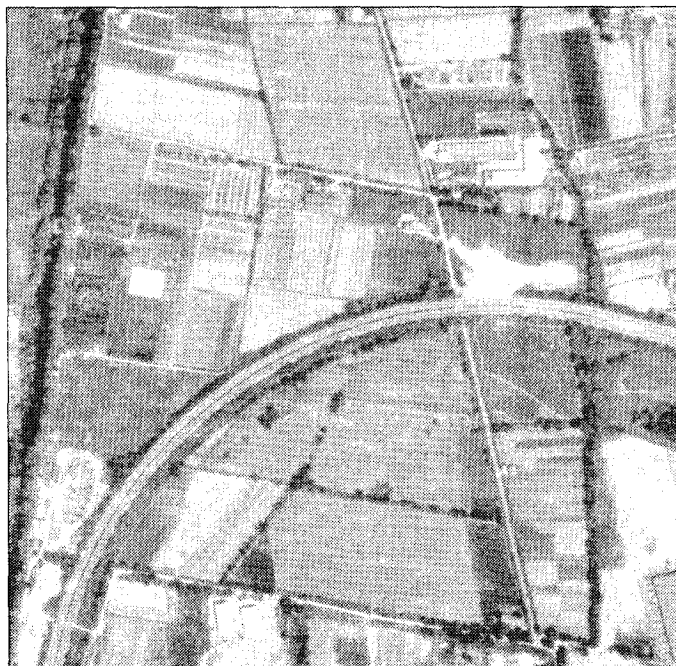


Figure 5. — Image initiale.

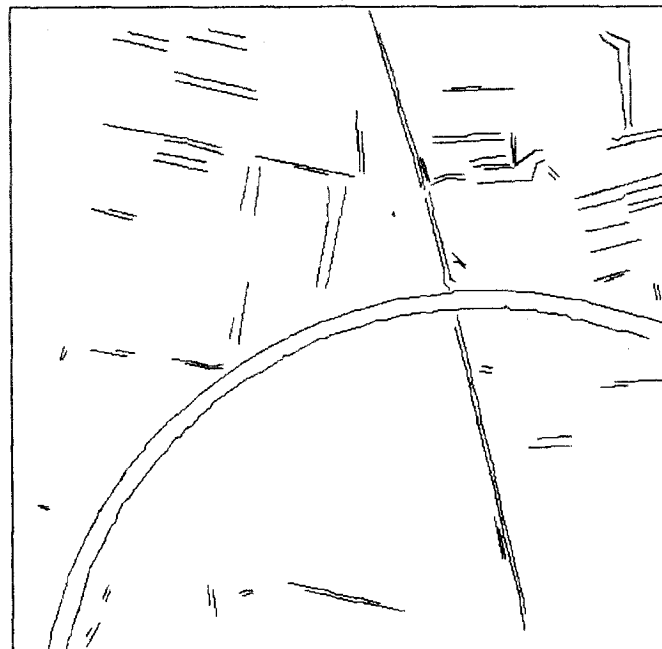


Figure 7. — Les tronçons détectés.



Figure 6. — Segments extraits à partir de l'image 5.

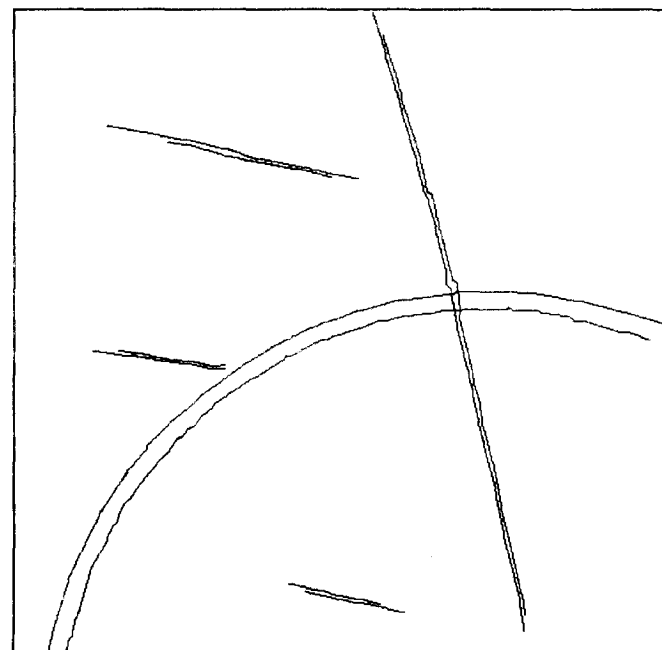


Figure 8. — Amélioration par connexion/suppression.

noter que ces informations restent locales au spécialiste et ne sont pas transmises au niveau scène. A partir de ces segments le spécialiste construit des hypothèses de rubans (fig. 7) puis les améliore par une étape de connexion (fig. 8). Seules les meilleures hypothèses sont transmises à la scène. L'espace de recherche, pour la localisation des

routes évidentes, est limité par l'utilisation d'une contrainte sur la longueur des objets recherchés. En pratique, la construction d'une route repose obligatoirement sur l'utilisation d'une chaîne de contour suffisamment longue.

Les résultats de cette étape sont présentés sur deux autres images :

- La figure 9 correspond à une zone suburbaine.

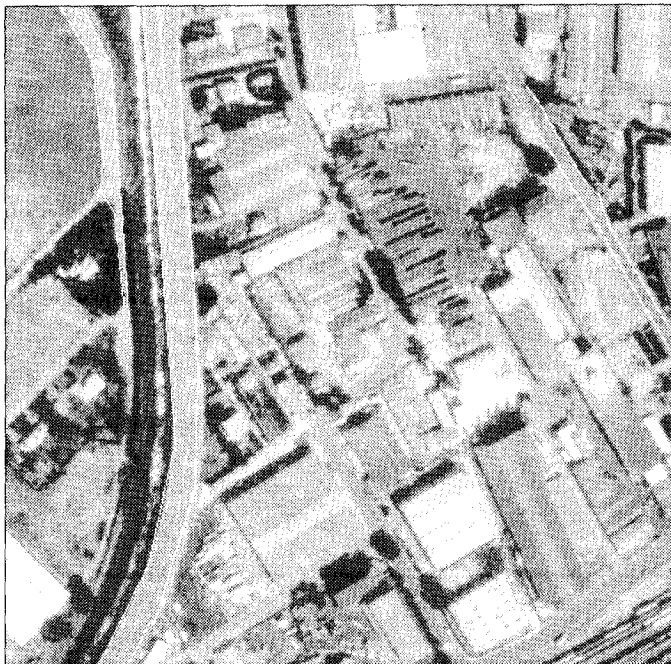


Figure 9. — Résultats obtenus dans une zone suburbaine.

• La figure 10 correspond aux segments obtenus sur une image de résolution différente (40 cm par pixel). Le même spécialiste est utilisé mais la valeur du paramètre résolution est modifiée. Les résultats sont présentés sur les figures 11 et 12. On note l'intérêt que présente la phase de connexion sur cette image.

Les rivières

Le spécialiste « rivière » utilise le même type d'algorithme que le spécialiste « route ». Seules les valeurs des paramètres ont été adaptées pour prendre en compte l'irrégularité des berges et pour orienter la recherche vers les zones les



Figure 10. — Les segments détectés sur l'image.



Figure 11. — Les tronçons détectés.

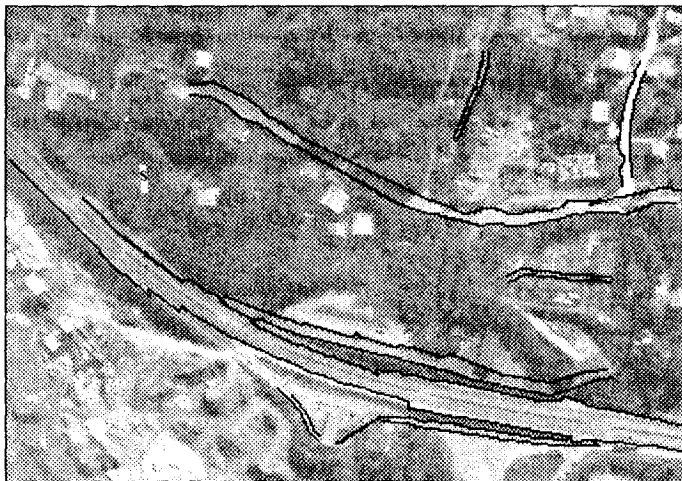


Figure 12. — Amélioration par connexion/suppression.

plus sombres. A partir des segments de la figure 6 le spécialiste a détecté deux hypothèses (dont une qui correspond à une rangée d'arbre) (voir fig. 13). Une anomalie sur la rivière (voir fig. 14) a provoqué une coupure. Néanmoins les deux parties détectées sont suffisamment longues et cohérentes pour que la phase de connexion corrige ce problème (voir fig. 15).

Les bâtiments

La figure 17 présente les résultats obtenus dans le cadre de la recherche de bâtiments (la figure 16 correspondant à l'image initiale).

Le spécialiste utilise une approche différente de celle évoquée précédemment : il repose non plus sur une segmentation-contour mais une segmentation-région qui, dans le cadre des bâtiments évidents est plus adaptée.

En effet, l'algorithme de segmentation utilisée [22] permet de ne considérer que les zones « claires » de l'image.

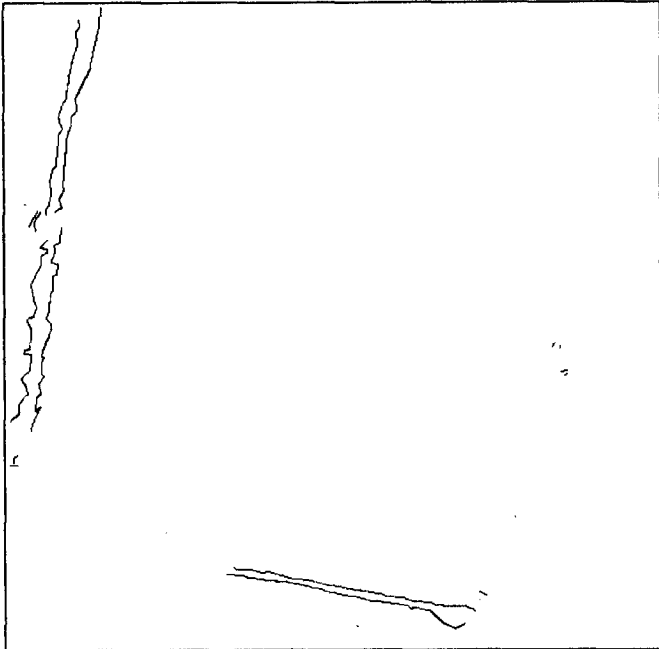


Figure 13. — Les rivières détectées à partir de l'image 6.

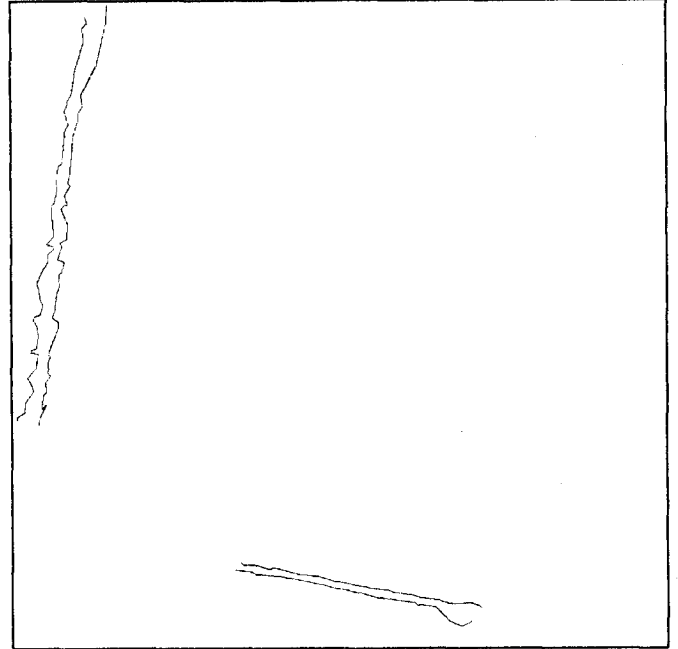


Figure 15. — Amélioration des rivières par connexion.



Figure 14. — Une « anomalie » sur la rivière.



Figure 16. — Image initiale (recherche de bâtiments).

Un processus de groupement, basé sur la recherche d'angles droits (voir *fig. 18*), détecte des formes rectangulaires.

Certaines des hypothèses émises sont fausses, une forme rectangulaire peut effectivement correspondre à un petit champ ou à un parking, une phase de validation sera donc nécessaire.

5.3. UTILISATION DU CONTEXTE

Les règles de démarrage ayant été exploitées, le superviseur oriente maintenant l'activité du système vers l'exploitation

des règles d'inférence ou de validation. Le système focalise donc son activité sur des zones réduites de l'image.

Les routes

Pour les routes, il existe une règle qui permet de « rechercher les véhicules sur les routes ». L'objectif de cette règle est l'inférence de nouveaux objets. Elle utilise une pré-

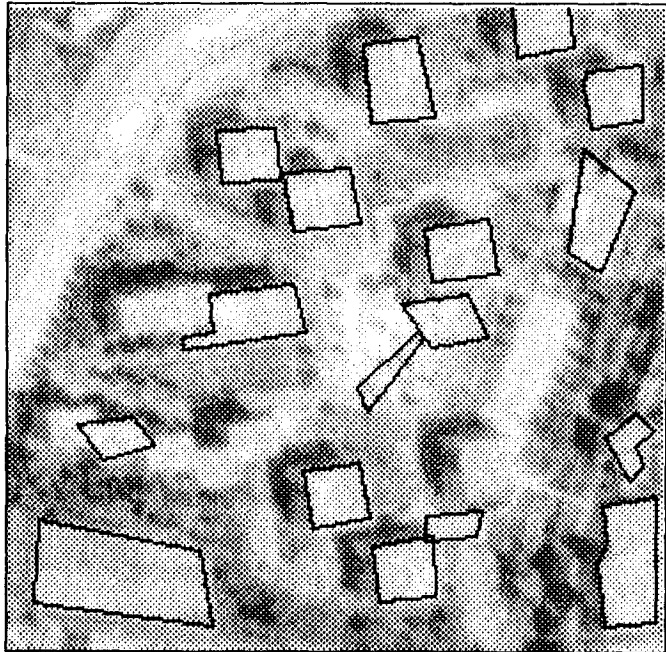


Figure 17. — Hypothèses de bâtiments.

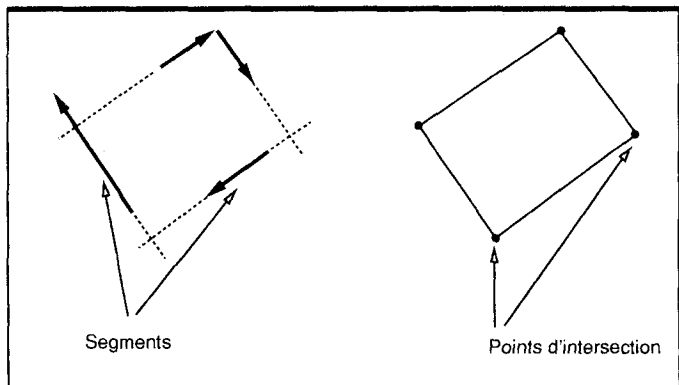


Figure 18. — Génération d'une hypothèse de bâtiment.

condition qui permet de rechercher les véhicules sur les axes routiers les plus importants (les plus larges). La définition de la zone de recherche est simple puisqu'elle est liée à la surface de la route (voir fig. 19 b).

Le superviseur détermine alors le ou les spécialistes à activer. Actuellement, il n'existe qu'un seul spécialiste capable de répondre à la requête de recherche de voitures. Ce spécialiste crée une image à partir de la zone de recherche en préservant les valeurs des pixels de la route (fig. 19 b). Il utilise ensuite une segmentation en régions qui lui permet d'obtenir les résultats de la figure 19 c (respectivement 20 à partir de l'image 8).

Suite aux réponses du spécialiste, le superviseur exécute les actions concernant la règle ; ces actions consistent à ajouter les véhicules à la scène et à établir un lien avec la

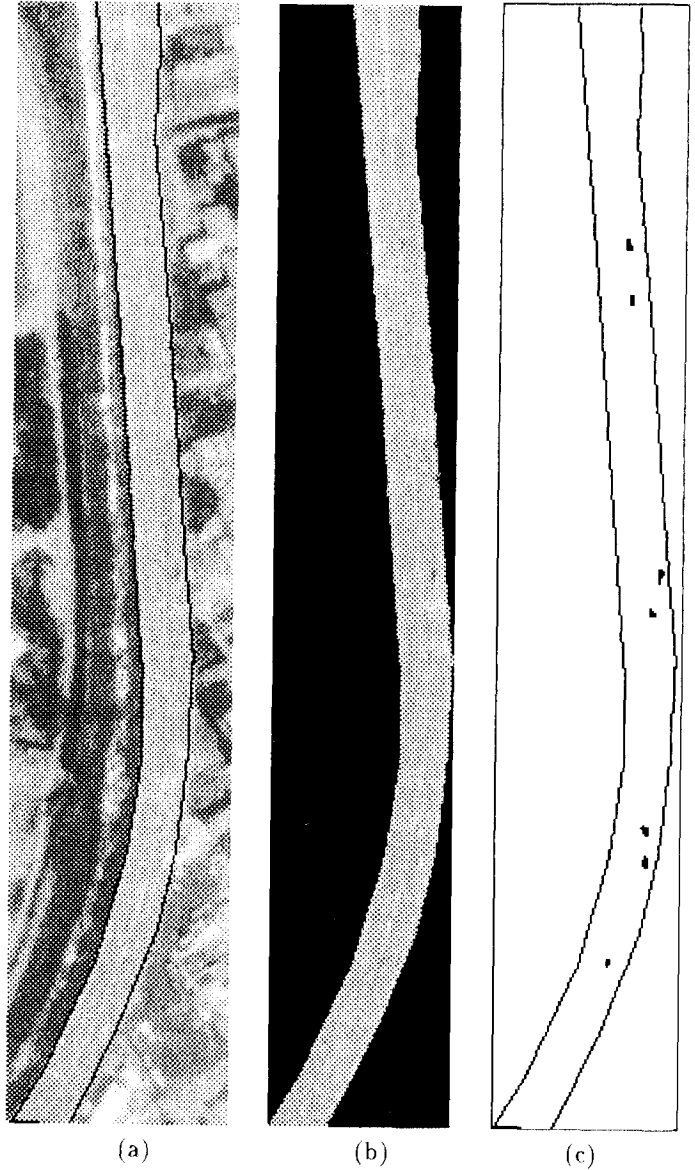


Figure 19. — Recherche de voitures sur une route de l'image 9.

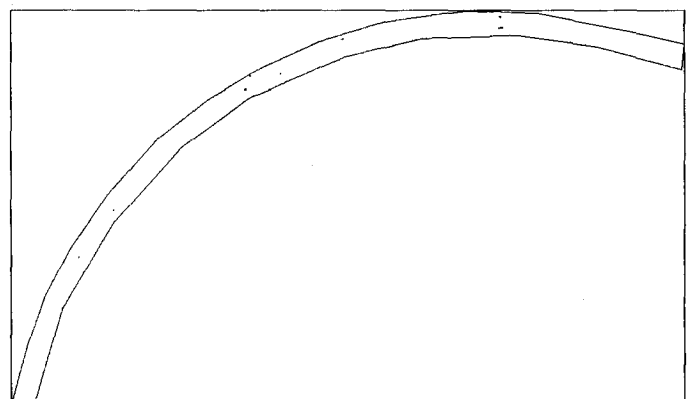


Figure 20. — Recherche de voitures sur une route détectée sur l'image 8.

route. Mais d'autres règles sont envisageables pour l'étude des routes :

- Les plus essentielles concernent le prolongement des routes. Si une hypothèse de route s'arrête ailleurs que sur le bord de l'image, alors de nombreuses informations de contexte peuvent être envisagées : existence d'un tunnel, masquage par des arbres ou plus généralement une segmentation de l'image insuffisante.

Pour illustrer la potentialité de cette approche, nous avons calculé, pour les extrémités des routes qui s'interrompent, de nouveaux segments en utilisant la même méthode que pour les routes évidentes mais avec des valeurs de paramètres permettant d'obtenir plus de détails (voir fig. 21).

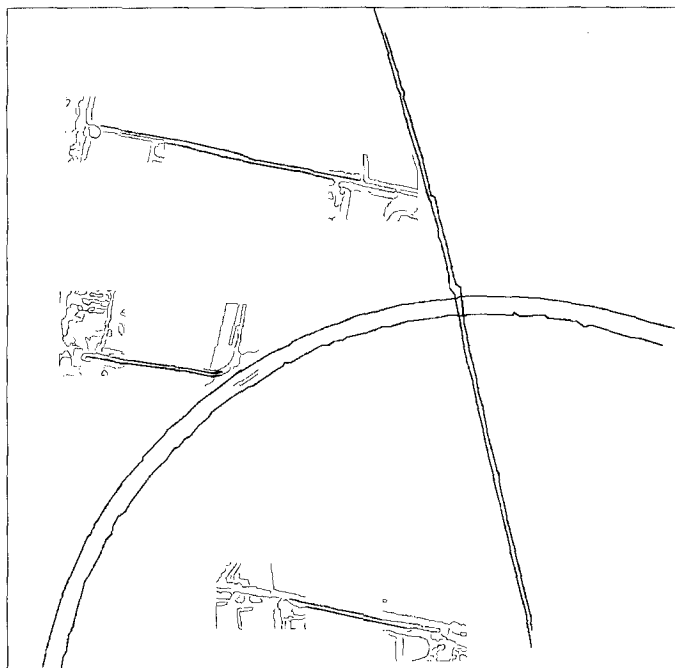


Figure 21. — Segmentation image « guidée » par les hypothèses de routes.

- Dans certaines zones urbaines ou agricoles, les routes présentent des symétries (alignement) importantes avec les bâtiments ou les champs. Cette particularité peut être exploitée facilement.

Les bâtiments

Le spécialiste de détection des bâtiments utilise des informations qui ne permettent pas d'assurer une confiance importante dans les hypothèses émises. Pour résoudre ce problème, nous avons introduit, dans le modèle des bâtiments, une règle dont l'objectif est la validation de ces hypothèses. Cette règle est basée sur le contexte suivant « il existe une ombre près d'un bâtiment ». Une précondition d'utilisation de cette règle dépend, bien entendu, des conditions d'acquisition de l'image (présence ou non de soleil). Cette règle est déclenchée pour chacune des hypothèses émises précédemment. Le contrôleur de localisation calcule alors la zone de recherche en fonction de la

forme du bâtiment et de l'orientation du soleil. La figure 22 a, b, c illustre la détermination de la zone de recherche qui est calculée en tenant compte de la hauteur maximum d'un bâtiment. Ce paramètre étant défini, le superviseur sélectionne alors un spécialiste capable de détecter une ombre (la fig. 22 d montre l'ombre qui a été détectée dans le cas précédent). Une fois que le spécialiste a répondu (il peut détecter plusieurs ombres pour un même bâtiment), le superviseur vérifie que les réponses vérifient les contraintes exprimées par la règle. Ces contraintes permettent de vérifier la cohérence (localisation et forme) entre les ombres et le bâtiment. Pour les objets qui vérifient ces contraintes, la partie action de la règle est déclenchée ; cette action consiste à établir un lien entre les couples d'objets ombre/bâtiment et à mettre à jour le coefficient de vraisemblance. La figure 24 montre les couples d'objets jugés cohérents.

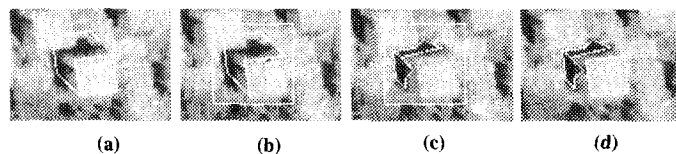


Figure 22. — Détermination de la zone de recherche de l'ombre.

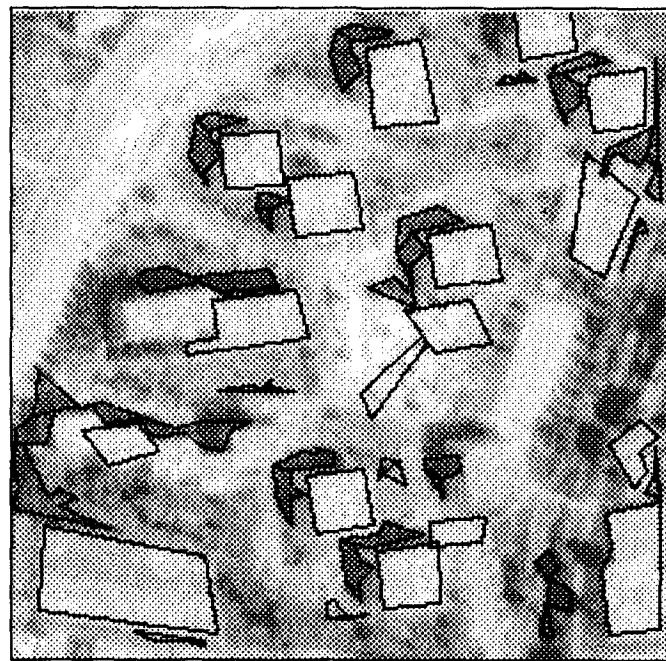


Figure 23. — Validation par recherche des ombres.

D'autres règles peuvent ensuite être appliquées. Par exemple, mettre en œuvre une stratégie permettant d'utiliser les bâtiments précédents pour en détecter de nouveaux. Cette stratégie s'exprime par deux règles :

1. Demander de « chercher des bâtiments proches d'un autre bâtiment » et définir un objet « groupe de bâtiments ».

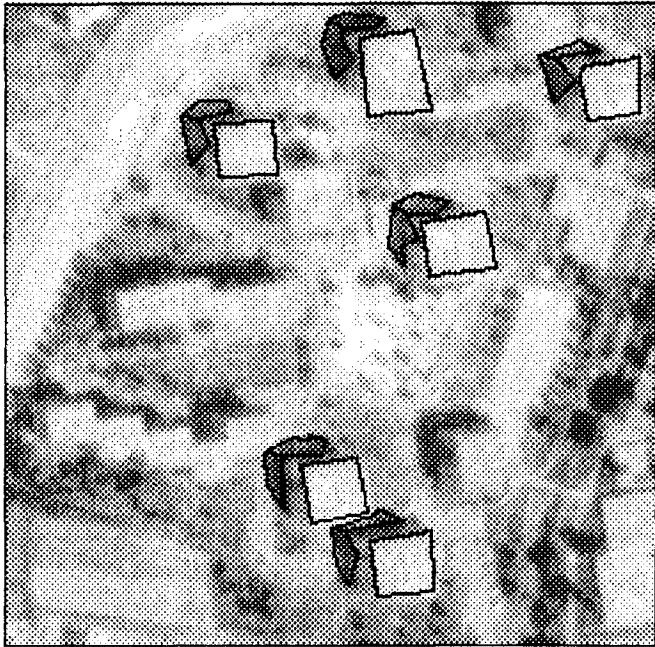


Figure 24. — Bâtiments validés.

Dans ce cas, nous précisons dans les informations de contrôle de la règle que seul le contrôleur de scène doit répondre.

2. Utiliser comme objet déclencheur « un groupe de bâtiments ». Si les bâtiments du groupe présentent des caractéristiques communes (pré-conditions de la règle) alors des bâtiments similaires peuvent être recherchés.

De manière plus générale, l'architecture du système peut également permettre la mise en œuvre de toute autre stratégie. Nous pourrions, par exemple, commencer à chercher d'abord toutes les ombres puis vérifier s'il existe un bâtiment à proximité en définissant uniquement deux nouvelles règles !

6. Conclusion

Nous venons de présenter MESSIE, architecture multi-spécialistes pour l'interprétation d'images, et l'avons appliquée pour l'interprétation en imagerie aérienne.

Cette architecture a été mise en œuvre pour répondre à la diversité de la connaissance à exprimer ; en particulier pour répondre à la modélisation des objets physiques, modélisation qui repose sur une organisation selon quatre points de vue. L'architecture du système MESSIE est basée sur le modèle du tableau noir dont l'intérêt dans le cadre de l'analyse de scène n'est plus à démontrer. Cette architecture facilite le développement et la maintenance du système en proposant de répartir la connaissance dans différentes bases de connaissances.

La connaissance est répartie en deux niveaux hiérarchiques. Le premier niveau décrit la connaissance liée à la stratégie,

à la scène et aux relations entre les objets. Cette connaissance est respectivement exploitée par le superviseur, le contrôleur de scène et le contrôleur de localisation. Le deuxième niveau correspond à la connaissance relative aux points de vue forme et aspect des objets ; elle est mise en œuvre par de bases de connaissances appelées spécialistes.

D'un point de vue stratégique, la répartition des connaissances entre les différents spécialistes et contrôleurs offre des possibilités intéressantes que nous avons pour le moment exploitées en mettant en œuvre un processus basé sur le principe suivant : le mécanisme de compréhension d'une scène complexe peut être vu comme un processus hiérarchique. Nous recherchons les objets remarquables puis, en focalisant la recherche, des objets plus complexes sont inférés ou validés.

Reproduire cette démarche dans le cadre d'un système automatique d'interprétation nous semble indispensable :

- d'une part pour pallier les problèmes liés à la quantité d'informations à traiter (il faut que le système focalise son attention sur une partie réduite de l'image),
- d'autre part pour pallier les difficultés de compréhension : l'identification de certains objets ne peut se faire qu'après la reconnaissance des objets avoisinants. C'est notamment le cas lorsque la qualité des données n'est pas suffisante ou lorsque l'analyse doit prendre en compte des phénomènes tels que les reflets ou le masquage.

La séparation de la connaissance de haut niveau en trois modules (contrôleur de scène, contrôleur de localisation et superviseur) facilite l'étude des différents problèmes à résoudre. Néanmoins, la réalisation d'un système d'interprétation est un sujet très vaste et le prototype qui a été développé n'est qu'un premier pas vers une solution.

L'ampleur du problème nous a conduit à aborder certains problèmes de façon très fragmentaire :

- la vraisemblance des hypothèses d'objets : elle doit permettre de quantifier numériquement l'ensemble des informations symboliques qui interviennent : contexte de création de l'objet, méthode de détection, relations avec les objets avoisinants.

- La planification : actuellement l'analyse de la scène est guidée par des règles relativement élémentaires. Il serait fortement souhaitable d'améliorer le contrôleur de localisation afin qu'il puisse construire automatiquement des plans plus complexes à partir de cette connaissance fragmentaire.

Au-delà de ces difficultés, l'un de nos objectifs consiste à démontrer les potentialités de MESSIE dans d'autres applications. Dès à présent, une nouvelle application est en cours de développement, elle concerne l'interprétation de données issues de capteurs SPOT et SAR [28] [4]. Cette application devrait permettre de démontrer les potentialités de l'architecture proposée dans le cadre de la fusion multi-capteurs pour laquelle l'approche multi-spécialistes semble particulièrement adaptée. Nous profitons de cette nouvelle application, pour réécrire une nouvelle version avec un Noyau Multi Spécialistes (NMS) [9] [6] développé à l'INRIA. En effet, le développement de MESSIE nous a montré l'importance et l'énorme quantité de travail de type génie logiciel qu'une telle entreprise nécessitent, et par là-

même, l'obligation de s'appuyer sur un noyau performant et maintenu. De plus, l'utilisation d'un outil de pilotage d'algorithmes comme OCAP1 [5] est envisagé à très court terme pour améliorer les performances des chaînes de traitements complexes mises en œuvre par les spécialistes.

Manuscrit reçu le 4 juillet 1991.

Remerciements

Nous tenons à remercier à travers Mr P. Nonin, la Société THOMSON-CSF Prime Contractor du projet Esprit SPAN N° 1588, pour nous avoir fourni les données image et pour les fructueux échanges que nous avons eus avec lui et son équipe. Nous tenons à remercier aussi Mademoiselle I. Herlin pour l'aide précieuse apportée à la relecture de ce manuscrit ainsi que les membres du comité de lecture pour leurs critiques et remarques constructives.

BIBLIOGRAPHIE

- [1] T. O. BINFORD, Survey of Model-Based Images Analysis Systems, *Int. J. Robot. Res.*, Vol. 1, N° 1, pp. 18-64, Spring 1982.
- [2] J. CHAILLOUX, Le-Lisp de l'INRIA, *Manuel de référence, Le-Lisp 15.2*, Mai 1986.
- [3] V. CLÉMENT, G. GIRAUDON, S. HOUZELLE, Interpretation of Remotely Sensed Images in a Context of Multisensor fusion, *In 2nd ECCV Conference*, San Margherita, Italy, Mai 1992.
- [4] V. CLÉMENT, G. GIRAUDON, S. HOUZELLE, A Knowledge-Based Interpretation System for Fusion of SAR and SPOT Images, *Proc. of IGARSS*, Houston, Texas, May 1992.
- [5] V. CLÉMENT, M. THONNAT, Pilotage de procédures de traitement d'images pour la description morphologique de galaxies, *Numéro spécial TS/IA*, 1992.
- [6] V. CLÉMENT, G. GIRAUDON, S. HOUZELLE, A Multi-Specialist Architecture for Sensor Fusion in Remote Sensing, *11th ICPR*, The Hague, 1992.
- [7] J. P. COQUEREZ and J. DEVARIS, Détection de contours dans les images aériennes : nouveaux opérateurs, *In Traitement du Signal*, Vol. 2, N° 1, 1985.
- [8] O. CORBY, Un tableau réflexif pour la coopération de bases de connaissances. Thèse de Doctorat, Université de Nice, juin 1988.
- [9] J. L. ANTOINE, O. CORBY, M. PUGELLI, Erasme a multi expert system for pavement rehabilitation, *International Conference on Applications of Advanced Technology in Transportation Engineering*, San Diego, février 1989.
- [10] R. DERICHE, Optimal Edge Detection Using Recursive Filtering, *In International Journal of Computer Vision*, pp. 167-187, 1987.
- [11] J. DESACHY, ICARE : An Expert for Automatic Mapping from Satellite Imagery, *In Proceedings of the NATO Workshop, Mapping and Spatial Modelling for Navigation*, pp. 181-192, Denmark, August 1989.
- [12] M. DESVIGNES, Étude de l'influence du contexte dans un Système d'Interprétation d'une Séquence d'Images. Réalisation à l'aide d'un blackboard, Thèse de l'Université de Caen, Spécialité Sciences, juin 1990.
- [13] B. A. DRAPPER, R. T. COLLINS, J. BROLIO, A. R. HANSON, E. M. RISEMAN, Issues in the development of a blackboard-based schema system for image understanding, *Blackboard systems*, pp. 189-298, Addison Wesley, 1988.
- [14] B. A. DRAPPER, R. T. COLLINS, J. BROLIO, A. R. HANSON, E. M. RISEMAN, The Shema System, *Intern. Journal of Comp. Vision*, 2, pp. 209-250, Kluwer Academic Publishers, janvier 1989.
- [15] P. FUA, Une approche variationnelle pour la reconnaissance d'objets, Thèse de Doctorat, Orsay, September 1989.
- [16] P. FUA, A. J. HANSON, Objective Functions for Feature Discrimination in *IJCAI-89 Conference*, pp. 1596-1602, Detroit USA, August, 1989.
- [17] P. GARNESON, G. GIRAUDON, Spatial Context in an Image Analysis System, pp. 579-582, *ECCV*, Antibes 1990.
- [18] P. GARNESON, MESSIE : un système d'analyse de scène, application à l'imagerie aérienne, Thèse de Doctorat, Université de Nice Sophia-Antipolis, juillet 1991.
- [19] P. GAUSSIER, J. P. COQUEREZ and S. PHILIPP, Systèmes d'Interprétation mixte : réseaux de neurones/Système-Expert appliqué à l'interprétation d'Images aériennes, *8e Congrès RFIA*, pp. 477-488, Lyon novembre 1991.
- [20] J. F. GILMORE, S. P. ROTH, S. D. TYNOR, A Blackboard System for Distributed Problem Solving, *Blackboard Architectures and Applications*, pp. 371-394, Academic Press, 1989.
- [21] G. GIRAUDON, Chaînage efficace de contours, *Rapport de Recherche N° 605 INRIA*, février 1987.
- [22] G. GIRAUDON, P. MONTESINOS, Coopération Contour-Région pour la Détection Infrarouge, *6e Congrès RFIA*, Antibes, novembre 1987.
- [23] A. R. HANSON, E. M. RISEMAN, VISIONS : A Computer System for Interpreting Scenes, New York Academic Press, pp. 303-333, 1978.
- [24] ALLEN R. HANSON, EDWARD M. RISEMAN, The VISIONS image Understanding System, *Advances in Computer Vision*, pp. 1-114, C. M. BROWN ed., Erlbaum Assoc., 1987.
- [25] B. HAYES-ROTH, The Blackboard Architecture : A General Framework for Problem Solving ?, Stanford University, Report HPP-83-30, May 1983.
- [26] F. HAYES-ROTH, A Blackboard Architecture for Control, *Artificial Intelligence 26(2)*, pp. 251-321, 1985.
- [27] B. HAYES-ROTH, M. V. JOHNSON, A. GARVEY, M. HEWETT, Building Systems in the BB* Environment, *Blackboard Systems*, pp. 543-560, Addison-Wesley, 1988.
- [28] S. HOUZELLE, G. GIRAUDON, Data fusion using SPOT and SAR images for bridge and urban area extraction, *IGARSS*, Vol. 3, pp. 1455-1458, Espoo Finland, juin 1991.
- [29] Y. HSIEH, F. PERLANT and D. M. MCKEOWN, Recovering 3D Information from Complex Aerial Imagery, *In Proc. of ICPR*, pp. 136-146, Atlantic City, June 1990.
- [30] A. HUERTAS, W. COLE and R. NEVATIA, Detecting Runways in Aerial Images, *In Proc. of 6th NCAI*, AAAI, pp. 272-297, July 1987.
- [31] A. HUERTAS and R. NEVATIA, Detecting Building in Aerial Images, *In ICGCV*, Vol. 41, N° 2, pp. 131-152, February 1988.
- [32] J. M. HULLOT, CEYX version 15, Programmer en Ceyx, *Rapport Technique INRIA*, février 1985.
- [33] R. B. IRWIN and D. M. MCKEOWN, Methods for exploiting the Relationship between Buildings and their Shadows in Aerial Imagery, *IEEE Trans. on Systems, Man and Cybernetics*, Vol. 19, N° 6, November 1989.
- [34] H. LAASRI, B. MAITRE, Coopération dans un univers multi-agents basée sur le modèle du black-board : études et réalisations. Thèse, Université de Nancy I, 1989.
- [35] D. G. LOWE, Perceptual organisation and Visual Recognition, 1985.

- [36] W. LUO and H. MAITRE, Using Surface Model to Correct and Fit Disparity data in Stereo Vision, *In Proc. of ICPR*, pp. 60-64, Atlantic City, June 1990.
- [37] H. MAITRE, Utilisation de l'imagerie aérienne et satellitaire pour l'aménagement du territoire, *Tutorial 8e Congrès AFCET*, Lyon, novembre 1991.
- [38] T. MATSUYAMA, V. SHANG-SHOUQ HWANG, SIGMA, A knowledge-Based Aerial Image Understanding System, *Advances in Computer Vision and Machine Intelligence*, Editor Martin D. Levine, Plenum Press, New York, 1990.
- [39] D. M. McKEOWN, W. A. HARVEY, J. McDERMOTT, Rule-Based Interpretation of Aerial Imagery, *in PAMI*, Vol. 7, N° 5, pp. 570-585, September 1985.
- [40] D. M. McKEOWN, J. L. DENLINGER, Cooperative Methods for Road Tracking in Aerial Imagery, *in Proc. of IUW*, 1990, pp. 662-672.
- [41] R. MOHAN, R. NEVATIA, Perceptual Grouping for the Detection and Description of Structures in Aerial Images Iris Report, N° 225, University of Southern California, 1989, *IEEE-PAMI*, Vol. 11, November 1989.
- [42] P. MONTESINOS, Une méthodologie en vue de l'interprétation automatique d'images aériennes : application à la détection des bâtiments, Thèse de Doctorat, Université de Nice Sophia-Antipolis, février 1990.
- [43] M. NAGAO, T. MATSUYAMA, A Structural Analysis of Complex Aerial Photographs, New York : Plenum 1980.
- [44] A. M. NAZIF, M. D. LEVINE, Low Level Segmentation : An Expert System, *in PAMI*, Vol. 6, n° 5, pp. 555-577, September 1984.
- [45] B. NICOLIN and R. GLABER, A Knowledge-Based System for the Analysis of Aerial Images, *in IEEE Geoscience and Remote Sensing*, Vol. 25, N° 3, May 1987.
- [46] H. P. NU, Blackboard systems, *AI magazine*, 1986.
- [47] G. REYNOLDS, N. IRWIN, A. HANSON, E. RISEMAN, Hierarchical Knowledge-directed Object Extraction Using a Combined Region and Line Representation, *Proc. of Work. on Comp. Vision Repres. and Cont.*, Silver Spring, pp. 238-247, 1984.
- [48] T. SMITH, Requirements and Principles for the Implementation and Construction of Large-scale Geographic Information Systems, *in Int. J. geographical information systems*, Vol. 1, N° 1, pp ; 13-31, 1987.
- [49] M. ZAHID, J. P. COQUEREZ and S. PHILIPP, Localisation et Reconstruction de Bâtiments à partir d'un couple Stéréoscopique d'Images Aériennes, *8e Congrès RFIA*, pp. 1099-1104, Lyon, novembre 1991.