



HAL
open science

Vision pour la robotique en environnement naturel

Patricia Lasserre

► **To cite this version:**

Patricia Lasserre. Vision pour la robotique en environnement naturel. Automatique / Robotique. Université Paul Sabatier - Toulouse III, 1996. Français. NNT: . tel-00139846

HAL Id: tel-00139846

<https://theses.hal.science/tel-00139846>

Submitted on 3 Apr 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Année 1996

Thèse

préparée au
Laboratoire d'Analyse et d'Architecture des Systèmes du CNRS

en vue de l'obtention du
Doctorat de l'Université Paul Sabatier de Toulouse

Spécialité : Robotique

par **Patricia LASSERRE**

Vision pour la robotique en environnement naturel

Soutenue le Jeudi 26 Septembre 1996 devant le jury:

MM.	Georges	GIRALT	Président
	Maurice	BRIOT	Directeur de thèse
	Dominique	BARBA	Rapporteur
	Michel	CATTOEN	Rapporteur
	Jack-Gérard	POSTAIRE	Rapporteur
	Mallik	GHALLAB	Examineur

Rapport LAAS N° 96423

Thèse préparée au Laboratoire d'Analyse
et d'Architecture des Systèmes du CNRS
7, Avenue du Colonel Roche
31077 Toulouse Cedex 4 (France)

A mes parents, mon frère,
ma famille et mes amis,
Yves...

Avant-propos

Les travaux présentés dans ce mémoire ont été effectués au Laboratoire d'Analyse et d'Architecture des Systèmes, LAAS-CNRS, dirigé par Monsieur Alain Costes, que je remercie pour son accueil.

Mes remerciements vont aussi à Monsieur Georges Giralt, qui était alors directeur du groupe Robotique et Intelligence Artificielle (dirigé à présent par Monsieur Malik Ghallab), de m'avoir permis d'effectuer mes travaux dans son groupe. Je tiens aussi à lui exprimer toute ma gratitude pour l'honneur qu'il m'a fait en présidant le jury de ma thèse.

Je remercie également Maurice Briot, grâce à qui j'ai découvert la Vision par Ordinateur. Il m'a encadrée au cours de ces années et j'ai pu apprécier sa disponibilité, surtout dans la phase finale de rédaction du manuscrit.

Parce que juger le travail des autres est une tâche difficile, je voudrais remercier l'ensemble des membres du jury et plus particulièrement Messieurs Dominique Barba, Michel Cattoen et Jack-Gérard Postaire, rapporteurs de ma thèse, pour l'intérêt qu'ils ont porté au manuscrit et pour leurs critiques constructives. Merci aussi à Monsieur Malik Ghallab de m'avoir réservé la journée du 26 septembre.

Une thèse est rarement le travail d'une personne mais reflète plutôt le résultat de discussions et coopérations avec les gens qui nous entourent. Il m'est difficile d'énumérer ici toutes les personnes qui, par leur compétence ou par leur présence (au cours des moments difficiles, jusqu'au petit coup de pouce final), m'ont aidée durant ces quatre années. Merci à tous pour votre précieux soutien !

L'ambiance de travail dans le groupe ne serait pas ce qu'elle est sans les «à-côtés» perpétués au fil des ans par les différentes générations de thésards. Longue vie donc aux «pauses cafés» et «salle Flavien», aux soirées brésiliennes, mexicaines, palestiniennes et j'en passe... Et que dure encore cette ouverture d'esprit procurée par cet environnement si particulier où se cotoient des gens d'horizons différents.

Un petit clin d'œil plus personnel à ceux avec qui j'ai partagé des moments intenses : aux ex-DEA et thésards de ma génération, aux anciens et néo-«zombies» du labo, aux «voileux» sur «Jacby», à ceux du CdT, aux «randonneurs du dimanche», aux «amis de Pierrette»,...

Enfin je réserve mes derniers mots pour ceux à qui je dédie ma thèse et qui, par leur confiance, leur patience et tant d'autres choses encore..., ont été mon réconfort quotidien et ma source d'énergie.

Patricia

Table des matières

Table des matières	i
Liste des figures	v
Liste des tableaux	ix
Présentation générale	1
I Le contexte	5
I.1 La vision et la robotique mobile en environnement d'extérieur	6
I.2 Les travaux menés au LAAS en environnement naturel	8
I.3 EDEN et ADAM	9
I.3.1 Cadre général	9
I.3.2 Le projet EDEN	11
I.4 Conclusion	12
II Vision : l'approche proposée	13
II.1 Position du problème	13
II.1.1 Modélisation de l'environnement dans le cadre d'EDEN	13
II.1.2 Les limites actuelles	15
II.2 Pourquoi la vision?	15
II.2.1 Une meilleure connaissance de l'environnement «proche»	16
II.2.2 Connaissance de l'environnement lointain	18
II.2.3 En conclusion...	18
II.3 Approche proposée	19
II.3.1 Présentation générale	19
II.3.2 Un nouveau modèle	20

II.3.3	Quelles classes dans la base de données?	21
II.3.4	Quels attributs pour la caractérisation?	22
II.3.5	Construction du Modèle Nominatif de Régions	22
II.4	Conclusion	25
III	La stéréovision	27
III.1	Présentation générale	27
III.1.1	Principe de la stéréovision	27
III.1.2	Les principales approches	29
III.2	Méthode utilisée	30
III.2.1	Rectification	31
III.2.2	Corrélation	37
III.2.3	Reconstruction 3D	40
III.3	Résultats expérimentaux	41
III.4	Conclusion et perspectives	45
IV	L'extracteur de régions	47
IV.1	Présentation générale	48
IV.2	La segmentation d'images couleur	50
IV.2.1	Les méthodes existantes	51
IV.2.2	Notre approche	53
IV.2.3	Génération des classes de régions élémentaires	57
IV.2.4	Conclusion	68
IV.3	La caractérisation des régions	69
IV.3.1	Les attributs colorimétriques	69
IV.3.2	Les attributs de texture	75
IV.3.3	Conclusion	79
IV.4	Conclusion	81
V	Identification des régions	83
V.1	Objectifs et principe	83
V.2	Génération de la base de connaissances	85
V.2.1	L'analyse factorielle discriminante	86
V.2.2	Le prototypage	87

V.3	Classification des régions	89
V.3.1	La classification bayésienne	89
V.3.2	Validation expérimentale	90
V.3.3	Conclusion	99
V.4	Evaluation de l'identification et de la segmentation	100
V.5	Fusion des régions et construction du modèle	104
V.5.1	Fusion des régions connexes	104
V.5.2	Vérification de la cohérence et construction du modèle	105
V.6	Conclusion	109
Conclusion générale		111
A	Segmentation d'images intensité	115
B	Etude sur les attributs de texture	117
Références bibliographiques		121

Table des figures

I.1	Le robot mobile ADAM	9
I.2	Une scène de EDEN	10
II.1	Un environnement mal connu	16
II.2	Les différentes représentations de l'environnement	19
II.3	Construction du Modèle Nominatif de Régions	23
II.4	Les différentes résolutions des images	24
III.1	Principe de la stéréovision	28
III.2	Le modèle caméra CAHV	32
III.3	Construction du repère rectifié	33
III.4	Déformation des images rectifiées	33
III.5	Choix des contraintes pour la rectification	35
III.6	Principe de la corrélation	38
III.7	Idée de base du critère de vérification	38
III.8	Corrélation et vérification en une seule passe	39
III.9	Un exemple complet sur une paire d'images stéréoscopiques	43
III.10	Résultat de la stéréocorrélation sur deux exemples différents	44
IV.1	Schéma fonctionnel de l'extracteur de régions	49
IV.2	Une scène de EDEN	54
IV.3	Image IV.2 segmentée	54
IV.4	Partie de l'image IV.2	54
IV.5	Image IV.4 segmentée	54
IV.6	Génération des cellules	56
IV.7	Histogramme $H(c)$ et $H_{cum}(c)$	58
IV.8	Image 1 couleur	60
IV.9	Image 2 couleur	60
IV.10	Image 3 couleur	60
IV.11	Image 4 couleur	60
IV.12	Image 5 couleur	60
IV.13	Image 6 couleur	60

IV.14	Image 1 segmentée	61
IV.15	Image 2 segmentée	61
IV.16	Image 3 segmentée	61
IV.17	Image 4 segmentée	61
IV.18	Image 5 segmentée	61
IV.19	Image 6 segmentée	61
IV.20	Génération des seuils de séparation des classes	64
IV.21	Histogrammes sur les trois plans couleur de l'image IV.24	65
IV.22	Image 1 couleur	66
IV.23	Image 2 couleur	66
IV.24	Image 5 couleur	66
IV.25	Image 7 couleur	66
IV.26	Image 8 couleur	66
IV.27	Image 9 couleur	66
IV.28	Image 1 segmentée	67
IV.29	Image 2 segmentée	67
IV.30	Image 5 segmentée	67
IV.31	Image 7 segmentée	67
IV.32	Image 8 segmentée	67
IV.33	Image 9 segmentée	67
IV.34	Représentation des classes dans RVB	71
IV.35	Représentation des classes dans rvb	72
IV.36	Représentation des classes dans ITS	73
IV.37	Représentation des classes dans rv	74
IV.38	Représentation de trois classes dans le repère teinte	75
IV.39	Représentation après analyse factorielle discriminante sur l'ensemble des attributs de texture	79
IV.40	Représentation des classes après analyse factorielle discriminante sur RVB et texture (2 axes discriminants conservés)	80
IV.41	Représentation des classes après analyse factorielle discriminante sur RVB et texture (3 axes discriminants conservés)	80
V.1	Déroulement de la phase d'identification	84
V.2	Représentation des classes après AFD sur l'ensemble des attributs (2 axes discriminants conservés)	88
V.3	Représentation des classes après AFD sur l'ensemble des attributs (3 axes discriminants conservés)	88
V.4	Image d'intensité de l'image IV.11	94
V.5	Résultat de l'identification de l'image V.4 (segmentation méthode 1)	94

V.6	Résultat de l'identification de l'image V.4 (segmentation méthode 2)	96
V.7	Evaluation de la segmentation (méthode 1) de l'image V.4	98
V.8	Identification des régions re-segmentées	98
V.9	Evaluation de la segmentation (méthode 2) de l'image V.4 et identification des régions re-segmentées	98
V.10	Evaluation de la segmentation (méthode 2) et identification sur une image du même type	99
V.11	Evaluation de la segmentation et de l'identification de l'image V.4	102
V.12	Schéma de principe de fonctionnement du procédé d'évaluation	103
V.13	Fusion des régions de l'image V.5	105
V.14	Fusion des régions de l'image V.6	105
V.15	Principales erreurs détectables par le contexte et la position dans l'image .	106
V.16	Principe du relevé de la position d'une région	106
V.17	Fusion des régions de l'image V.13	107
V.18	Fusion des régions de l'image V.14	107
V.19	Image intensité d'une scène	108
V.20	Résultat pour l'image V.19	108
V.21	Image intensité de l'image IV.27	109
V.22	Résultat pour l'image V.21	109
A.1	Image 2 (méthode 1)	116
A.2	Image 2 (méthode 2)	116
A.3	Image 5 (méthode 1)	116
A.4	Image 5 (méthode 2)	116
A.5	Image 7 (méthode 1)	116
A.6	Image 7 (méthode 2)	116
B.1	Représentation des classes selon densité et granularité	118
B.2	Représentation des classes selon densité et discontinuité	118
B.3	Représentation des classes selon densité et courbure	118
B.4	Représentation des classes selon densité et rectitude	118
B.5	Représentation des classes selon densité et netteté	119
B.6	Représentation des classes selon granularité et discontinuité	119
B.7	Représentation des classes selon granularité et courbure	119
B.8	Représentation des classes selon granularité et rectitude	119
B.9	Représentation des classes selon granularité et netteté	119
B.10	Représentation des classes selon discontinuité et courbure	119
B.11	Représentation des classes selon discontinuité et rectitude	120
B.12	Représentation des classes selon discontinuité et netteté	120
B.13	Représentation des classes selon courbure et rectitude	120

B.14	Représentation des classes selon courbure et netteté	120
B.15	Représentation des classes selon courbure et rectitude	120

Liste des tableaux

V.1 Résultats de l'identification et de la segmentation (méthode 1)	93
V.2 Résultats de l'identification et de la segmentation (méthode 2)	96
V.3 Limites sur l'entropie et la probabilité pour chaque classe	101
B.1 Dispersion des classes selon les attributs de texture	117



Présentation générale

Robotique...

Dans le milieu industriel, robotique est synonyme de machine automatisée et d'optimisation du rendement de l'entreprise.

Au niveau médiatique, la robotique mobile revêt une toute autre signification. L'imagination prend le relais sur la réalité : les robots deviennent des humanoïdes, capables de parler, d'être des «humoristes», et de se déplacer avec aisance...

Mais la robotique mobile, c'est uniquement pour les chercheurs, une machine que l'on dote d'une «intelligence» relative. Ce terme d'«intelligence» reflète alors une connotation très différente. Il s'agit d'autonomie de déplacement, de capacité d'appréhender l'environnement et d'évoluer dans celui-ci sans intervention systématique de l'homme. On est loin ici des deux définitions précédentes : les humanoïdes restent du domaine du rêve, les machines automatisées existent depuis longtemps déjà.

L'objectif des scientifiques est tout autre : leur but est d'utiliser un robot mobile autonome pour aider l'homme dans des cas de figures où il lui est difficile d'agir comme l'exploration planétaire, l'intervention sur des sites dangereux... Dans ce cadre, l'autonomie du robot réside dans sa capacité de réaction et d'action dans un environnement qui peut être inconnu *a priori*.

Cet objectif requiert des capacités de raisonnement pour effectuer la mission demandée. En particulier, il sera nécessaire de modéliser l'environnement à partir de données issues des capteurs du système de perception du robot. Cette étape est primordiale puisqu'elle sera à la base des décisions futures qui devront être prises pour accomplir la tâche fixée. Elle consiste en effet à fournir une description synthétique de l'environnement sur laquelle reposera le raisonnement pour établir le déplacement du robot, sa localisation dans l'environnement... Plusieurs représentations coopèrent généralement pour apporter l'information utile à la tâche en cours d'exécution (navigation ou localisation), chacune d'elles ayant des besoins spécifiques.

C'est dans le cadre de la modélisation que notre travail se situe. La plupart des représentations utilisées en environnement naturel se basent essentiellement sur des modèles géométriques et topologiques pour décrire l'environnement. La télémétrie reste généralement privilégiée alors que d'autres capteurs tels que les caméras vidéo pourraient apporter

une sémantique riche pour ce type d'environnements.

La vision par ordinateur couvre un large spectre d'applications allant de l'industrie, en passant par l'imagerie médicale ou la surveillance, jusqu'à la robotique. Elle est néanmoins peu utilisée en robotique mobile d'extérieur. Le coût calculatoire des traitements en imagerie limite en effet son utilisation. Les progrès techniques en électronique et en informatique permettent cependant d'espérer une augmentation significative de l'exploitation d'un tel capteur, surtout dans des environnements naturels où la végétation peut être luxuriante et les couleurs variées.

Ce mémoire montrera l'intérêt d'un tel capteur pour la modélisation de l'environnement aussi bien dans les tâches de navigation que de localisation. Il s'organise en cinq chapitres selon la description précisée ci-dessous.

Le premier chapitre constitue un état de l'art en robotique mobile d'extérieur. Après une présentation rapide des différents intérêts et thèmes relatifs à la robotique mobile, nous précisons les capteurs utilisés pour modéliser l'environnement, dans différentes réalisations de robots dans le monde. Nous faisons ensuite un bref rappel des travaux en environnement naturel incluant le LAAS ainsi qu'une présentation du projet interne au groupe Robotique et Intelligence Artificielle (RIA), EDEN, dans lequel s'insère notre travail.

Le deuxième chapitre de ce mémoire s'intéresse plus particulièrement à la position générale du problème et à l'approche que nous proposons. Un inventaire des représentations exploitées pour modéliser l'environnement est donné ainsi que les limites actuelles du système de perception. Nous précisons donc comment la vision peut être exploitée pour apporter une amélioration significative pour la modélisation de l'environnement et développons l'approche générale que nous avons adoptée.

Les trois autres chapitres sont relatifs aux travaux menés en vision par ordinateur.

Le chapitre III est consacré à un domaine connu et souvent utilisé en robotique mobile, la stéréovision. Nous nous intéressons essentiellement à la technique de stéréocorrélation qui est adaptée au type de scènes que nous traitons. Nous y présentons comment nous avons amélioré la rapidité d'obtention de l'information tridimensionnelle sans dégrader la qualité du résultat, c'est-à-dire en conservant le maximum de critères pour valider l'appariement.

Les deux suivants sont liés à la modélisation de l'environnement et au moyen d'obtenir une information primordiale pour la navigation et la localisation du robot : l'identification de la nature des éléments contenus dans la scène.

Le chapitre IV précise comment extraire les régions de l'image et comment les caractériser à partir d'attributs de couleur et de texture. Nous y présentons en particulier une technique originale de segmentation basée sur l'étude des histogrammes sur chacune des composantes. Deux méthodes sont présentées, l'une fournissant des régions colorimétrique-

ment homogènes et qui est intéressante pour la construction de la base de connaissances utilisée lors de la phase d'identification ; l'autre qui permet d'extraire les grandes régions de l'image et qui est plus intéressante lors de la reconnaissance. L'identification des régions passe d'abord par la caractérisation de ces dernières. La deuxième partie de ce chapitre est donc consacrée à la présentation des attributs de couleur et de texture utilisés et à leur intérêt pour la caractérisation.

Le chapitre V traite de l'identification proprement dite de la nature des différentes régions. A partir des attributs développés pour caractériser les régions, nous utilisons la classification bayésienne pour déterminer la nature de ces dernières. Une analyse détaillée des résultats sert de base pour déterminer la qualité des résultats de la segmentation et de l'identification. Cette analyse permet aussi de montrer l'intérêt d'associer des informations contextuelles pour valider globalement les résultats obtenus. Dans ce but, des relations sur la position des régions dans l'image ou les unes par rapport aux autres sont introduites. Elles sont utilisées pour définir les règles de base auxquelles doivent répondre les régions et leur nature pour être validées. Nous verrons qu'ainsi, nous pouvons non seulement détecter certaines erreurs d'identification, mais aussi redéfinir la nature de la région en cause.

Nous concluons finalement ce mémoire de thèse en relatant les différentes contributions apportées et leur intérêt dans la construction du modèle de l'environnement. Nous proposons plusieurs perspectives intéressantes comme la construction d'un système plus complexe intégrant l'extraction d'amers (objets particuliers) lointains et le suivi d'objets en dynamique... afin d'améliorer la connaissance globale de l'environnement et d'envisager des traitements rapides pour la vision.

Chapitre I

Le contexte

La principale ligne actuelle de recherche pour la robotique mobile en environnement naturel est de réussir à faire déplacer le robot de manière autonome dans son environnement pour lui faire accomplir une tâche. Loin d'être simple, cet objectif demande au robot des capacités décisionnelles et perceptuelles importantes, l'environnement du robot étant souvent inconnu ou partiellement connu.

Le déplacement du robot est en effet nécessaire quel que soit le problème posé. Deux grandes problématiques peuvent être distinguées pour les robots mobiles autonomes en environnement naturel :

- Le suivi de route : il s'agit ici d'aide à la conduite automobile ; Le conducteur serait prévenu de dangers potentiels grâce à la surveillance de sa conduite.
Ce type de demande nécessite bien entendu une capacité de réaction très grande, des capacités de perception et d'interprétation de l'environnement très rapides en particulier pour détecter si le véhicule reste bien dans l'axe, s'il y a des obstacles sur la route (voitures...)
Bien que ces problèmes soient loin d'être simples, ils ne nécessitent pas obligatoirement la mise en œuvre de structures algorithmiques lourdes, et au contraire, favorisent les méthodes réactives.
- La robotique d'intervention : le robot, dans ce cas, doit remplacer l'homme que ce soit à cause de conditions d'accès ou de travail difficiles ou dangereuses, ou même impossibles parfois. On y retrouve donc les thèmes suivants :
 - La sécurité civile : lors de catastrophes naturelles ou provoquées par l'homme, l'intervention de celui-ci sur les sites peut être dangereuse. L'utilisation d'un robot doté d'autonomie est dans ce cas souhaitée.

- L'exploitation de ressources naturelles : dans les mines, en milieu sous-marin ou sur terre (forêt, ...), l'exploitation peut être difficile. L'homme n'a pas toujours accès facilement à ces sites ou est soumis à des dangers constants.
- L'exploration planétaire : la Lune, Mars ou l'Antarctique sont des zones d'accès difficile ou très coûteux. La découverte de ces sites nécessite l'intervention de robots mobiles autonomes qui effectueraient les expérimentations.

Contrairement au suivi de route, ces applications nécessitent une autonomie décisionnelle et perceptuelle importante dans la mesure où l'homme sera sur des sites distants du robot. D'autre part, les environnements sont complexes, peu connus (ou inconnus), peu structurés (ou pas du tout) et le robot doit pouvoir adapter son comportement à cet environnement.

Le type d'environnement dans lequel devra se déplacer le robot détermine bien entendu le type de capteurs à utiliser pour la perception. L'utilisation de caméras (couleur ou noir et blanc) est souvent limitée même si elle n'est pas remise en cause. La difficulté d'exploitation des données qu'elle fournit, même si ces données sont très riches, fait qu'à ce jour, l'utilisation de la vision reste souvent réduite lorsqu'il s'agit d'applications réelles.

1.1 La vision et la robotique mobile en environnement d'extérieur

La mise en œuvre de robots mobiles autonomes en environnement naturel a commencé à se développer dans les années 80 avec l'apparition de projets concernant les deux problématiques du suivi de route et des robots d'intervention.

Pour des robots d'intervention, la vision n'est pas obligatoirement primordiale pour le déplacement. Suivant le type d'environnement dans lequel il évolue, elle prendra plus ou moins d'importance : il est clair que l'intérêt *a priori* de la couleur, par exemple, sur des sites lunaires ou martiens, est réduit. En effet, dans le cas particulier de missions d'exploration planétaire, l'autonomie du robot réside essentiellement dans sa capacité de déterminer les zones navigables de celles qui ne le sont pas et l'information qui semble la plus naturelle est d'utiliser l'information tridimensionnelle. Celle-ci peut être fournie directement par télémétrie laser ou à partir d'un système de vision stéréoscopique. Différents laboratoires se sont intéressés à cette approche.

Les travaux utilisant la télémétrie laser

À l'université de Carnegie-Mellon, plusieurs projets ont intégré la télémétrie laser pour percevoir l'environnement : les projets AMBLER puis APEX ont permis le développement de robots à pattes. Ces robots ont été prévus pour se déplacer sur des surfaces très irrégulières telles que celles de Mars [Simmons 92, Krotkov 94]. Parallèlement à ce projet, un véhicule tout-terrain équipé d'un télémètre laser a été utilisé pour des expérimentations en environnement naturel [Langer 94]. NAVLAB II utilise uniquement un télémètre laser pour mesurer la forme du terrain autour de lui [Stentz 95]. D'autres capteurs lui ont été associés pour améliorer sa capacité de navigation (centrale

inertielle, odométrie). Ce projet a été supporté en partie par le projet UGV (Unmanned Ground Vehicle).

Le projet ALV (Autonomous Land Vehicle) a donné lieu à beaucoup de travaux dans le domaine de la robotique mobile. En particulier, on peut citer les travaux de [Daily 88]; leur système de navigation (avec utilisation des données issues d'un télémètre laser pour générer une carte d'élévation du terrain) a été testé sur le site de Denver.

Les travaux exploitant la vision achrome

Toujours dans le cadre d'ALV, d'autres travaux ont intégré à la fois les données issues du télémètre et d'une caméra (image intensité) pour classifier les régions étiquetées obstacles en objets artificiels ou naturels [Asada 88, Asada 90].

Au Jet Propulsion Laboratory (JPL), l'appui de la NASA a permis le développement de divers projets [Wilcox 92]. Le robot ROBBY, un véhicule d'exploration planétaire, a été équipé d'un système stéréoscopique. Ce système a été testé en environnement extérieur [Matthies 92, Grandjean 93].

Par la suite, une famille de micro-robots (Rocky family) a été développée, chacun réalisant une tâche spécifique [Stone 93]. La stéréovision est aussi utilisée sur ces robots.

Les micro-robots et l'approche comportementaliste sont à l'initiative du MIT. Ce laboratoire s'intéresse actuellement à l'utilisation de ces robots pour des explorations planétaires [Malafeew 94]. La vision y joue un rôle pour la navigation du robot [Gavin 94].

D'autres travaux, menés en Europe, utilisent aussi l'approche vision achrome pour la navigation du robot. MARSOKHOD, en particulier, est un châssis de conception originale (châssis articulé à six roues coniques) [Kemurdjian 92, Kemurdjian 95], prévu pour se déplacer à la surface de Mars. Une version de ce robot est utilisée au Centre National d'Etudes Spatiales (CNES) [Proy 93].

Citons aussi le projet PANORAMA (ESPRIT II) dont l'objectif était le développement de machines robotisées opérant en milieu partiellement non structuré (environnements d'extérieur) [Halme 93]. Dans ce projet la vision s'est restreinte essentiellement à des tâches de suivi de route ou de reconnaissance d'amers particuliers pour la navigation.

Dans le cadre de PROMETHEUS, la vision achrome est aussi utilisée par U. Solder et V. Greafe pour détecter des obstacles situés sur la route [Solder 93] à des distances variables.

Les travaux utilisant la vision trichrome

Par ailleurs, à l'université de Carnegie-Mellon, le projet ALV a permis l'équipement d'un véhicule commercial pour effectuer le suivi de route. NAVLAB est équipé d'une caméra couleur et d'un télémètre laser [Thorpe 91a, Thorpe 91b]. La vision est utilisée pour détecter la route et la suivre tandis qu'un système de vision 3D gère les problèmes

d'évitement d'obstacles [Thorpe 88]. NAVLAB a d'abord été testé sur route avec succès. Il a par la suite été utilisé en environnement naturel en utilisant uniquement le télémètre laser pour modéliser l'environnement [Langer 95, Kweon 92].

Au niveau européen, cette problématique intervient dans le projet PROMETHEUS : à l'École des Mines de Paris, plusieurs travaux ont été menés dans ce domaine aussi bien concernant la modélisation de la route [Lapierre 94] que concernant la détection de cette dernière et des panneaux de signalisation [Baghdassarian 95].

En conclusion ...

On peut remarquer que la vision s'appuie essentiellement sur des images achromes pour les applications d'exploration planétaire. En revanche, pour le suivi de route, la couleur est exploitée car elle permet de restreindre la recherche (la route est peu colorée par rapport au reste de l'image ; les panneaux ont des couleurs plus saturées que les autres éléments de la scène).

1.2 Les travaux menés au LAAS en environnement naturel

Les premiers travaux menés dans le groupe en robotique mobile en environnement naturel ont débuté en 1987 avec le projet Euréka AMR (Advanced Mobile Robot) [Lacombe 91]. Ce projet a permis la construction d'un démonstrateur ADAM (Advanced Demonstrator for Autonomy and Mobility) par FRAMATOME et MATRA MARCONI Space. Dans ce projet, le groupe RIA a participé à la conception et l'intégration du système de perception notamment par la modélisation du terrain à partir de données télémétriques. ADAM fut en prêt au LAAS à partir d'octobre 1992. Avec son arrivée, un projet interne au groupe, EDFN, a été mis en place. Ce projet sera détaillé dans la section suivante.

Le deuxième projet ayant impliqué le LAAS est le projet VAP (Véhicule Automatique Planétaire), lancé en 1989 par le CNES, et dont le but était de développer les concepts de base pour un robot d'exploration planétaire (Mars mais aussi la Lune) et d'en prouver la faisabilité [Giralt 92]. Cette première phase de ce projet a été réalisée en association avec le consortium RISP (Robots d'Intervention sur Site Planétaire) comprenant les quatre principales organisations françaises de Recherche et Développement travaillant en robotique : le Commissariat à l'Énergie Atomique, le Centre National de la Recherche Scientifique, l'Institut National de Recherche en Informatique et Automatique et l'Office National d'Études et de Recherches Aérospatiales.

Fin 1993, le projet VAP s'est transformé en projet IARES (Illustrateur Autonome de Robotique mobile pour l'Exploration Spatiale). Son objectif est de valider les concepts fonctionnels proposés dans VAP [Boissier 91].

L'implication du LAAS dans ces deux projets a été très forte et a concerné différents aspects (modélisation de l'environnement, navigation, stratégies de déplacements, structure

décisionnelle. . .).

Depuis peu, l'arrivée d'un nouveau châssis LAMA (Lavochkin Alcatel Model Autonomous), prêté par ALCATEL Espace, va permettre de poursuivre la problématique abordée dans EDEN avec le démonstrateur ADAM.

I.3 EDEN et ADAM

L'arrivée de ADAM au LAAS a permis la mise en place d'un projet interne au groupe RIA, EDEN (Expérimentation de Déplacement en Environnement Naturel) dont l'objectif scientifique était le développement, l'intégration et la mise en œuvre de capacités de navigation en environnement naturel.

C'est dans le cadre de ce projet que notre travail s'insère.

I.3.1 Cadre général

ADAM et son système de perception

ADAM, présenté sur la figure I.1, est un robot imposant par sa taille (environ 3x2x2 mètres) et sa masse (800 kg). Le châssis, de fabrication russe, est équipé de six roues non orientables commandables séparément. La gestion des déplacements du robot est incluse dans un module spécifique réalisé dans le cadre d'AMR.



FIG. I.1 – *Le robot mobile ADAM*

A son arrivée au LAAS, le robot a été équipé d'un système de perception comprenant :

- un télémètre laser à deux axes qui fournit une image de points 3D (utilisé donc comme une caméra laser 3D),

- deux bancs stéréoscopiques, l'un orienté vers le bas et dédié à la détection d'obstacles en mouvement, l'autre visant la scène et dédié à la modélisation.

Les caméras couleur que nous avons utilisées sont uniquement celles visant la scène. Ce sont des caméras CCD MICAM 650. Le signal vidéo est numérisé par une chaîne d'acquisition Datacube qui fournit des images de dimensions 576x502. L'objectif utilisé a une focale de 8mm. La balance des blancs est réglée automatiquement. Nous n'intervenons pas dans le processus d'acquisition. Lors du départ de ADAM, un autre système d'acquisition, basé sur l'utilisation d'un camescope SONY CCD-V6000E de focale 8mm a été utilisé. Les images obtenues après numérisation sont de taille 604x804.

L'environnement de EDEN

Le terrain d'expérimentation sur lequel se déplace ADAM est un terrain de petite dimension (20x35 mètres). Un certain nombre d'obstacles composés de rochers isolés ou de tas de cailloux y ont été déposés. Sur les bords de ce terrain se trouvent des arbres qui peuvent être eux aussi isolés ou sous forme de haies. Le terrain est composé essentiellement de zones d'herbe et d'un chemin de terre. Certaines parties de ce terrain sont en pente et constituent des zones accidentées pour le robot. La figure I.2 montre une scène du terrain sur lequel il évolue.



FIG. I.2 – Une scène de EDEN

I.3.2 Le projet EDEN

La mission à effectuer est de réaliser de façon autonome la tâche canonique «aller à - (but)». Ce but peut être spécifié par une position à atteindre ou par une zone particulière identifiable par le robot ; dans le cas d'EDEN, il s'agit d'une cible hexagonale dont le modèle est fourni au robot.

La réalisation de cette tâche se fait par des déplacements successifs du robot jusqu'à ce qu'il atteigne le but fixé. Les différentes étapes se déroulent selon le schéma global suivant :

1. perception de l'environnement,
2. recherche de la cible : si celle-ci n'est pas en vue, le robot connaît la direction dans laquelle elle se trouve. Dans le cas contraire, la position du but sera modifiée si nécessaire ;
3. construction du modèle de l'environnement à partir des données tridimensionnelles et mise à jour du modèle global ;
4. planification du chemin à prendre (sous-buts) sur la partie connue de l'environnement : cette étape pourra être suivie d'une étape de planification de la trajectoire si le terrain à traverser est accidenté ou comporte des obstacles ;
5. exécution du mouvement.

Ces différentes étapes seront effectuées jusqu'à ce que le robot arrive devant la cible.

Pour réaliser une telle tâche, le robot doit être à même de modéliser l'environnement à partir des données issues de ses capteurs, de localiser sa position dans l'environnement et de déterminer le déplacement à effectuer.

Cet objectif a permis l'intégration et le développement des principaux thèmes abordés dans le groupe :

- la modélisation de l'environnement à partir de données télémétriques : avec la construction incrémentale de modèles du terrain à partir d'une carte d'élévation [Nashashibi 93] ou d'une carte de régions étiquetées en fonction de la navigabilité [Lacroix 94] et la modélisation incrémentale de l'environnement avec une description de la scène par objets [Betge-Brezetz 95],
- la localisation du robot peut être réalisée à partir de la carte d'élévation sur la base d'une modélisation par B-splines et extraction de pics [Fillatreau 93] ou par corrélation [Nashashibi 93], ou par appariement d'objets [Betge-Brezetz 96],
- la planification de chemin pour définir les sous-buts [Lacroix 95] et de trajectoires dans le cas d'un terrain accidenté ou plan [Dacre-Wright 93],
- mais aussi l'architecture logicielle embarquée [Fleury 96, Chatila 93] et la structure décisionnelle...

- et enfin le traitement d'images, peu développé jusqu'alors et dans lequel s'intègrent nos travaux avec la stéréovision [Lasserre 95] actuellement en cours d'intégration pour le nouveau robot LAMA, la segmentation d'images couleur [Lasserre 96] et l'identification des régions segmentées. D'autres travaux sont aussi actuellement à l'étude comme le suivi d'amers naturels [Vandapel 96].

I.4 Conclusion

Au cours de ce chapitre nous avons relaté les deux grands axes de la robotique mobile en environnement d'extérieur. L'un concerne l'aide à la conduite automobile, l'autre la robotique d'intervention.

Les capteurs de perception diffèrent en fonction de l'objectif fixé. Dans le cadre du suivi de route, la vision et la couleur apparaissent comme les moyens privilégiés pour percevoir. En revanche, dans le cadre de la robotique d'intervention, et en particulier pour l'exploration planétaire, la vision reste très peu employée : son utilisation est généralement limitée à la vision achrome. Elle est uniquement envisagée pour remplacer le télémètre laser mais est rarement exploitée pour apporter des informations complémentaires et pour l'interprétation de la scène. Or, suivant le type d'environnement dans lequel doit se déplacer le robot, en particulier terrestre, des informations telles que la couleur ou la texture permettraient d'enrichir significativement la connaissance de l'environnement.

Au sein du laboratoire, le projet EDEN a déjà donné lieu à des expérimentations et ont permis de démontrer les capacités de déplacement de ce robot. Cependant, la perception de l'environnement est uniquement réalisée à partir de la télémétrie laser. L'environnement d'EDEN est pourtant très riche en végétation et la vision devrait y avoir sa place.

Chapitre II

Vision : l'approche proposée

La richesse des informations susceptibles d'être fournies par la caméra vidéo fait des images acquises par celui-ci une option sérieuse pour améliorer la connaissance de l'environnement.

Dans ce chapitre, nous nous attacherons donc tout d'abord à développer l'intérêt que peut avoir la vision dans le cadre du projet EDEN. Nous montrerons que le développement de l'autonomie du robot passe par l'utilisation de la vision achrome et trichrome. Nous proposerons ensuite notre approche de la vision. Celle-ci cherchera à répondre aux besoins spécifiés en tenant compte des différents travaux déjà menés dans le cadre de la modélisation de l'environnement.

II.1 Position du problème

L'autonomie du robot réside principalement dans sa capacité de percevoir et d'appréhender son environnement. La modélisation de l'environnement est en effet à la base des décisions sur le mode de déplacement du robot, sa localisation dans l'environnement, la planification de sa trajectoire...

II.1.1 Modélisation de l'environnement dans le cadre d'EDEN

La solution adoptée dans le cadre d'EDEN pour répondre à ces différents besoins est d'exploiter plusieurs représentations de l'environnement spécifiques à la tâche à réaliser. La construction de ces représentations dépend du mode de fonctionnement du robot.

Certaines seront construites à chaque étape, d'autres uniquement en cas de nécessité.

Trois modes de fonctionnement ont été définis :

- le mode «réflexe» : lorsque le terrain est considéré comme quasiment libre de tout obstacle ;
- le mode «planifié 2D» : qui est utilisé lorsque le terrain est essentiellement plan mais sur lequel des obstacles peuvent gêner le déplacement du robot ;
- le mode «planifié 3D» : lorsque la zone est accidentée.

La stratégie de déplacement du robot et son choix de fonctionnement se basent sur la représentation de l'environnement proposée par [Lacroix 95].

Elle fournit rapidement une description qualitative de l'environnement : l'image de points 3D est discrétisée en cellules dont la taille varie en fonction de la distance par rapport au capteur et des caractéristiques de ce dernier. Chaque cellule est affectée à une classe définissant le type de terrain : plat, accidenté, obstacle ou inconnu.

Le modèle de l'environnement est finalement composé de trois représentations dont les niveaux d'abstraction sont complémentaires : le «**bitmap**» qui regroupe l'ensemble des informations obtenues par la classification sur une zone particulière de l'image, le **carte de régions** et le **graphe de connexion** qui gère la globalité des différentes zones perçues. C'est donc à partir de cette modélisation que sera précisé s'il est nécessaire ou non de faire appel à d'autres représentations. Si les obstacles peuvent gêner le déplacement du robot ou que le terrain est trop accidenté, il est nécessaire de faire appel à une description plus détaillée du terrain, de le localiser dans l'environnement et de le recalculer si nécessaire.

Dans le cas d'un terrain très accidenté, une trajectoire précise doit être planifiée. Il est donc nécessaire de posséder une description surfacique du terrain. Celle-ci est fournie par le **Modèle Numérique du Terrain (ou MNT)** développé par [Nashashibi 93] par interpolation directe sur les données 3D.

La localisation du robot est assurée par la construction du **modèle B-splines** : à partir du MNT, des zones d'intérêt sont détectées afin d'effectuer le recalage du robot [Fillatreau 93].

Lorsque le terrain est plan mais que des obstacles sont présents, la planification d'une trajectoire doit aussi intervenir. L'utilisation du MNT n'est cependant pas nécessaire. Une autre représentation, demandant moins de volume mémoire, est mise en œuvre : il s'agit d'une représentation de l'environnement par objets développée par [Betge-Brezetz 96].

Le **modèle géométrique des objets** et le **modèle topologique** associé sont basés sur une représentation discrète de l'environnement. L'environnement est tout d'abord décomposé en parties du sol (considérées comme des plans) ou objets (leur forme est définie à partir de super-quadriques) que le modèle topologique permet de relier (des relations simples entre les différents objets telles que *devant*, *derrière*,... ont été définies).

L'ensemble de ces représentations permet ainsi de décrire l'environnement de manière

plus ou moins précise en fonction des nécessités. Un schéma regroupant toutes ces représentations est présenté un peu plus loin dans ce chapitre (voir section II.3.1 figure II.2).

II.1.2 Les limites actuelles

La modélisation de l'environnement telle qu'elle est actuellement n'exploite que l'information tridimensionnelle acquise par télémétrie laser. Deux problèmes essentiels apparaissent :

- le temps d'acquisition d'une image de points 3D à partir de la caméra laser placée sur ADAM est trop long : le télémètre laser requiert 20 ms pour acquérir un point soit plus de trois minutes pour obtenir une image de 100 lignes et 100 colonnes ;
- toutes les représentations proposées exploitent uniquement l'information tridimensionnelle : elles utilisent donc uniquement des informations d'ordre géométrique ou topologique. Or, nous montrerons dans la section suivante que ce type de modélisation peut s'avérer insuffisant.

Le problème de l'utilisation de la télémétrie laser n'est pas vraiment spécifique de ce projet. S'il est vrai que ce télémètre laser est très lent, il a l'avantage d'être précis dans le sens où la résolution de l'image est assez importante pour que l'image de points 3D soit assez dense. Un compromis se pose toujours entre la rapidité d'acquisition et la densité de points 3D.

D'autres limitations interviennent aussi comme son prix, son poids et sa consommation d'énergie. En revanche, ce capteur a l'avantage de ne pas dépendre de la luminosité ambiante, spécificité intéressante en environnement naturel.

Le deuxième problème posé, bien que spécifique de l'environnement dans lequel doit évoluer le robot, amène à une question plus générale : quelle connaissance doit avoir le robot de l'environnement qui l'entoure ? Doit-elle rester limitée à la zone dans laquelle il évolue et découvrir l'environnement au fur et à mesure de son avancée ? Peut-on avoir une vision plus globale de l'environnement ?

L'autonomie du robot dépend bien entendu de sa capacité à connaître au mieux l'environnement même si ce n'est pas toujours avec précision. Dans ce cas, d'autres capteurs doivent être exploités pour améliorer la connaissance de celui-ci.

Nous allons donc préciser pourquoi nous avons choisi le capteur caméra et comment il peut répondre à ces problèmes.

II.2 Pourquoi la vision ?

L'environnement sur lequel se déplace ADAM est un environnement non structuré que ce soit au niveau des objets ou du terrain à traverser. Dans ce contexte, le capteur caméra vidéo fournit des informations riches, mais plus bruitées que la caméra laser. Nous entendons par là que, pour des images vidéo, l'information utile est «noyée» dans la

masse et l'extraction de primitives appropriées ou d'attributs caractéristiques s'avère être généralement difficile. Ceci explique en partie les raisons pour lesquelles l'utilisation de ce capteur reste peu développée en robotique mobile.

Cependant, malgré les difficultés énumérées, le capteur caméra intéresse toujours la communauté scientifique :

- les informations issues des deux capteurs (caméra laser et vidéo) sont complémentaires,
- la caméra peut voir loin contrairement à la télémétrie laser (dont la portée est souvent limitée à une dizaine de mètres pour des raisons de sécurité oculaire),
- le coût d'une caméra vidéo est bien moindre...

Présentons tout d'abord les différents apports que ce capteur pourrait amener.

II.2.1 Une meilleure connaissance de l'environnement «proche»

Les étapes de localisation et de navigation nécessitent une bonne connaissance de l'environnement dans lequel évolue le robot.

Nous avons précisé précédemment que l'information tridimensionnelle apportée par le télémètre laser permet d'extraire des informations sur la géométrie du terrain, la forme et la surface des objets, mais ne permet pas toujours d'obtenir une description suffisante de l'environnement.

La figure II.1 montre deux cas où la modélisation de l'environnement, telle qu'elle est à ce jour, peut poser des problèmes : plusieurs objets peuvent avoir la même forme, le terrain peut s'avérer impraticable même s'il est plat...

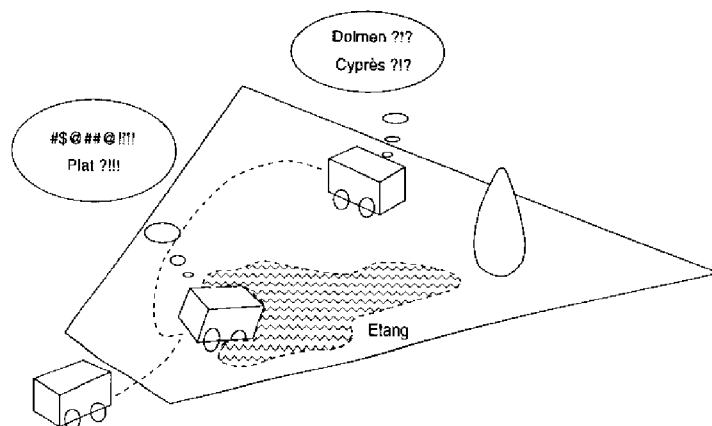


FIG. II.1 – Un environnement mal connu

Finalement, tous les objets sont considérés comme identiques dans la mesure où l'on ne connaît pas leur nature : un objet est seulement défini par le fait qu'il ne s'agit pas du sol ; de même, le sol n'est pas considéré comme un obstacle lorsqu'il est plat alors qu'il peut le devenir.

Pour améliorer la connaissance du terrain ou des objets situés dans l'environnement, il est nécessaire d'ajouter d'autres informations telles que la couleur ou la texture. Cela permet en effet d'apporter une information de type sémantique qui peut augmenter les capacités à identifier la nature des objets ou du terrain à traverser : distinction entre objets (rochers, arbres...), entre différents types de terrains (herbe, terre...).

Nommer un objet ou le sol a plusieurs intérêts :

- la tâche de navigation sera plus fiable : l'apport de la vision peut permettre d'améliorer la définition de la « praticabilité » du terrain en tenant compte du fait que l'on se déplace sur de l'herbe ou de la terre...
- la tâche de localisation sera plus précise car elle permettra de mieux distinguer les différents objets de la scène,
- il sera possible d'envisager des amers ou cibles naturelles de type arbre ou roche lorsque celles-ci sont discriminantes (un arbre dont le feuillage serait différent des autres, un tas de roches...).
- un modèle sémantique de la scène pourrait être construit : les éléments de la scène seraient nommés par des identificateurs symboliques simples ou composés afin de faciliter la communication avec le niveau décisionnel. Ces identificateurs indiqueront la nature ou des caractéristiques spécifiques intéressantes (« arbre », « roche noire »).

L'objectif qu'il faudrait finalement atteindre est d'affiner la connaissance de l'environnement afin de trouver les amers appropriés à la localisation, c'est-à-dire qui remplissent les conditions suivantes :

- être discriminants par rapport aux autres objets de la scène : par exemple une seule roche de couleur noire dans l'environnement alors que les autres sont grises.
- avoir une incertitude faible à la fois sur leur reconnaissance et sur leur position, le procédé de localisation étant incrémental.

La discrimination des objets ou du sol passe d'abord par l'identification de leur nature avant de rechercher leurs spécificités.

L'ensemble de ces informations (3D, couleur, texture) est complémentaire. Il permet de mieux décrire les objets constituant la scène : le modèle associé sera donc plus riche. Il permettra une meilleure localisation du robot et la stratégie de navigation sera plus adaptée au robot puisque la nature du terrain à traverser sera connue.

II.2.2 Connaissance de l'environnement lointain

Le capteur caméra vidéo a une qualité non négligeable : il a une profondeur de champ de vue telle qu'il peut « voir loin ». Si l'on souhaite que le robot soit autonome, il faudra lui indiquer des cibles naturelles, qu'elles soient éloignées (pour indiquer une direction à prendre), ou proches (pour atteindre un but précis). Cette étape nécessitera obligatoirement la prise en compte d'un tel capteur.

L'extraction d'amers proches peut se faire en coopération avec l'information tridimensionnelle obtenue par ailleurs. En revanche, les amers lointains ne pourront être obtenus qu'à partir d'une image issue d'une caméra.

Considérons l'exemple suivant où, dans une scène, on apercevrait au loin une haie d'arbres dont un domine par sa hauteur. Il serait intéressant, au niveau du modèle, d'une part de faire apparaître cette information et d'autre part d'indiquer approximativement sa distance par rapport au robot.

Si nous supposons que la région correspondant à la haie a pu être extraite, nous pouvons imaginer qu'il sera possible de fournir ces deux indications :

- il est possible d'obtenir l'information tridimensionnelle par triangulation à partir d'un système stéréoscopique : la stéréovision par corrélation sur quelques points de la ligne d'horizon ou la stéréovision en utilisant des primitives de plus haut niveau pourraient permettre d'estimer la distance ;
- les caractéristiques (colorimétriques et peut être de texture) pourraient servir à déterminer la nature de la région.

II.2.3 En conclusion...

La limite actuelle de l'utilisation de la vision dans des applications de robotique mobile n'est pas seulement due à la difficulté de traitement des données mais aussi à leur faible rapidité d'exécution.

Les progrès actuels en informatique et en électronique (parallélisme, carte câblée...) permettent d'espérer dans un avenir proche des coûts calculatoires faibles pour le traitement des images. L'utilisation de la vision devrait alors s'accroître de manière significative dans ce type d'applications.

Nous sommes convaincus de l'intérêt d'une telle approche et en particulier de l'approche couleur lorsque le robot doit se déplacer dans un environnement naturel. Les informations de couleur et de texture que peut fournir le capteur caméra vidéo sont indispensables pour acquérir l'autonomie nécessaire dans ce type d'environnement. Ces informations sont complémentaires de celles fournies par le télémètre laser. Elles apportent un niveau d'abstraction supplémentaire au modèle par identification de la nature des éléments constituant la scène.

D'autre part, la capacité de ce capteur à « voir loin » est un atout majeur pour l'évolution future des travaux dans ce domaine.

II.3 Approche proposée

Le choix de l'approche vision à développer doit répondre à plusieurs conditions :

- répondre en premier lieu au problème de l'acquisition de l'information tridimensionnelle,
- deuxièmement, tenir compte des deux aspects, connaissance de l'environnement proche et lointain dans le choix des traitements futurs,
- et enfin, l'approche développée doit pouvoir s'intégrer aux travaux déjà menés en modélisation de l'environnement dans le cadre d'EDEN.

Nous allons à présent montrer comment l'approche proposée prend en compte ces critères.

II.3.1 Présentation générale

Le schéma de la figure II.2 montre les différentes représentations de l'environnement utilisées dans le cadre d'EDEN.

Ces différentes représentations ont été détaillées au début de ce chapitre (en section II.1.1) et sont représentées en noir. Les flèches indiquent les données nécessaires à la construction des représentations.

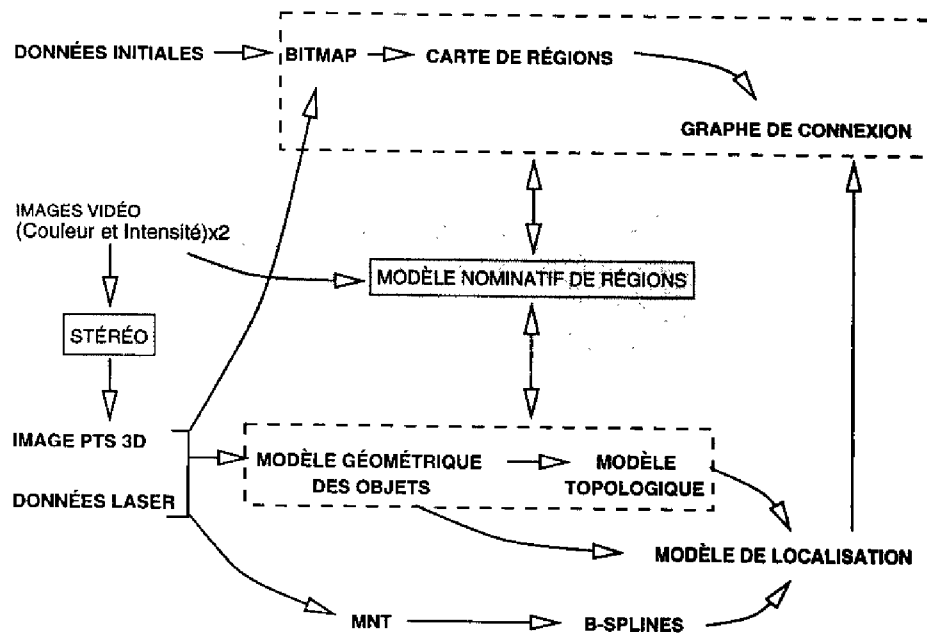


FIG. II.2 - Les différentes représentations de l'environnement

Nous avons représenté sur ce même schéma, en gris, comment peut intervenir la vision. Deux axes se distinguent :

1. fournir l'information tridimensionnelle à partir d'un système stéréoscopique;
2. apporter une nouvelle représentation de l'environnement qui enrichisse les autres représentations existantes : le **modèle nominatif de régions**.

Il consisterait à fournir une description très synthétique de l'image en dégagant les principaux éléments constituant l'image ainsi que leur nature (roche, arbre...). La nature des éléments constituant l'environnement est en effet la principale information manquant aux modèles existants.

Plusieurs questions se posent lorsqu'on souhaite développer un tel système :

- doit-on définir un nouveau modèle, ou doit-on compléter les modèles existants? En particulier, donner la nature suppose que l'on a des régions à identifier. D'où doit-on les extraire?
- quelles classes définir dans la base de connaissances?
- quels attributs choisir pour caractériser les régions?

Nous allons donc répondre à ces questions. Notons déjà que la nécessité de travailler sur des régions va privilégier deux des représentations existantes comme l'on peut le voir sur la figure II.2 : le *modèle « bitmap »* et le *modèle géométrique des objets* auxquels nous ferons de nombreuses références tout au long de ce manuscrit.

II.3.2 Un nouveau modèle

L'extraction des régions peut se faire par divers traitements et à partir de différentes données :

- A partir de l'image de points 3D : le modèle par objets fournit des régions étiquetées sol ou objets. Cependant, nous l'avons déjà précisé, cette décomposition est insuffisante et demande une nouvelle segmentation pour préciser les différentes parties du sol. La discrétisation en cellules est plus intéressante car les régions sont plus petites (de l'ordre de 50x50 pixels) mais assez significatives pour être par la suite identifiées.

L'inconvénient lié à ces deux méthodes est cependant qu'elles ne s'intéressent pas à toute la scène. Les régions éloignées pour lesquelles l'information tridimensionnelle n'est pas connue ne sont pas prises en compte. Dans ce cas, il est nécessaire d'utiliser une autre source d'information.

- A partir d'une image caméra : l'intérêt principal est bien entendu que l'ensemble de la scène est segmenté contrairement au cas précédent. La difficulté réside par contre dans le choix du critère de segmentation pour obtenir la description recherchée.

Au vu de ces remarques, nous pouvons déjà voir la nécessité de définir un nouveau modèle si l'on souhaite pouvoir exploiter au mieux le capteur caméra. L'extraction des régions à partir d'une image vidéo peut être obtenue dans notre cas, soit à partir d'une image intensité, soit à partir des trois plans d'une image couleur. L'utilisation de la luminance n'est pas toujours suffisante pour obtenir une bonne description de la scène comme nous le montrerons en annexe A. Nous avons donc choisi de segmenter l'image à partir des informations colorimétriques.

En revanche, nous devons prendre en compte la possibilité d'exploiter les régions définies par ces deux modèles. Ceci imposera que les algorithmes puissent être utilisés aussi bien sur la totalité de l'image que sur une partie de celle-ci mais aussi que l'identification des régions soit effectuée indépendamment du type de segmentation utilisée.

II.3.3 Quelles classes dans la base de données ?

L'environnement et le but poursuivi interviennent dans le choix des classes. Elles doivent, en effet, être adaptées à l'application pour laquelle elles sont produites. La navigation et la localisation, suivant les niveaux auxquels on va intervenir, ne nécessitent pas obligatoirement les mêmes classes :

- Au niveau de la localisation : l'intérêt principal est d'identifier la nature des objets pour ensuite rechercher leurs spécificités. Compte tenu de l'environnement dans lequel doit se déplacer le robot, les classes à définir correspondent aux principaux types d'éléments constituant les scènes.
- Au niveau de la navigation, deux possibilités peuvent être envisagées :
 1. redéfinir des classes du type «navigable», «praticable» et «non-praticable» directement : ceci suppose qu'on limite la base de données à un type de robot, d'où une simplicité de mise en œuvre non négligeable. En revanche, la base de données est à redéfinir sitôt que les capacités de navigation du robot sont différentes ;
 2. conserver les différentes classes concernant les types de terrain traversés (par exemple «herbe», «gravier», «terre», etc.), puis redéfinir les nouvelles classes de «praticabilité» en combinant les premières classes à celles fournies par le «bitmap» concernant la navigabilité (exprimée sous la forme «plat», «accidenté» ou «obstacle») et aux limites du robot.

Cette dernière idée est la plus intéressante car elle est générique. Outre qu'elle permet de conserver la même base de données pour la modélisation et la navigation, elle amène une souplesse importante au système : il sera facile de passer d'un environnement à un autre, ou de modifier le type de robots utilisés, sans avoir à redéfinir entièrement la base de données et les classes ; il suffit de redéfinir les règles de «praticabilité». Mais c'est aussi une méthode lourde à mettre en œuvre et demandant une réflexion approfondie pour la définition des règles.

Pour toutes les raisons citées précédemment, nous avons choisi de nous placer dans le cas de la deuxième option.

Le terrain sur lequel ADAM se déplace est un terrain accidenté comportant des rochers qui sont des obstacles infranchissables pour le robot et donc à contourner. Les images des scènes correspondantes comportent essentiellement de l'herbe, de la terre, des rochers, des arbres et le ciel ; nous avons donc défini ces cinq éléments comme étant les classes à générer dans la base de connaissances.

II.3.4 Quels attributs pour la caractérisation ?

Les régions issues de la segmentation doivent ensuite être caractérisées. Nous avons choisi de compléter les informations de couleur en ajoutant des informations de texture.

L'utilisation de la texture est intéressante pour améliorer la caractérisation des régions car elle représente une information non prise en compte par ailleurs.

En revanche, nous ne sommes pas convaincus de l'intérêt de segmenter à partir des attributs que nous avons choisis. Sans entrer dans la définition de ceux-ci, plusieurs remarques permettent de préciser nos raisons :

- les attributs utilisés se calculent sur une fenêtre 16x16. La segmentation résultante est donc grossière. Ceci n'est pas gênant en première approximation mais peut le devenir si l'on souhaite affiner le modèle ;
- l'image doit être traitée à la pleine résolution : en liaison avec la remarque précédente, pour que l'information de texture puisse avoir un sens, il est difficile d'imaginer de travailler sur l'image réduite.
De plus, le nombre d'attributs calculés est important (6 attributs définis - cf. chapitre IV, section IV.3.2) et ils sont complémentaires. Segmenter à partir de l'un d'eux n'est pas suffisant. Les temps de calcul pour la segmentation seraient donc importants.
- la couleur est le seul critère que l'on puisse considérer comme indépendant de la distance à laquelle sont perçus les différents objets.

Cette approche de segmentation par attributs de texture a été testée par R. Murrieta [Murrieta Cid 95] et confirme les remarques faites ici.

Nous avons donc choisi de segmenter l'image à partir de la couleur uniquement, puis de compléter la caractérisation grâce à la texture.

II.3.5 Construction du Modèle Nominatif de Régions

Nous savons à présent quelles seront les principales caractéristiques de ce modèle. Les raisons développées précédemment nous incitent à construire le modèle en suivant le fonctionnement présenté sur la figure II.3.

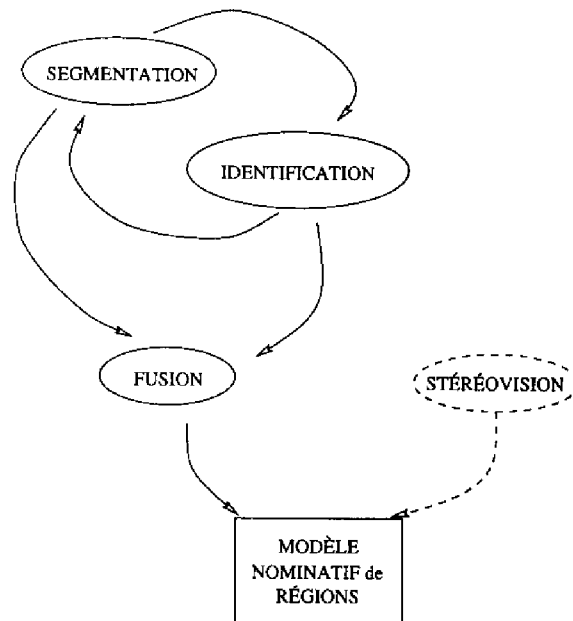


FIG. II.3 – Construction du Modèle Nominatif de Régions

Le processus «segmentation-identification» devrait boucler sur lui-même jusqu'à ce que les régions soient identifiées avec une confiance suffisante, et que la segmentation n'apporte plus d'information.

A l'issue de cette étape, les régions adjacentes de même nature et qui seraient identifiées de manière sûre seraient fusionnées.

Le modèle serait donc une image de régions dont on connaîtrait la nature et la confiance que l'on peut accorder à ce résultat.

L'information tridimensionnelle obtenue par stéréocorrélation pourrait être associée à ces régions.

L'idée principale étant définie, nous devons à présent tenir compte de différents aspects liés aux besoins définis.

Un aspect d'adaptabilité

L'identification des régions est indépendante de la technique de segmentation employée. En revanche, les régions obtenues seront de nature différente et de tailles différentes suivant qu'elles auront été obtenues à partir de l'information tridimensionnelle (stéréocorrélation) ou de la couleur.

De même, la segmentation couleur devra pouvoir être effectuée à différents niveaux : sur la totalité de l'image pour en dégager les principales composantes, sur une zone spécifique proche (pour l'affinement d'une région par exemple) ou lointaine (si l'on souhaite unique-

ment s'intéresser à cette zone).

Finalement, le modèle final pourra être directement associé à un modèle déjà existant, et donc compléter son information, ou être indépendant, dans le cas où la segmentation couleur serait utilisée.

Un aspect multi-résolution

Les différents algorithmes doivent pouvoir exploiter les images à différentes résolutions :

- la stéréocorrélation utilise généralement les images intensité à faible résolution afin de limiter le temps de calcul. La résolution choisie devra dans un premier temps fournir un nombre de points 3D proche de celui fourni par la caméra laser. Par la suite, si une zone de l'image avait besoin d'être affinée, on pourrait choisir de se placer à une résolution supérieure ;
- la segmentation couleur doit, elle aussi, être faite en utilisant une résolution inférieure. Il n'est pas nécessaire d'avoir une énorme résolution pour dégager les principales composantes de l'image. De plus, les temps de calcul deviendraient trop importants pour que le procédé de segmentation soit exploitable ;
- l'identification des régions, au contraire, a besoin du maximum d'informations pour caractériser au mieux les régions issues de la segmentation. Dans ce cas, on se ramène à la pleine résolution de l'image.

La figure II.4 récapitule les différentes précisions fournies ici.

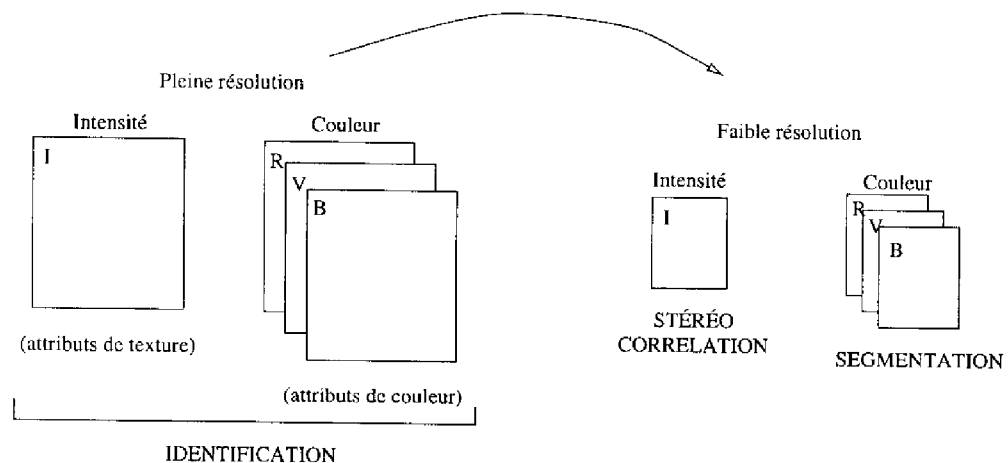


FIG. II.4 - Les différentes résolutions des images

Les images initiales sont de tailles 576x502 (pour les paires stéréoscopiques) ou 604x804 (cas du deuxième système d'acquisition).

En réduisant quatre fois ces images, on se retrouve, pour la faible résolution, avec des images de tailles plus raisonnables (respectivement de 144x125 et de 151x201 pour les deux systèmes d'acquisition).

Finalement, le modèle fourni sera de même taille que la segmentation produite.

II.4 Conclusion

Nous avons présenté au cours de ce chapitre les raisons pour lesquelles nous avons choisi la vision pour enrichir le modèle de l'environnement.

Elles sont principalement de deux ordres :

- la vision peut apporter l'information tridimensionnelle de manière plus dense et plus rapide que le télémètre laser,
- elle permet aussi d'identifier la nature des éléments contenus dans la scène à partir d'informations sur la couleur et la texture de ces derniers.

Ces raisons nous amèneront à travailler dans ces deux directions en développant tout d'abord un algorithme de stéréovision par corrélation.

Puis, nous apporterons une nouvelle représentation de l'environnement de manière à enrichir celles existantes. Cette représentation décrira la scène en régions dont la nature sera identifiée. Ce type d'information sera utile aussi bien pour la localisation que pour la navigation du robot dans son environnement.

La suite de ce manuscrit sera donc consacrée au développement de ces deux parties. Le prochain chapitre traitera le problème de la stéréovision. Les deux chapitres suivants présenteront plus précisément les phases de construction du modèle.

Le chapitre IV détaillera les différentes techniques utilisées pour segmenter l'image ainsi que la façon dont les régions obtenues seront caractérisées.

Le dernier chapitre (chapitre V) s'intéressera plutôt à l'identification des régions et à la construction du modèle final.



Chapitre III

La stéréovision

La stéréovision est une technique connue depuis longtemps mais où des améliorations doivent encore être apportées sur le temps de calcul, la précision des données....

Dans ce chapitre, nous présenterons brièvement les différentes approches en stéréovision ainsi que les domaines dans lesquels elles sont utilisées. La stéréocorrélation est plus adaptée en environnement naturel où les images sont texturées. Nous développerons donc cette méthode et préciserons les apports que nous avons amenés pour améliorer la rapidité de l'algorithme. Nous verrons en particulier que les images peuvent être rectifiées avec un minimum de déformations et qu'il est possible de conserver plusieurs critères lors de l'appariement sans pour autant augmenter le temps de calcul.

III.1 Présentation générale

III.1.1 Principe de la stéréovision

La stéréovision est basée sur le principe que l'information de profondeur peut être obtenue par triangulation à partir de deux images ayant une partie commune dans leur champ de vue.

Avant de préciser comment l'information tridimensionnelle est obtenue, nous souhaitons rapidement présenter sur quel concept se base ce type de traitement. Le schéma de la figure III.1 illustre rapidement comment la variation des coordonnées d'un pixel et de son correspondant dans les deux images est liée à la position et donc à la profondeur de l'objet auquel il correspond.

Sur ce schéma est représentée une vue de dessus du champ de vue des deux caméras (représentées par leur centre optique C_1 et C_2 et les plans images - seules les colonnes des

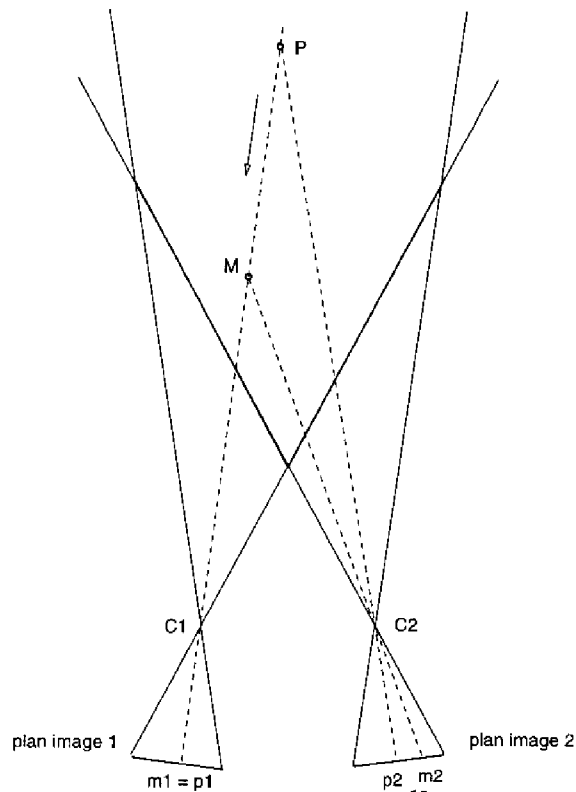


FIG. III.1 – Principe de la stéréovision

images apparaissent par projection).

Soit P un point d'un objet de la scène situé à une distance donnée des caméras. Il se projette respectivement en p_1 et p_2 sur les plans image des caméras 1 et 2.

Supposons à présent que l'on rapproche cet objet jusqu'à ce que le point P se trouve en M et de telle manière que sa projection dans le plan image 1 reste sur le même indice colonne comme le montre la figure.

On peut immédiatement remarquer que dans le plan image de la caméra 2, ce point va voir son indice colonne varier (différence entre m_2 et p_2). Ainsi, *une variation de distance du point dans la scène correspond à une variation de l'indice colonne de la projection dans le plan image*. Ce principe peut s'appliquer globalement sur toute l'image.

Quittons le domaine dynamique pour revenir au statique et intéressons nous à présent uniquement au point M et à sa projection sur les deux plans images m_1 et m_2 . Les coordonnées des points projetés diffèrent dans les deux images. Cette différence est aussi appelée *disparité*. A partir du modèle des caméras et de leur position dans un repère lié au monde (phase de calibration [Garric 96]), on peut montrer que cette grandeur est inversement proportionnelle à la profondeur. En déterminant la disparité de tous les pixels communs aux deux images, on pourra obtenir une carte dense de points 3D.

III.1.2 Les principales approches

Différentes approches sont utilisées en stéréovision. Elles consistent à utiliser des primitives différentes pour l'appariement (régions, contours, pixels). Suivant le type de scènes traitées, elles sont plus ou moins adaptées.

Utilisation de primitives de haut niveau

Elles sont généralement utilisées dans le cas d'environnements structurés. Ce type de méthodes consiste en effet à appairier des primitives de haut niveau comme les régions ou les segments. Elles sont particulièrement adaptées pour des scènes d'intérieur où les contours sont essentiellement des segments de droite, mais le sont peu dans le cas d'environnements naturels :

- les régions peuvent changer de forme d'une image à l'autre : un exemple simple est de considérer deux tas de cailloux situés à des distances différentes. Il peut se trouver que sur la première caméra, les deux tas soient fusionnés car l'un est derrière l'autre, alors que sur la seconde, la distinction existe ;
- les objets naturels ont rarement des surfaces planes ;

Nous n'insisterons donc pas davantage sur ce type de méthodes. Le lecteur intéressé pourra trouver dans [Ayache 89] une revue détaillée des méthodes existantes.

Ces méthodes ont été étudiées dans le groupe dans le cas d'environnements d'intérieur : A. Robert de St Vincent a proposé un système de stéréovision basé sur l'appariement de segments afin de construire une carte planaire de l'espace occupé dédiée à la navigation [Robert de Saint Vincent 86] ; T. Baron et V. Coutance ont étudié la coopération entre régions couleurs et contours pour améliorer l'appariement [Baron 91, Coutance 91].

Notons que ces techniques pourraient toutefois être appliquées aux scènes d'extérieur dans certains cas particuliers : si certains objets sont structurés (par exemple un bâtiment), ces traitements seront plus adaptés que la stéréocorrélation.

Il est aussi envisageable de considérer, lorsque les régions sont éloignées, qu'elles se situent dans un plan vertical (par exemple pour une haie d'arbres). Il sera alors possible d'utiliser ce type de méthodes pour donner une approximation grossière du plan dans lequel elles se trouvent. Il faudra cependant que le nombre de régions candidates soit limité et qu'elles puissent être bien caractérisées pour éviter les erreurs d'appariement. Pour cela, on pourrait s'appuyer sur un système hiérarchique comme celui proposé dans [Marapane 94].

Stéréocorrélation

La stéréocorrélation se base sur l'appariement de pixels pour obtenir une carte dense de points 3D. Elle est particulièrement adaptée lorsque les images à traiter sont texturées comme dans le cas d'images de scènes naturelles. Elle est d'ailleurs largement utilisée en robotique mobile en environnement naturel [Boissier 92, Matthies 92] car elle permet

d'obtenir des images de points 3D plus denses que celles fournies par la télémétrie laser dans des temps réduits.

Les problématiques liées à la stéréocorrélation sont de deux ordres :

1. Améliorer et évaluer la précision du résultat : différentes possibilités permettent d'améliorer la qualité du résultat.

Tout d'abord, les pré-traitements de l'image, laplacien de gaussienne ou équivalent [Bracco 92] permettent de faire ressortir la texture de l'image.

L'appariement peut aussi être conditionné par plusieurs critères : vérification de l'appariement par un deuxième passage en inversant le rôle des deux caméras [Fua 93], utilisation de fenêtres adaptatives [Kanade 94] ou d'une troisième caméra pour valider l'appariement [Hotz 91, Ayache 88]...

D'autres solutions consistent enfin à filtrer l'image de disparité pour ôter les mauvais appariements : filtre médian, modélisation de la carte de disparité [Eastman 87, Maître 92]...

2. Améliorer les temps de calcul : un compromis est nécessaire entre la qualité requise par les traitements de localisation et de navigation.

Au JPL, Matthies a proposé un système de stéréocorrélation très rapide, basé sur une étude des images à faible résolution [Matthies 92] et sur l'utilisation d'une technique probabiliste pour estimer les valeurs de disparité. L'intérêt de sa méthode est qu'elle fournit directement une information sur la validité de l'algorithme.

Notre objectif a donc été de développer un algorithme qui tienne compte de ces deux aspects. L'algorithme que nous avons implémenté prend donc en compte différents critères pour l'appariement, en particulier le double passage pour valider l'appariement proposé par [Fua 93]. Nous montrerons que l'utilisation de ce test n'implique pas un doublement du temps de calcul. D'autres améliorations seront aussi apportées pour minimiser le temps de calcul et améliorer la qualité de l'image de disparité.

III.2 Méthode utilisée

Le procédé de stéréocorrélation peut se décomposer en trois phases :

- a- la rectification : elle permet de limiter la recherche de l'appariement à une ligne de l'image.

Pour cela, les deux plans images sont ramenés, en utilisant la géométrie épipolaire, dans un même repère tel que les deux caméras virtuelles soient parfaitement alignées ;

- b- la corrélation : elle correspond à la recherche des appariements des pixels des deux images.

Une fenêtre de corrélation permet de rechercher le meilleur appariement possible à partir d'un critère de ressemblance. La différence entre les indices colonne des deux

pixels appariés est appelée aussi *disparité*.

A l'issue de cette étape, une image de disparité est donc obtenue ;

- c- la reconstruction 3D: elle s'effectue à partir de l'image de disparité. Une simple relation de triangulation permet en effet de lier profondeur et disparité.

Nous allons donc à présent développer le raisonnement employé pour chacune de ces étapes et préciser les améliorations que nous avons apportées [Lasserre 95].

III.2.1 Rectification

Réaliser un alignement parfait des deux caméras est intéressant à deux niveaux : à l'étape de corrélation, les pixels correspondants seront sur la même ligne dans les deux images ; la relation qui lie disparité et profondeur sera très simple et permettra un calcul très rapide lors de la reconstruction.

L'obtention d'un alignement parfait des deux caméras est réalisé en utilisant les contraintes de la géométrie épipolaire et les modèles des caméras fournis par la calibration. Le problème consiste en fait à déterminer un modèle de caméra virtuel et un repère commun aux deux caméras.

Ce type de traitement introduit des déformations de l'image. Notre objectif a été de déterminer les meilleurs paramètres de rectification pour que les images soient altérées le moins possible et que la surface de l'image soit exploitée au mieux.

Le modèle de caméra

Comme dans [Ayache 88] nous avons utilisé le modèle *pinhole* obtenu par calibration. Ce modèle est un modèle linéaire qui ne prend pas en compte les distorsions des images (les images pourront cependant être corrigées lors de l'étape de rectification à partir des paramètres de distorsion fournis par la calibration).

Il est constitué par quatre vecteurs C , A , H et V , ou douze paramètres.

La projection d'un point 3D P sur le plan image en un pixel (i, j) est donnée par les équations suivantes :

$$i = \frac{(P - C).V}{(P - C).A} \quad \text{et} \quad j = \frac{(P - C).H}{(P - C).A} \quad (\text{III.1})$$

où C est le centre optique de la caméra, A est la direction de l'axe optique et H et V définissent le plan focal.

Le modèle est invariant à un coefficient près de A , H et V ; nous considérerons donc que A est un vecteur unitaire.

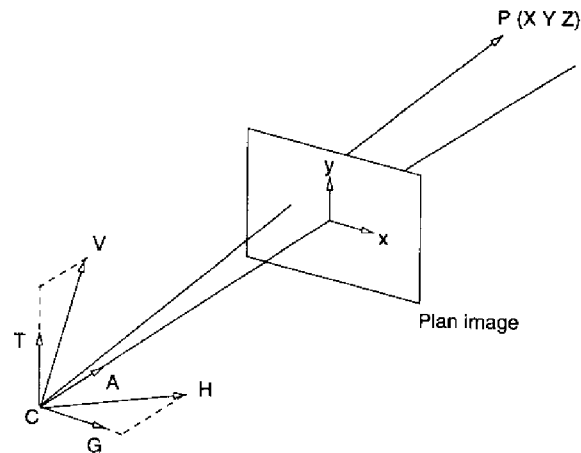


FIG. III.2 - Le modèle caméra CAHV

L'axe optique, défini par C et A , est projeté sur le plan image en un pixel de coordonnées $(H.A, V.A)$ que nous appellerons par la suite *centre de l'image*.

Les deux vecteurs $H - (H.A).A$ et $V - (V.A).A$ représentent alors les directions principales de la matrice CCD selon les facteurs d'échelle horizontaux et verticaux correspondants.

G et T sont les vecteurs unitaires correspondant à ces deux directions (voir figure III.2). Remarquons toutefois que G et T ne sont pas nécessairement orthogonaux.

Détermination du repère rectifié

Soient $C_1A_1H_1V_1$ et $C_2A_2H_2V_2$ les modèles respectifs des caméras gauche et droite. Nous cherchons finalement à représenter ces deux caméras virtuelles dans un repère commun tel qu'elles soient alors considérées comme deux caméras identiques parfaitement alignées.

L'alignement parfait des caméras virtuelles peut être réalisé sous les conditions suivantes :

- A est orthogonal à C_1C_2 (les lignes épipolaires sont parallèles) ;
- T est orthogonal à C_1C_2 (les lignes épipolaires sont horizontales).

De plus, pour simplifier par la suite les équations de triangulation, nous ajouterons la contrainte que G soit orthogonal à T . Donc G doit être colinéaire à C_1C_2 .

En outre, pour améliorer les résultats de la corrélation, il est nécessaire que les deux images soient les plus semblables possibles aux images originales. Pour cela, le repère AGT commun aux deux caméras devra être aussi proche que possible des deux repères caméra $A_1G_1T_1$ et $A_2G_2T_2$.

G , A et T sont obtenus de la manière suivante (voir figure III.3) :

- G est le vecteur unitaire colinéaire à C_1C_2 ,

- A est le vecteur unitaire orthogonal à G , proche de la bissectrice de A_1 et A_2 ,
- T est issu du produit vectoriel de A et G .

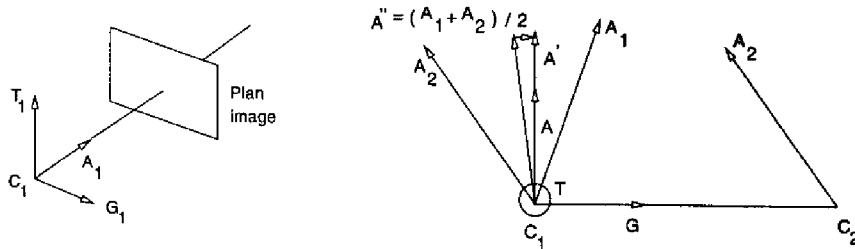


FIG. III.3 - Construction du repère rectifié

Détermination des paramètres de rectification

Généralement, les images rectifiées conservent les facteurs d'échelle des images originales mais elles sont très déformées par le changement de repère et sont plus grandes que les images originales pour conserver l'ensemble de l'image (voir figure III.4).

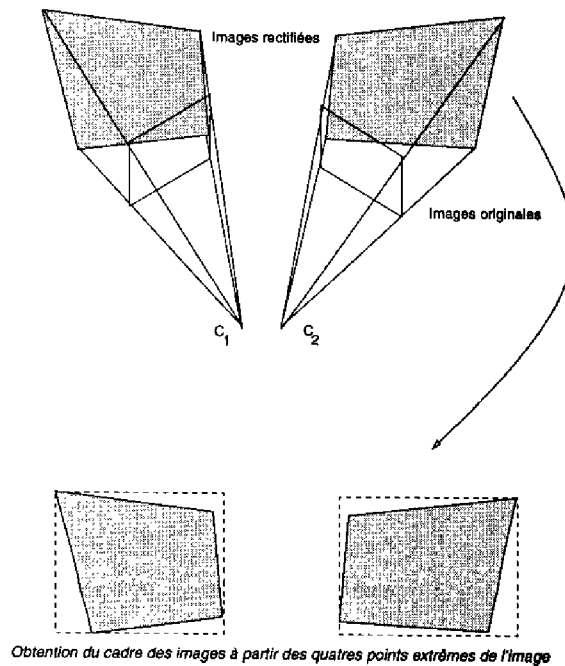


FIG. III.4 - Déformation des images rectifiées

Une autre solution consiste à préférer conserver des dimensions identiques pour les images originales et rectifiées et à déterminer les facteurs d'échelle et les centres des nou-

velles images. C'est la solution que nous avons retenue. Bien entendu, ceci suppose que certains pixels ne se trouveront plus à l'intérieur du cadre ; il faudra donc choisir ces paramètres de façon à maximiser la place occupée par l'image dans le cadre.

D'autre part, certaines contraintes devront être respectées :

- le facteur d'échelle vertical k'_i (correspondant aux lignes) et la coordonnée verticale du centre optique i'_0 devront être identiques pour les deux images rectifiées afin de vérifier la contrainte de correspondance ligne-à-ligne ;
- Le facteur d'échelle horizontal k'_j (correspondant aux colonnes) devra être identique pour les deux images. Ceci est intéressant pour deux raisons :
 1. les deux images seront ainsi plus semblables et c'est préférable pour la phase de corrélation ;
 2. cela permettra d'avoir des équations de triangulation plus simples ;
- Les coordonnées horizontales des deux centres optiques , j'_{01} et j'_{02} , devront être différentes dans les deux images.

Précisons à présent comment ces paramètres interviennent sur la forme de l'image rectifiée.

La droite 3D support d'un pixel (i_1, j_1) de l'image gauche peut être obtenue à partir des équations de projection :

$$P = C_1 + \alpha D_1(i_1, j_1)$$

avec $D_1(i_1, j_1) = (i_1 A_1 - V_1) \wedge (j_1 A_1 - H_1)$ (III.2)

De plus, ce même point P se projette sur l'image virtuelle en un pixel (i'_1, j'_1) selon les équations suivantes :

$$\begin{aligned} i'_1 &= \frac{k'_i(P - C_1)T}{(P - C_1)A} + i'_{01} \\ j'_1 &= \frac{k'_j(P - C_1)G}{(P - C_1)A} + j'_{01} \end{aligned} \quad \text{(III.3)}$$

Ces équations (III.2) et (III.3) permettent d'exprimer la relation entre un pixel de l'image originale et de l'image virtuelle, et en fonction de k'_i, k'_j, i'_0 et j'_{01} (toujours pour la caméra gauche) :

$$\begin{aligned} i'_1 &= k'_i \cdot K_i^1(i_1, j_1) + i'_{01} \\ \text{où} \quad K_i^1(i_1, j_1) &= \frac{D_1(i_1, j_1) \cdot T}{D_1(i_1, j_1) \cdot A} \end{aligned} \quad \text{(III.4)}$$

$$j'_1 = k'_j \cdot K_j^1(i_1, j_1) + j'_{01}$$

où
$$K_j^1(i_1, j_1) = \frac{D_1(i_1, j_1) \cdot G}{D_1(i_1, j_1) \cdot A} \quad (III.5)$$

La forme de l'image rectifiée est représentée dans ces équations par les coefficients $K^1(i, j)$ des quatre points extrêmes de l'image (l'exposant 1 correspond à la caméra gauche). Ces huit valeurs déterminent le polygone à l'intérieur duquel se trouve l'image.

Nous pouvons à présent déterminer les facteurs d'échelle et les centres des images qui maximisent la partie utile de l'image dans le cadre. Pour cela, nous recherchons les lignes et colonnes minimum et maximum correspondant à ce qui pourra être au mieux corrélé dans les images rectifiées (voir figure III.5).

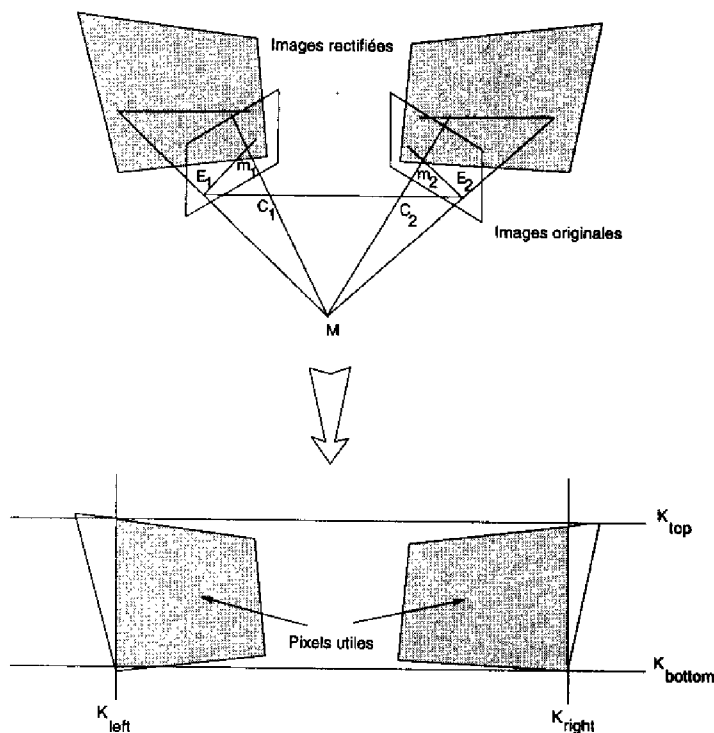


FIG. III.5 - Choix des contraintes pour la rectification

- Sur les lignes, nous déterminons les valeurs de $K_i^{1,2}(i, j)$ pour $i = 0, i = \mathcal{D}_i, j = 0$ et $j = \mathcal{D}_j$, pour les deux images ($\mathcal{D}_i, \mathcal{D}_j$ sont les dimensions de l'image et les indices 1, 2 représentent les caméras gauche et droite). On aura finalement :

$$K_{top} = \max(\min(K_i^1(0, 0), K_i^1(0, \mathcal{D}_j)), \min(K_i^2(0, 0), K_i^2(0, \mathcal{D}_j)))$$

$$\mathcal{K}_{bottom} = \min(\max(K_i^1(\mathcal{D}_i, 0), K_i^1(\mathcal{D}_i, \mathcal{D}_j)), \max(K_i^2(\mathcal{D}_i, 0), K_i^2(\mathcal{D}_i, \mathcal{D}_j)))$$

Comme le montre la figure III.5, cela revient à déterminer les plus petites (\mathcal{K}_{top}) et plus grandes (\mathcal{K}_{bottom}) lignes communes aux deux images rectifiées. D'autre part, nous souhaitons que ces valeurs soient les limites des images rectifiées, d'où en utilisant les relations exprimées par les équations III.4 :

$$k'_i \mathcal{K}_{top} + i'_0 = 0$$

$$k'_i \mathcal{K}_{bottom} + i'_0 = \mathcal{D}_i$$

La résolution de ces équations fournit les valeurs du facteur d'échelle vertical k'_i et de la coordonnée verticale des deux origines i'_0 :

$$i'_0 = \frac{\mathcal{D}_i}{1 - \frac{\mathcal{K}_{bottom}}{\mathcal{K}_{top}}}$$

$$k'_i = \frac{\mathcal{D}_i}{\mathcal{K}_{bottom} - \mathcal{K}_{top}}$$

- Le même principe est appliqué sur les colonnes pour déterminer le facteur d'échelle horizontal et les coordonnées horizontales des origines. Dans ce cas, seul le facteur d'échelle doit être identique pour les deux images ; les deux coordonnées des origines doivent être différentes. Il suffit pour cela, d'introduire cette différence dans le calcul de la disparité avant la triangulation.

Pour réduire le nombre de pixels non utilisés dans les images, nous avons choisi d'éliminer les colonnes de l'image gauche rectifiées et les colonnes de l'image droite rectifiée qui ne pourraient pas être entièrement corrélées (voir figure III.5).

De plus, la recherche du correspondant lors de l'étape de corrélation, se fait sur un intervalle de disparité $[d_{min}, d_{max}]$ (ou de distance correspondant $[D_{min}, D_{max}]$) fourni par l'utilisateur. La définition de la disparité est donnée par :

$$d = j'_1 - j'_2$$

Ces contraintes sont nécessaires pour calculer le critère de corrélation sur tout l'intervalle de disparité pour tous les pixels de l'image (voir équations III.5) :

- la première colonne de l'image rectifiée gauche devra être au moins la colonne d_{max} , donc on aura :

$$j'_1 = k'_j \mathcal{K}_{left} + j'_{01} = d_{max}$$

avec

$$\mathcal{K}_{left} = \max(K_j^1(0, 0), K_j^1(\mathcal{D}_i, 0))$$

- de même, la dernière colonne de l'image rectifiée droite sera au mieux la colonne $\mathcal{D}_j - d_{max}$, soit :

$$j'_2 = k'_j \cdot \mathcal{K}_{right} + j'_{02} = \mathcal{D}_j - d_{max}$$

avec

$$\mathcal{K}_{right} = \min(K_j^2(0, \mathcal{D}_j), K_j^2(\mathcal{D}_i, \mathcal{D}_j))$$

Nous avons choisi de fixer $d_{max} = 0$; cela signifie que nous aurons un intervalle de disparité négatif.

D'autre part, les équations de triangulation (voir III.2.3) fournissent la relation suivante :

$$d_{max} = \frac{k'_j \cdot B}{D_{min}} + (j'_{01} - j'_{02}) = 0$$

où B est la base du système stéréoscopique (distance entre les centres optiques des deux caméras, $B = \overline{C_1 C_2}$).

Ces trois relations permettent de déterminer les valeurs de k'_j , j'_{01} et j'_{02} , en fonction de l'intervalle de distance choisi. On obtient ainsi :

$$k'_j = \frac{\mathcal{D}_j}{\mathcal{K}_{right} - \mathcal{K}_{left} + \frac{B}{D_{min}}}$$

$$j'_{01} = -k'_j \cdot \mathcal{K}_{left}$$

$$j'_{02} = \mathcal{D}_j - k'_j \cdot \mathcal{K}_{right}$$

De plus, nous savons que :

$$d_{min} = k'_j \cdot B \left(\frac{1}{D_{max}} - \frac{1}{D_{min}} \right)$$

Finalement, la connaissance du repère A, G, T , des facteurs d'échelle, et des coordonnées des origines des images dans les caméras rectifiées permettra de déterminer les deux matrices rectifiées.

La construction des images rectifiées se fait alors par interpolation sur les quatre voisins les plus proches dans l'image originale (les matrices rectifiées fournissant la transformation d'un repère à l'autre).

III.2.2 Corrélation

Le principe de base de la corrélation est de rechercher le correspondant d'un pixel de l'image gauche dans l'image droite.

L'étape de rectification permet de limiter la recherche à la ligne correspondante et à l'intervalle de disparité choisi (soit finalement sur un certain nombre de colonnes uniquement).

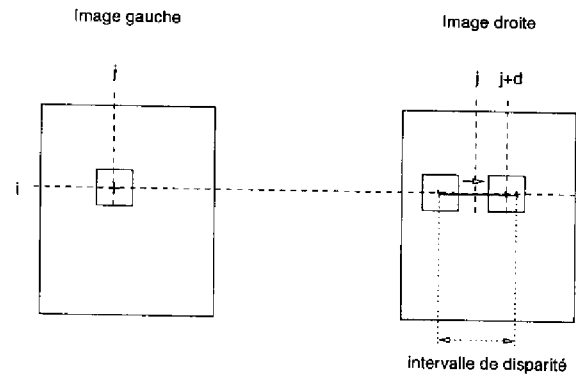


FIG. III.6 - Principe de la corrélation

Pour un pixel donné de l'image gauche, un critère sera calculé sur une fenêtre centrée autour du pixel. L'appariement entre ce pixel et l'un des pixels de l'image droite sera réalisé en comparant la ressemblance sur la fenêtre de corrélation (voir figure III.6).

Plusieurs critères peuvent être utilisés pour mesurer la ressemblance entre les deux fenêtres [Hotz 91]. Parmi ceux-ci, nous avons choisi d'utiliser le critère suivant (à maximiser) :

$$C = \frac{\sum_{i,j} I_1(i,j) \cdot I_2(i,j-d)}{\sqrt{\sum_{i,j} I_1^2(i,j) \cdot \sum_{i,j} I_2^2(i,j-d)}}$$

où $I_1(i,j)$ et $I_2(i,j)$ représentent les niveaux de gris des pixels respectivement pour les images gauche et droite (i et j décrivant les fenêtres de corrélation) et d la disparité (différence entre les indices colonne des deux images).

L'idée de base, proposée dans [Fua 93] et que nous avons reprise, consiste à introduire après cette phase une vérification supplémentaire en considérant que les deux images peuvent jouer un rôle symétrique (voir figure III.7).

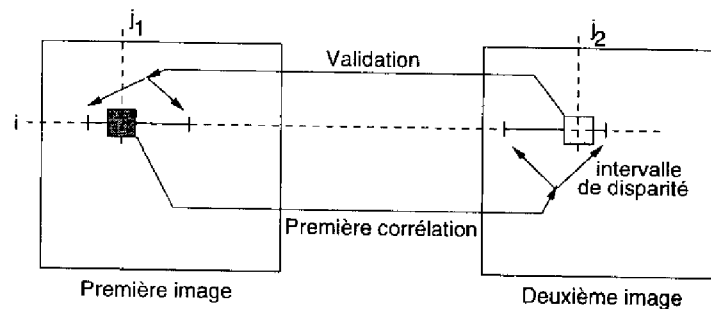


FIG. III.7 - Idée de base du critère de vérification

Dans un premier passage, le meilleur appariement possible d'un pixel (i, j_1) de l'image

gauche est détecté dans l'image droite (pixel (i, j_2)). L'idée est ensuite de vérifier la cohérence du résultat en recherchant à présent le meilleur appariement possible pour ce pixel (i, j_2) dans l'image gauche.

Ce critère de validité est particulièrement intéressant en comparaison avec ceux généralement utilisés (seuillage sur le critère de corrélation, forme du pic de corrélation...). Il se base uniquement sur la distance entre les deux appariements trouvés et a donc l'avantage d'être indépendant de l'intervalle de disparité choisi ou des caractéristiques de l'image (texture, contraste...). Il est en outre bien adapté pour détecter les occultations particulièrement sur les bords de l'image.

Ce test de validité est réalisé dans [Fua 93] en relançant deux fois le procédé de corrélation et en inversant le rôle des deux images. Cette manière de procéder fait que cette technique est très coûteuse en temps de calcul.

En réalité, nous avons montré qu'il est possible de réaliser les deux phases en une seule passe, ce qui permet de réduire de moitié le temps de calcul et de revenir à des temps du même ordre que lorsqu'un seul passage était effectué.

Pour chaque ligne de la première image, le critère de corrélation est calculé pour tous les pixels de la ligne et pour chaque disparité comprise dans l'intervalle par un algorithme classique de propagation (on pourra trouver le détail du principe dans [Bracco 92]).

Les valeurs du critère sont conservées dans un tableau dont les dimensions correspondent d'une part à la ligne traitée (l'axe des abscisses fournit les indices des colonnes) et d'autre part à l'intervalle de disparité comme le montre la figure III.8.

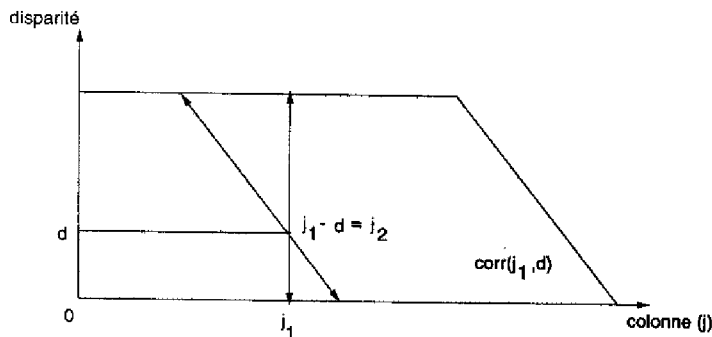


FIG. III.8 – Corrélation et vérification en une seule passe

Rechercher dans ce tableau le meilleur appariement d'un pixel donné (i, j_1) de la première image consiste à rechercher, selon la colonne j_1 la valeur maximum du critère. Supposons qu'il soit trouvé à une disparité d , soit dans la seconde image aux coordonnées $(i, j_2) = (i, j_1 - d)$.

Considérons à présent, la diagonale passant par (j_1, d) (voir figure III.8) et dont l'équation est $j_1 - d = j_2$. Sur cette droite, la valeur de j_2 reste la même tandis que j_1 et d varient. Finalement, scruter les valeurs du critère sur cette droite revient à rechercher, sur tout l'intervalle de disparité, pour le pixel (i, j_2) trouvé dans la seconde image, le meilleur

correspondant dans la première image.

Réaliser le test de validité se fera donc très simplement en vérifiant que le pixel trouvé sur cette diagonale se trouve bien sur la colonne j_1 à un seuil de distance près (généralement on accorde une erreur de 1 ou 2 pixels).

La dernière étape consiste ensuite à affiner les valeurs obtenues par approximation sub-pixellique de la disparité. Elle consiste à déterminer, par interpolation sur trois points autour de la disparité obtenue, la disparité optimale qui maximise le critère (on suppose pour cela que le critère suit une loi polynômiale du second degré).

III.2.3 Reconstruction 3D

La reconstruction 3D de l'image se fait à partir de l'image de disparité obtenue à l'étape précédente et des matrices de projection P_1 et P_2 des deux caméras virtuelles. Ces matrices (de dimension 3x4), déduites de l'équation III.1, s'expriment sous la forme :

$$P_1 = \begin{pmatrix} V & -C_1.V \\ H_1 & -C_1.H_1 \\ A & -C_1.A \end{pmatrix} \quad \text{et} \quad P_2 = \begin{pmatrix} V & -C_2.V \\ H_2 & -C_2.H_2 \\ A & -C_2.A \end{pmatrix}$$

$$\text{où} \quad V = k'_i.T + i'_0.A \quad \text{et} \quad H_{1,2} = k'_j.G + j'_{01,02}.A$$

Ces deux matrices ne diffèrent en fait que sur la seconde ligne. En effet, H_1 et H_2 prennent en compte les coordonnées horizontales des centres des deux images (j'_{01} et j'_{02}). Par ailleurs, on peut noter que $C_1.A = C_2.A$ et que $C_1.V = C_2.V$ puisque A et V sont orthogonaux à C_1C_2 .

La projection d'un point 3D P sur les deux images rectifiées sont respectivement les points

$$p_1 = \begin{pmatrix} i_1 \\ j_1 \end{pmatrix} \quad \text{et} \quad p_2 = \begin{pmatrix} i_2 \\ j_2 \end{pmatrix}.$$

$$\text{Ceci peut aussi s'exprimer par :} \quad \begin{pmatrix} i_{1,2}.w \\ j_{1,2}.w \\ w \end{pmatrix} = \begin{pmatrix} V & -C_{1,2}.V \\ H_{1,2} & -C_{1,2}.H_{1,2} \\ A & -C_{1,2}.A \end{pmatrix} \begin{pmatrix} P \\ 1 \end{pmatrix}$$

On obtient donc le système d'équations suivant :

$$(V - i_1.A).P = C_1.(V - i_1.A)$$

$$(H_1 - j_1.A).P = C_1.(H_2 - j_1.A)$$

$$(V - i_2.A).P = C_2.(V - i_2.A)$$

$$(H_2 - j_2.A).P = C_2.(H_2 - j_2.A)$$

où les première et troisième équations sont identiques puisque les images sont alignées ($i_1 = i_2$).

Au lieu de résoudre ce système pour chaque pixel, il est possible de se placer dans un repère où la solution sera simple, puis de transformer par la suite, le résultat dans le repère

de notre choix. Ce repère est en fait le repère rectifié (A, G, T) défini précédemment. Dans ce dernier, la solution correspond simplement aux équations de triangulation pour deux caméras parfaitement alignées. Ainsi, on aura :

$$\begin{aligned} X &= (P - C_1).T \\ Y &= -(P - C_1).G \\ Z &= (P - C_1).A \end{aligned} \quad \text{ou} \quad \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} T \\ -G \\ A \end{pmatrix} (P - C_1)$$

Finalement, on obtient la relation suivante entre le point 3D P et les coordonnées des pixels correspondants dans les deux images rectifiées :

$$Z = \frac{k'_j B}{d - (j'_{01} - j'_{02})}$$

$$X = (i_1 - i'_0).Z/k'_i$$

$$Y = (j'_{01} - j_1).Z/k'_j$$

avec $d = j_1 - j_2$, qui correspond à la disparité et,

$$P = \begin{pmatrix} T & -G & A \\ k'_i & -k'_j & \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + C_1$$

Comme on peut le remarquer, l'équation de triangulation est simple et ne demande pas d'inversion de matrices.

III.3 Résultats expérimentaux

Cet algorithme a été implémenté et testé à partir d'images acquises sur le système stéréoscopique du robot ADAM¹. Les deux images subissent initialement deux pré-traitements :

1. elles sont tout d'abord amenées à une résolution inférieure pour travailler avec des images de dimension 144x125. On obtient ainsi un compromis satisfaisant entre le nombre de pixels à corrélérer (temps de calcul) et la densité des points dans l'image de profondeur.
2. nous appliquons ensuite sur ces images un masque pseudo-laplacien obtenu par l'utilisation de fenêtres de tailles 7x7 et 3x3 (différence entre les deux fenêtres) afin d'augmenter les transitions présentes dans l'image.

Les améliorations apportées au cours de l'étape de rectification permettent d'augmenter le nombre de pixels corrélés dans l'image, la géométrie des images originales étant mieux conservée (voir figure III.9-c). Ces contraintes de similarité ne sont pas prises en compte

¹ La distance entre les deux caméras placées sur ADAM est de l'ordre de 0,6 mètre et leur angle de convergence est d'environ 10 degrés.

dans [Ayache 88] où l'altération des images dépend totalement du choix du repère de calibration.

Comme le montrent les images de disparité et de profondeur (figures III.9-d et e, III.10-1c et d, 2c et d), le nombre de pixels corrélés, généralement plus de 80% des pixels contenus dans la partie commune aux deux images, et par là, le nombre de points 3D est important ; ce qui nous permet d'avoir une carte dense. La plupart des pixels qui n'ont pas pu être corrélés sont éloignés des caméras.

D'autre part, l'étape de corrélation, regroupant en un seul passage appariement et vérification, a permis de diminuer significativement les temps de calcul (pratiquement de moitié par rapport à l'utilisation en deux passages du test de validation). Ils sont généralement de l'ordre de 4 secondes sur une station Sparc20. Naturellement, ces temps de calcul restent à améliorer. Nous donnerons, un peu plus loin dans cette section, quelques directions prises en ce sens.

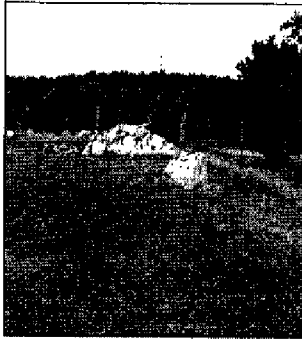
D'un autre côté, le choix du repère rectifié pour les calculs de triangulation permet de choisir par la suite le repère dans lequel visualiser la carte de profondeur : il suffira d'une simple multiplication par matrices de transformation. Dans notre cas, la carte de profondeur est présentée dans le repère de la caméra gauche.

Certaines erreurs d'appariement subsistent dans l'image, en particulier pour des pixels correspondant à des objets éloignés. On peut considérer que la distance critique au-delà de laquelle l'information est trop erronée est autour de 12 mètres (compte tenu de la méthode et des caractéristiques des caméras utilisées).

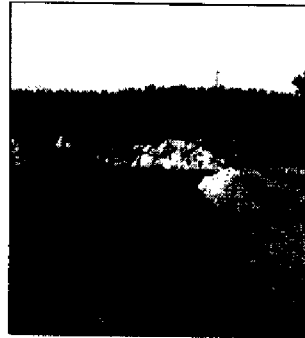
Nous fournissons à l'issue du procédé le pourcentage de pixels qui n'ont pas été corrélés suite à chaque critère d'appariement. Le nombre de paires stéréoscopiques à notre disposition ne permet cependant pas d'analyser ces résultats : ce travail d'analyse devrait être mené sur des images du même type pour en tirer une conclusion significative, ce qui n'était pas notre cas (images très variées comme le montrent les images des figures III.9 et III.10).

Une étude a néanmoins été réalisée par C. Parra [Parra 95] sur des images acquises au CNES. L'évaluation de l'algorithme portait sur la qualité des images de disparité en fonction du critère de corrélation choisi (deux critères ont été testés parmi ceux présentés dans [Hotz 91]), de la taille de la fenêtre de corrélation (3x3, 7x7 ou 11x11), et de l'intervalle de disparité - à pleine et demi-résolution (de 0 à 52 pour la pleine résolution ou le même intervalle mais en prenant une valeur de disparité sur deux pour la demi-résolution).

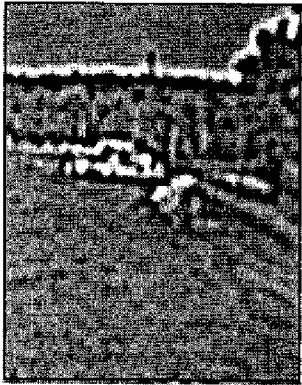
Il s'avère que le choix de notre critère est justifié car il permet d'avoir un bon compromis entre le nombre de pixels corrélés et le temps de calcul associé. L'utilisation de l'algorithme de propagation permet de rendre le calcul pratiquement indépendant de la taille de la fenêtre de corrélation. De même, l'étude montre que la demi-résolution sur la disparité doit être utilisée avec précaution : les temps de calcul sont plus faibles (de l'ordre de 5



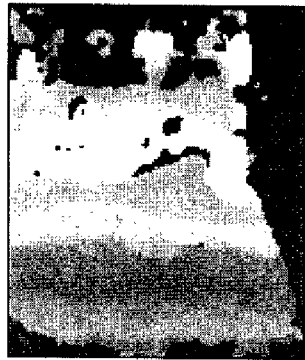
a- Image caméra gauche



b- Image caméra droite



c- Image rectifiée gauche
(après pré-traitements)



d- Carte de disparité



e- Image de profondeur

FIG. III.9 – *Un exemple complet sur une paire d'images stéréoscopiques*

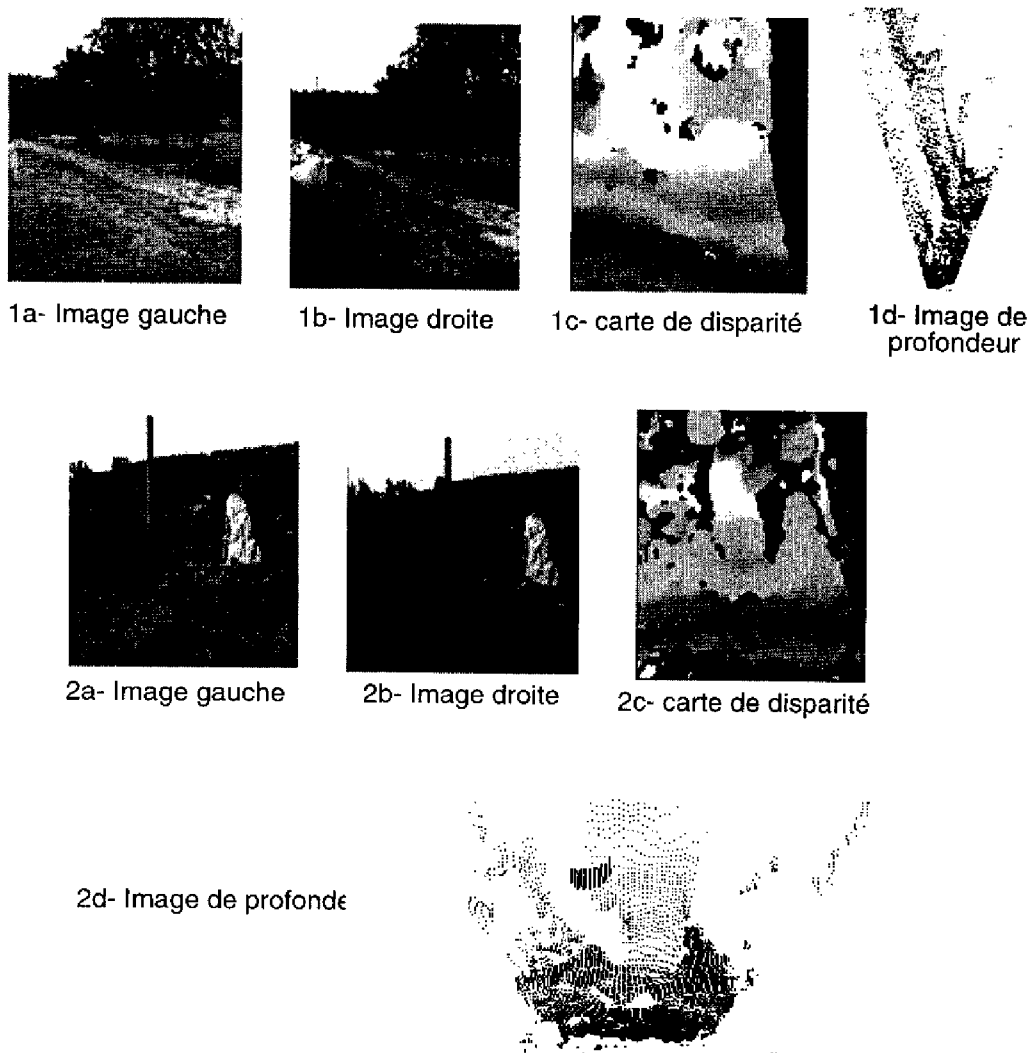


FIG. III.10 - *Résultat de la stéréocorrélation sur deux exemples différents*

secondes pour des images de 280x384) au détriment de la précision de la carte de disparité.

Pour améliorer la qualité de l'image de disparité, plusieurs filtres ont été testés dans le groupe. S. Lacroix [Lacroix 95] utilise un filtre médian pour éliminer les petites erreurs d'appariement. [Parra 95] a aussi étudié ce problème. Il utilise les propriétés de continuité de la disparité sur chaque colonne de l'image pour vérifier sa validité (modélisation de

l'évolution de la disparité par une fonction analytique [Eastman 87]).

Notons aussi que notre algorithme a été modifié pour pouvoir travailler sur une zone de l'image [Fouchet 94]. L'idée reste la même, il suffit simplement de limiter la zone de calcul au cadre englobant la région à corrélérer, puis de conserver, les calculs terminés, uniquement les valeurs pour les pixels contenus dans la région. Les temps de calcul pour la corrélation sont de l'ordre de 0,6 seconde pour une région de taille 50x50.

D'autres travaux, actuellement en cours, consistent maintenant à intégrer cet algorithme sur carte DSP afin de diminuer les temps de calcul. L'objectif est d'atteindre un temps de l'ordre de 2 secondes pour l'ensemble du procédé.

III.4 Conclusion et perspectives

Les deux problématiques actuelles en stéréocorrélation concernent la rapidité d'exécution des algorithmes et l'amélioration de la qualité de la carte de disparité obtenue. Nous avons répondu à ces deux objectifs de plusieurs manières :

- par réduction de l'altération des images rectifiées,
- par modification du traitement de l'appariement pour prendre en compte différents critères en une seule passe,
- par calcul de l'image de profondeur directement dans le repère de rectification.

A l'issue du procédé, un pourcentage de réussite sur chaque critère employé est fourni. Ce type d'information pourrait permettre de déterminer les raisons pour lesquelles l'appariement n'a pas donné de résultats sur certaines zones de l'image. On pourrait ainsi envisager d'appliquer d'autres traitements, plus spécifiques. Différentes approches pourraient être exploitées :

- utiliser la couleur sur une zone de l'image où le problème de similarité serait trop important,
- utiliser la coopération avec le laser sur une zone particulière de la scène lorsque la corrélation ne donne pas de résultats. Nous pensons en particulier à des problèmes liés à la luminosité comme des zones d'ombres dans l'image. Cela demande, par contre, de connaître précisément la position de la platine laser par rapport aux caméras ;
- appliquer la stéréovision avec des primitives de haut niveau : ceci peut être exploité à deux niveaux. Lorsque des zones structurées apparaissent dans la scène comme par exemple des bâtiments (association de segments et de régions) ou lorsque des zones sont trop éloignées (et donc homogènes) où l'on pourrait exploiter la ligne d'horizon (association de contours particuliers). Il faut pour cela avoir un modèle de la scène qui fournisse des régions qui soient bien caractérisées et qui correspondent aux principales composantes de la scène ; ce qui nous amène à la deuxième partie de ce manuscrit, le modèle nominatif de régions.

Chapitre IV

L'extracteur de régions

Construire un modèle nominatif des régions constituant la scène demande préalablement d'en extraire les principales composantes et de connaître les caractéristiques de chacune d'elles. Ces deux étapes sont importantes car elles pourront servir de base à différents traitements de plus haut niveau :

- l'interprétation de la scène,
- la détection d'amers,
- ou le suivi d'une région au cours du déplacement.

L'objectif est donc de construire un extracteur de régions afin d'obtenir les grandes régions de l'image, puis de caractériser chacune d'elles à l'aide d'un certain nombre d'attributs.

Dans ce chapitre, nous présenterons brièvement les différentes techniques de segmentation 3D déjà exploitées dans le groupe ainsi que celles utilisées pour segmenter une image vidéo. Ces premières ne permettent pas d'analyser la scène complète car les données utilisées sont tridimensionnelles et concernent uniquement la zone proche du robot. Nous présenterons ensuite la technique que nous avons développée pour segmenter l'image à partir des informations colorimétriques. Une dernière partie sera consacrée à la présentation des différents attributs que nous utilisons pour caractériser ces régions dans le but final d'identifier leur nature.

IV.1 Présentation générale

Comme nous l'avons vu précédemment dans le chapitre II (section II.1.1), l'extraction des régions et leur caractérisation sont actuellement obtenues par différents traitements en fonction du but poursuivi :

- Le robot doit se déplacer : [Lacroix 95] utilise la projection sur le sol d'une grille cartésienne pour déterminer les cellules initiales de la segmentation.

Ces régions sont par la suite caractérisées à partir d'attributs spécifiquement liés à l'information tridimensionnelle : la densité relative des points dans la cellule, la variance sur l'altitude des points, la variation maximale de l'altitude, le vecteur normal moyen, la variance sur ce vecteur.

Un dernier attribut indépendant des caractéristiques du terrain intervient dans la classification, la distance de la cellule au capteur, pour préciser la fiabilité des mesures. Il produit finalement une carte de régions indiquant les zones planes de celles qui sont accidentées ou qui sont des obstacles.

- Le robot doit se localiser : [Betge-Brezetz 96] segmente l'image 3D en parties du sol ou objets.

Le sol (ou les parties du sol) sont définis par «une grande surface d'orientation uniforme». Lorsque le sol est extrait, les objets correspondent alors finalement à tout ce qui intersecte le sol.

Ces objets sont ensuite caractérisés par leur forme à l'aide de super-quadriques et reliés entre eux dans la scène par des relations topologiques telles que *dessus*, *dessous*, *derrière* ...

Ce type de segmentation ne tient compte que de critères géométriques sur les objets ou le terrain. En outre, les images de segmentation 3D produites ne permettent pas d'analyser les zones lointaines du robot.

Segmenter les images issues du capteur caméra permettra de remédier à ces problèmes. Cependant, la segmentation que nous allons produire doit pouvoir s'intégrer avec les différents procédés existants dans la mesure où elle peut apporter des informations aussi bien pour la localisation que la navigation.

La figure IV.1 présente le schéma fonctionnel de l'extracteur tel que nous l'avons envisagé. Le schéma fait apparaître les fonctions déjà existantes de *segmentation 3D* et de *discrétisation 3D* développées respectivement par S. Betge et S. Lacroix et présente comment vont pouvoir s'intégrer les fonctions de segmentation d'images couleur et de caractérisation des régions que nous avons développées.

On peut remarquer que pour l'instant, la segmentation couleur serait uniquement exploitée pour la modélisation de l'environnement. En revanche, la caractérisation des régions

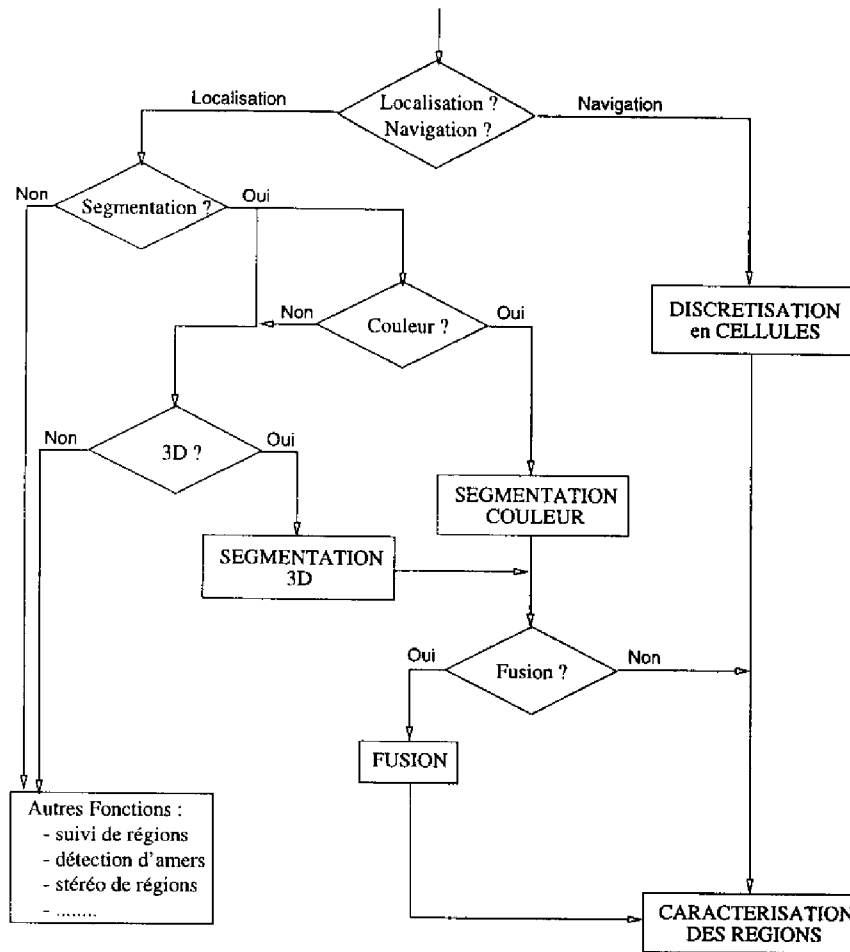


FIG. IV.1 – Schéma fonctionnel de l'extracteur de régions

peut être effectuée aussi bien sur les régions obtenues lors d'une tâche de navigation ou de localisation. Le fonctionnement de ces deux tâches se déroulera de la manière suivante :

- La segmentation couleur : à partir des images couleurs, les principales régions composant l'image sont extraites. Cette étape sera principalement utilisée dans la phase initiale de modélisation de l'environnement. Nous ne souhaitons pas en effet relancer tout le processus de segmentation chaque fois que le robot avance, mais simplement lorsque cela s'avère vraiment utile : si le robot avance en ligne droite, il peut être plus adéquat d'appliquer une technique de suivi plutôt que de re-segmenter. En revanche, si le robot effectue une rotation importante, ou s'il se déplace sur un terrain très abrupt, le point de vue risque d'être vraiment différent et le suivi moins

adapté. Une nouvelle phase de segmentation doit alors être envisagée.

La phase de segmentation peut être réalisée en plusieurs étapes : un premier passage permet de déterminer les plus grandes zones caractéristiques décrivant l'image. Les régions qui n'auront pas pu être correctement identifiées pourront, par la suite, être à nouveau segmentées.

Notons toutefois que dans le cas de la modélisation, une phase préalable de fusion devrait intervenir : les frontières issues des régions couleurs et des régions 3D ne se superposent pas toujours bien.

La coopération de ces deux types de segmentation nécessite la définition d'une stratégie pour extraire les limites de chaque région. Bien que nous n'ayons pas développé cette fonction, elle devra cependant être intégrée pour gérer l'ensemble des informations. Nous supposons pour l'instant que les régions seront uniquement issues du processus de segmentation couleur.

Il serait cependant possible dès à présent de considérer que les régions proches sont d'abord segmentées selon leur profondeur, puis selon leur couleur. Nous n'entrerons toutefois pas dans ce processus puisque notre objectif est de fournir des informations sur toute la scène (proche et éloignée).

- **La caractérisation :** chaque région de l'image est caractérisée par sa couleur et sa texture. Suivant que l'on est dans une phase de localisation ou de navigation, les régions proviendront respectivement de la segmentation couleur (et uniquement par elle pour l'instant) ou de la discrétisation de l'image 3D (les cellules fournies sont de tailles suffisantes pour être considérées comme des régions à caractériser). Ces régions et leurs caractéristiques serviront de base pour l'identification.

Nous allons donc présenter, à présent, les différentes techniques existant pour segmenter une image couleur ; nous verrons quelle approche nous avons choisie et les résultats que nous avons obtenus.

La deuxième partie sera consacrée à la caractérisation des régions issues de cette segmentation. Nous préciserons les attributs choisis pour caractériser les régions et montrerons leur intérêt pour la suite de ce travail.

IV.2 La segmentation d'images couleur

De manière générale, deux approches sont utilisées pour segmenter une image : l'approche contours et l'approche régions. Ces approches sont duales dans la mesure où un contour définit la frontière d'une région. Il faut toutefois noter qu'en pratique, compte tenu des algorithmes généralement mis en œuvre, les résultats obtenus par ces deux méthodes sont souvent différents pour des scènes relativement complexes.

L'extraction des régions à partir de contours pour des images d'extérieur comme celles que nous avons à traiter n'est pas envisageable : les images de gradient sont trop bruitées

pour que les contours significatifs puissent être correctement extraits. Nous nous sommes donc uniquement intéressés aux approches régions.

IV.2.1 Les méthodes existantes

Segmenter une image en régions consiste à obtenir une partition de l'image en zones répondant à des critères d'*homogénéité* et de *connexité*:

Soit X une image et $f(x, y)$ la fonction image associée: cette dernière peut aussi bien être scalaire (comme dans le cas des images noir et blanc) ou vectorielle (dans le cas des images couleur).

L'homogénéité se définit alors comme un prédicat $P(S)$ sur un sous-ensemble S de l'image X par la relation suivante:

$P(S)$ est vrai si S est homogène,

$P(S)$ est faux sinon.

Une segmentation de X correspondra donc à une partition de X en n sous-ensembles S_1, \dots, S_n tels que:

$$X = \bigcup_{i=1}^n S_i, \quad (\text{IV.1})$$

$$S_i \cap S_j = \emptyset \text{ pour tout } i \neq j, \quad (\text{IV.2})$$

$$P(S_i) = \text{vrai pour tout } i, \quad (\text{IV.3})$$

$$P(S_i \cup S_j) = \text{faux pour tout } i \neq j, S_i \text{ et } S_j \text{ étant adjacents dans } X. \quad (\text{IV.4})$$

Généralement, on distingue deux grandes classes de méthodes pour la segmentation en régions:

- Les méthodes globales: elle s'intéressent dans un premier temps uniquement à des critères d'uniformité, sans prendre en compte la connexité des pixels dans l'image. La formation des régions n'est donc obtenue qu'après analyse de la connexité (voisinage des pixels) dans l'image étiquetée.

On y retrouve essentiellement les méthodes de classification *supervisées* (approches statistiques) et *non supervisées* («clustering», méthodes de seuillage).

La classification se base sur une étude statistique de la répartition des pixels: chaque élément de l'image se distingue par un nuage dans l'espace des caractéristiques et le partitionnement de cet espace produit l'étiquetage recherché.

La classification sera dite *supervisée* lorsque les classes sont connues. Dans ce cas, une phase d'apprentissage permet de définir la partition des différentes classes dans l'espace des caractéristiques.

Dans le cas de méthodes *non supervisées* (ou automatiques), où les classes sont inconnues, la détermination de la partition se fait alors à l'aide d'une fonction de ressemblance (définie par le concepteur) dans l'espace des caractéristiques.

- Les méthodes locales (aussi appelées *méthodes ensemblistes*) : prennent en compte à la fois l'homogénéité et la connexité pour définir la partition. A l'issue du processus, on a donc directement la segmentation de l'image recherchée.

Elles s'organisent en trois grandes techniques :

- la séparation : elle part d'une partition initiale de l'image satisfaisant la relation IV.4 et produit une partition satisfaisant IV.3. L'image est donc initialement partitionnée en un petit nombre de régions. Ces régions seront subdivisées tant qu'elles ne satisferont pas au critère d'homogénéité ;
- la fusion : elle part inversement d'une partition satisfaisant IV.3 pour produire finalement une partition satisfaisant IV.4 ; dans ce cas, au contraire, l'image est partitionnée au départ en un grand nombre de petites régions qui seront fusionnées jusqu'à ce que le critère d'homogénéité ne soit plus satisfait ;
- la séparation-fusion : elle s'inspire des deux techniques précédentes ; la partition initiale est arbitraire et ne satisfait aucun des deux critères et la méthode produit finalement, après une phase de séparation puis de fusion, une partition satisfaisant les deux conditions. Cette méthode met généralement en œuvre deux critères d'homogénéité différents pour la séparation et la fusion (variance et moyenne par exemple).

Ces trois techniques mettent en œuvre un graphe d'adjacence pour permettre le regroupement des régions (prise en compte de la connexité) ; l'image est segmentée en régions de forme quelconque, chacune d'entre elles étant constituée de la réunion d'un certain nombre de régions élémentaires adjacentes.

Le lecteur intéressé pourra trouver dans [Pratt 91] et [Nalwa 93] des références sur chacune de ces méthodes.

En environnement d'extérieur, les principales méthodes utilisées sont les méthodes statistiques. Elles sont effectivement bien adaptées lorsque le but est d'identifier une zone spécifique de l'image. Ces méthodes de classification donnent de bons résultats dans le cadre du suivi de route, que la technique soit supervisée (comme dans [Baghdassarian 95] ou [Thorpe 88]) ou non supervisée [Lin 91], ou pour des applications de robotique agricole [Derras 94a, Derras 94b].

En revanche, lorsque le but est de fournir une description de la scène, les techniques non supervisées sont préférables (on ne sait pas *a priori* ce qui sera présent dans la scène). Les méthodes de classification, bien que très attractives, sont difficiles à mettre en œuvre. Ce genre de méthode demande en effet d'avoir des nuages distincts dans l'espace des caractéristiques ; ce qui, dans la réalité s'avère difficile dans le cas de scènes naturelles.

Les méthodes locales semblent dans ce cas plus intéressantes. Cet objectif de la segmentation reste cependant peu abordé: [Pujas 95] et [Panjwani 95] se sont intéressés à ces techniques et proposent de segmenter l'image selon la technique de séparation-fusion.

[Pujas 95] propose, avant de segmenter l'image, une classification préliminaire des pixels en fonction de leur degré de chromaticité. Cette classification permet de sélectionner les attributs pertinents pour la phase de segmentation.

Dans [Panjwani 95], la phase de séparation est suivie d'une classification des régions. Cette phase lui permet d'éliminer les très petites régions en regroupant les régions connexes ayant de fortes similarités. Une phase de fusion itérative est alors appliquée pour produire la segmentation finale. Les critères d'homogénéité utilisés au cours des différentes phases se basent sur un modèle markovien de la «texture couleur» (il entend par ce terme la prise en compte de l'organisation spatiale tridimensionnelle des pixels).

IV.2.2 Notre approche

Nous cherchons à obtenir une partition de l'image qui fournisse, en première approximation, une description synthétique de la scène perçue. L'utilisation de techniques non supervisées est donc préconisée qu'il s'agisse de méthodes globales ou locales. Nous nous sommes intéressés en particulier à la méthode locale de type «fusion» et à la méthode globale de type «clustering».

Position du problème

La méthode de fusion peut être décomposée en trois principales étapes :

1. partitionnement initial de l'image en un grand nombre de régions élémentaires généralement carrées,
2. fusion des régions élémentaires connexes satisfaisant au critère d'homogénéité pour obtenir les principales régions, et
3. élimination des petites régions restantes par fusion avec la région adjacente «la plus proche».

Un graphe d'adjacence permet de regrouper les régions en prenant en compte la connexité de celles-ci.

Les critères d'uniformité généralement utilisés sont détaillés dans [Orteu 89]. Cependant, la plupart de ces critères sont locaux ; ils introduisent une dépendance forte avec le graphe d'adjacence de telle sorte que le choix de la région élémentaire d'origine ou l'ordre de traitement des régions adjacentes influencent le résultat de la segmentation.

Les figures IV.3 et IV.5 montrent le résultat de la segmentation pour l'image entière (figure IV.2) et une partie de celle-ci (figure IV.4). Cette dernière a été obtenue pour les

points appariés lors de la stéréocorrélation (chapitre III). Les régions du chemin sont vraiment modifiées par la segmentation alors que les seuils de moyenne (pour le grossissement des régions) et d'élimination des petites régions sont identiques. Le choix du point de départ de la fusion et l'ordre de traitement des régions adjacentes modifient considérablement le résultat de la segmentation.

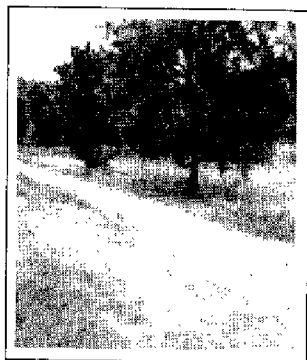


FIG. IV.2 – Une scène de EDEN

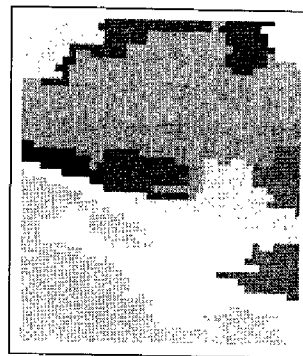


FIG. IV.3 – Image IV.2 segmentée



FIG. IV.4 – Partie de l'image IV.2

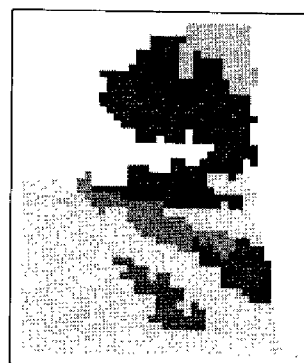


FIG. IV.5 – Image IV.4 segmentée

Ce type de résultat ne peut pourtant pas convenir. D'une part, les résultats sur toute l'image ne sont guère satisfaisants même s'ils sont cohérents. D'autre part, la comparaison avec la segmentation obtenue sur la partie corrélée indique d'autres limites. Lesquelles conserver dans ce cas et sur quels critères?

D'un autre côté, l'utilisation de méthodes du type «clustering» évite ce problème. En effet, ce type de méthode s'appuie sur le fait que les différentes régions sont caractérisées par un vecteur d'attributs. Cependant, dans le cas de scènes naturelles, les nuages de points dans l'espace des caractéristiques ne sont pas distincts (voir section IV.3.1). Cette condition est pourtant requise pour obtenir une bonne partition de l'image par ce type de

techniques.

Finalement, dans le contexte de scènes d'extérieur, ces deux méthodes ne peuvent nous donner entièrement satisfaction. En effet, d'une part nous souhaitons pouvoir appliquer la segmentation aussi bien sur l'image entière que sur une partie de celle-ci sans en modifier le résultat; ce qui ne peut pas être assuré par une méthode locale. D'autre part, la nature même des images traitées implique que nous ne puissions pas avoir dans l'espace de classification des nuages bien définis (zones où rochers et herbe sont mêlés, ...).

La méthode que nous avons développée conserve les avantages de chacune des deux méthodes en essayant de minimiser les inconvénients qui leur sont associés; elle s'inspire à la fois de la méthode de fusion (prise en compte de la connexité) et de la méthode du «clustering» afin de conserver un caractère global à la segmentation.

Présentation de la méthode

La structure générale de notre algorithme reprend celle classique des techniques de grossissement de régions. Cette approche s'inspire plus particulièrement des travaux menés par [Orteu 89] et [Coutance 91].

La taille des régions élémentaires carrées est déterminée en fonction de la taille des images et du niveau de finesse souhaité.

Un graphe d'adjacence (quatre adjacences possibles) est utilisé pour conserver la connexité des régions élémentaires à traiter ainsi que les caractéristiques colorimétriques de ces sous-régions. Ces caractéristiques sont les moyennes de chacune des composantes dans la région élémentaire :

$$m_{Re}(c) = \frac{1}{n_{Re}} \sum_{(i,j) \in Re} I(c, i, j)$$

où $m_{Re}(c)$ est la moyenne sur la composante c dans la région élémentaire Re (n_{Re} étant le nombre de pixels contenus dans Re) et $I(c, i, j)$ est la valeur de la composante c au pixel de coordonnées (i, j) dans Re .

Après une étude sur les différentes représentations possibles de la couleur, les composantes que nous avons utilisées sont directement les composantes Rouge, Verte et Bleue (R,V,B). Nous montrerons un peu plus loin dans ce chapitre (voir section IV.3.1) pourquoi nous avons fait ce choix.

Le critère d'uniformité utilisé pour la fusion des régions a été modifié : nous avons introduit une phase préliminaire d'affectation de chaque sous-région élémentaire à une classe. La génération automatique des classes s'inspire de celle présentée dans [Coutance 91] où l'espace de classification est partitionné en cellules cubiques. Dans notre cas, cependant, la taille de ces cellules n'est pas pré-définie : les limites des différentes cellules sont déterminées par l'étude de l'évolution de l'histogramme sur chacune des composantes colorimétriques. Le

nombre de seuils détectés déterminera le nombre de classes à créer ainsi que la taille des cellules (voir figure IV.6).

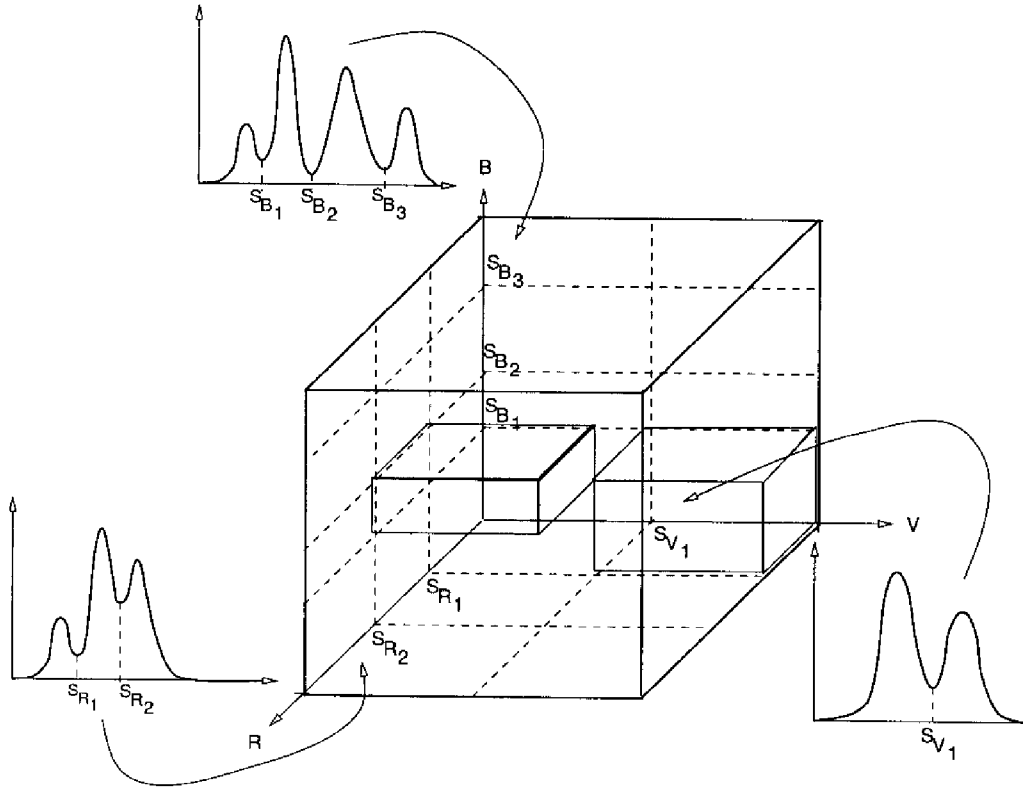


FIG. IV.6 – Génération des cellules

La partition de l'espace de classification obtenue est ainsi spécifique à l'image traitée; les différentes classes sont adaptées à l'image en cours de segmentation et non pas relatives à une référence fournie initialement.

Finalement, une région élémentaire sera affectée à la classe C si elle appartient à la cellule C définie par ses bornes (inférieure et supérieure) sur chacune des composantes colorimétriques. Le critère de fusion est alors :

Si deux régions élémentaires R_{e1} et R_{e2} sont adjacentes et appartiennent à la même classe C_i , alors R_{e1} et R_{e2} sont fusionnées.

L'intérêt est que l'on se ramène ainsi à un critère global où l'ordre d'adjacence des régions élémentaires et le point de départ de la fusion n'influencent plus le résultat de la segmentation.

La dernière étape, l'élimination des petites régions, est réalisée en utilisant le critère sui-

vant :

Soit R_e une région de taille inférieure au seuil fixé, R_1, \dots, R_n l'ensemble des régions adjacentes à R_e et $m_{R_e}, m_{V_e}, m_{B_e}$ et $m_{R_i}, m_{V_i}, m_{B_i}$ (i variant de 1 à n) les moyennes des composantes rouge, verte et bleue sur chacune de ces régions. R_e sera fusionnée à R_i si et seulement si :

$$\forall j \neq i, \quad \text{dist}(e, i) < \text{dist}(e, j)$$

dist étant la distance euclidienne entre les deux régions :

$$\text{dist}(e, k) = \sqrt{(m_{R_e} - m_{R_k})^2 + (m_{V_e} - m_{V_k})^2 + (m_{B_e} - m_{B_k})^2}$$

Nous venons de présenter rapidement comment se décompose la méthode. Nous allons maintenant préciser comment sont obtenues les classes des régions élémentaires et comment sont donc détectés les différents seuils de séparation.

IV.2.3 Génération des classes de régions élémentaires

La création des classes de régions élémentaires est basée sur l'étude des histogrammes sur chacune des composantes colorimétriques.

La détection des seuils sur chacune des composantes rouge, verte, et bleue a été envisagée de deux façons :

- les seuils sur chacune des composantes sont déterminés indépendamment les uns des autres. Le nombre de seuils fixés peut bien entendu être différent sur chacune d'elles ;
- le choix des seuils sur chacune des composantes dépend des seuils sur les autres composantes.

Le raisonnement est basé sur l'idée que les images que nous avons à segmenter ont *a priori* des variations chromatiques différentes sur chacune des composantes.

Dans le premier cas, on considère que chaque composante a le même poids. Dans le second cas, il est possible de considérer que, pour un même objet, les variations sur une composante sont plus pertinentes (car plus discriminantes) que sur l'une des deux autres composantes.

La détection finale des seuils passe par une phase préliminaire de *détection* des seuils possibles, et est suivie d'une phase de *sélection* des seuils pertinents. Cet objectif a été atteint par deux techniques différentes : l'une considère les caractéristiques locales des histogrammes cumulés, l'autre traite globalement le problème de la détection.

Après avoir présenté le principe associé à chacune d'elles et les résultats obtenus, nous finirons par une discussion sur leur utilité et dans quelle mesure il faudra choisir l'une ou l'autre.

Nous considérerons, pour simplifier la présentation de ces deux techniques, que nous travaillons dans un premier temps sur un seul histogramme ; nous préciserons, par la suite, les différentes possibilités testées lors de la transposition au cas tridimensionnel.

Méthode 1 : utilisation de critères locaux

Dans cette approche, nous avons préféré travailler sur l'histogramme cumulé.

Supposons que nous ayons une image d'intensité composée uniquement de trois objets. Soit $H(c)$ l'histogramme correspondant (figure IV.7).

Il est bien connu que le meilleur seuil pour séparer les deux classes (Classe 1 et Classe 2 sur notre schéma) est alors S_0 , correspondant au minimum local entre les deux classes (seuil optimal) [Duda 73]. En effet, S_0 correspond à la valeur où la probabilité d'erreur d'appartenance à l'une ou l'autre des classes est minimale [Postaire 87].

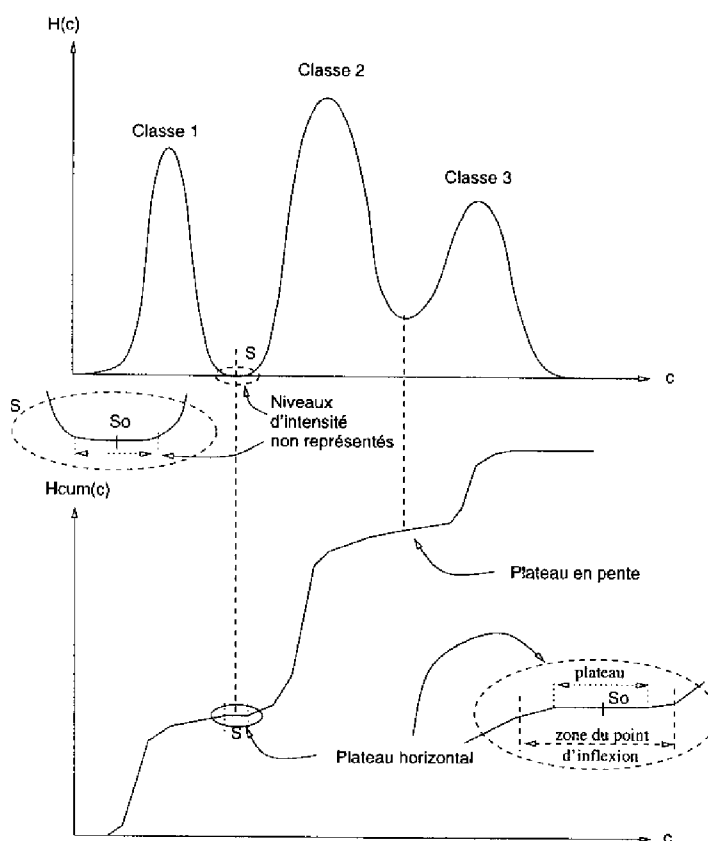


FIG. IV.7 - Histogramme $H(c)$ et $H_{cum}(c)$

Dans l'histogramme cumulé, ceci correspond à déterminer le point d'inflexion correspondant. Plusieurs informations peuvent être extraites de cet histogramme :

- la tangente au point d'inflexion : elle donne une indication qualitative du seuil ; plus la pente est faible, plus le seuil de séparation est sûr. Dans le cas général, une pente

nulle (comme dans le cas de S) indique en effet que certains niveaux d'intensité ne sont pas présents ;

- la largeur du plateau : précise le nombre de niveaux d'intensité qui stagnent à une valeur ; plus le plateau est large, plus les classes sont séparées ;
- la pente «de reprise» après le point d'inflexion : elle permet localement de déterminer l'importance de ce seuil ; une forte reprise après le point d'inflexion peut indiquer qu'une nouvelle classe apparaît.

Ceci nous a amenés à définir plusieurs critères :

1. *Détection préliminaire :*

- (a) Un seuil est détecté sitôt qu'un point d'inflexion apparaît c'est-à-dire dès que la pente diminue puis augmente (la méthode est détaillée dans [Lasserre 96]). Le nombre de seuils générés est plus ou moins important suivant le type d'images traitées.
- (b) Un seuil est détecté uniquement lorsque le plateau correspondant au changement de pente est supérieur ou égal à deux niveaux d'intensité. Ceci correspond à éliminer un grand nombre de seuils liés aux variations à l'intérieur d'une classe. Le nombre de seuils détectés est beaucoup plus limité.

2. *La phase de sélection :*

- (a) Affecte aux seuils détectés un coefficient de pondération sur la qualité du seuil. Ce coefficient est déterminé en calculant le rapport entre la tangente au point d'inflexion et la pente correspondant à la nouvelle progression (soit la prochaine classe). Seuils seront conservés les seuils dont le coefficient sera inférieur à un seuil fixé préalablement.
- (b) Dans le cas où les choix sur chaque composante sont interdépendants, le coefficient de pondération déterminé précédemment permet de classer par ordre d'importance les seuils sur chaque composante. Il s'agit alors de choisir les meilleurs seuils sur l'ensemble des composantes, soit en se fixant le nombre de seuils, soit en fixant le coefficient de pondération maximum.

L'ensemble de ces critères a été testé sur un grand nombre d'images plus ou moins complexes. Pour vérifier l'intérêt de chacun des critères employés nous nous sommes fixés une segmentation de référence. Celle-ci correspond uniquement à exploiter le critère (a) de la détection préliminaire, sans faire de phase de sélection par la suite.

La taille des régions élémentaires est de 3x3 pour les images IV.8 et IV.14 et de 2x2 pour les autres images présentées (IV.9 à IV.13 et IV.15 à IV.19).



FIG. IV.8 – Image 1 couleur



FIG. IV.9 – Image 2 couleur



FIG. IV.10 – Image 3 couleur



FIG. IV.11 – Image 4 couleur

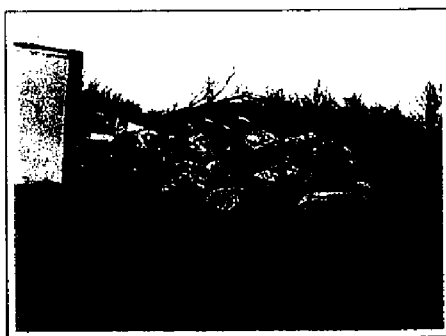


FIG. IV.12 – Image 5 couleur



FIG. IV.13 – Image 6 couleur

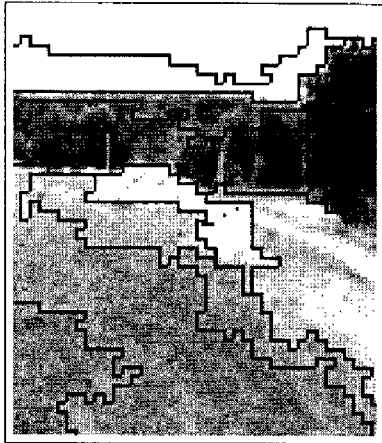


FIG. IV.14 – Image 1 segmentée



FIG. IV.15 – Image 2 segmentée

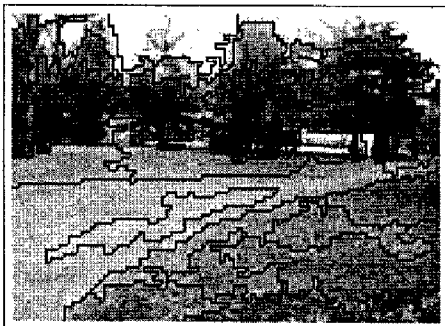


FIG. IV.16 – Image 3 segmentée

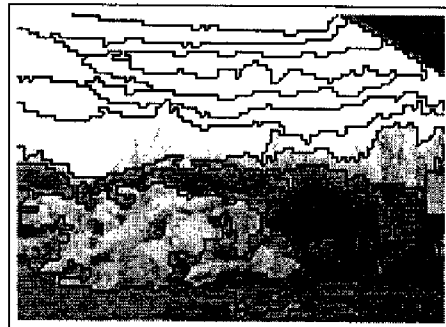


FIG. IV.17 – Image 4 segmentée



FIG. IV.18 – Image 5 segmentée

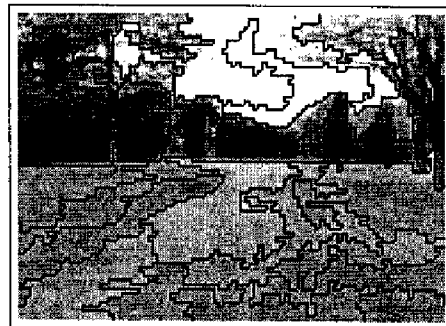


FIG. IV.19 – Image 6 segmentée

Les résultats obtenues lors de cette première segmentation font ressortir deux remarques importantes :

- Le nombre de régions à l'issue de la segmentation varie en fonction de la complexité des scènes perçues (voir figures IV.8 à IV.19).

La première image montre le cas d'une scène de EDEN (figures IV.8 et IV.14) ; le nombre de régions obtenu est faible (11 régions). Les autres images, au contraire, montrent des scènes plus complexes (éclairage varié, beaucoup d'éléments composent la scène) : le nombre de régions obtenu est beaucoup plus important. Les frontières obtenues restent cependant satisfaisantes dans la mesure où, globalement, les séparations entre les différents objets composant la scène sont détectées.

- Le nombre de seuils à conserver pour la segmentation doit rester assez important lorsque l'image est complexe. Le nombre de classes créées reste important et certaines régions très complexes (comme les arbres) sont donc essentiellement obtenues lors de l'étape d'élimination des petites régions ;

Nous avons ensuite comparé ces résultats à ceux issus d'une nouvelle segmentation prenant l'un des autres critères en compte. Il s'avère que les différents critères proposés ne permettent pas de dégager une amélioration significative pour l'ensemble des images testées :

1. Le critère (b) de la phase préliminaire de détection permet de ne pas tenir compte des variations à l'intérieur d'une même classe. Ce critère est intéressant dans le cas d'images très complexes (en particulier lorsque les variations de luminance sont importantes) mais devient trop sélectif dans le cas d'images plus homogènes (pas assez de seuils).
2. L'introduction de la phase de sélection (que ce soit dans le cas (a) ou (b)) associée au critère (a) de détection des seuils demande de déterminer une valeur critique du coefficient au-delà de laquelle on considère que les seuils détectés ne sont plus intéressants. Cette valeur critique diffère considérablement en fonction des images. Ce critère est donc relativement peu intéressant dans la mesure où il contraint l'utilisateur à intervenir dans l'étape de segmentation.

Finalement, tous ces critères de poids sont trop localisés au point détecté et ne sont pas liés à l'histogramme complet. Ils ne permettent pas, comme c'était le but, d'améliorer le résultat de la segmentation et de s'abstraire de l'utilisateur. C'est pourquoi nous avons

uniquement conservé la technique de base.

Les résultats de cette première segmentation (prenant en compte uniquement le critère (a) de la phase de détection) sont cependant intéressants car :

- elle est robuste malgré la variété des images traitées : elle assure la séparation correcte des différents éléments constituant la scène ;
- la technique est entièrement automatique : elle ne nécessite que de définir le seuil de taille pour l'élimination des petites régions (inférieur à 5% de la taille de l'image dans les cas traités). Ce seuil est initialement fixé et ne varie plus ; Notons toutefois que, par la suite, par coopération avec la segmentation 3D, il serait possible de déterminer ce seuil de taille en connaissant la taille des objets perçus.
- le nombre de classes change automatiquement en fonction de l'image : si l'on traite uniquement une partie de l'image, le nombre de classes est réduit et la segmentation produite fournit bien les frontières entre objets.

Cette technique est en fait mieux adaptée pour la caractérisation des régions car elle permet d'extraire de l'image les régions caractéristiques d'un type d'éléments (moins de dispersion dans l'homogénéité de la région). Nous la conserverons donc pour caractériser les différentes classes d'éléments contenus dans la scène.

En revanche, lors de l'identification, le but est d'extraire uniquement les grandes régions de l'image, quitte à fusionner les régions de peu d'intérêt. Dans ce cas, l'utilisation de critères locaux n'est plus suffisante lorsque les images deviennent complexes. Nous avons donc développé une autre technique, toujours basée sur l'étude de l'histogramme, mais qui exploite des critères se référant à l'ensemble de l'histogramme.

Méthode 2 : utilisation de critères globaux

Les critères utilisés pour détecter les seuils sur l'histogramme cumulé étaient uniquement des critères basés sur une étude locale autour du seuil à détecter. Or, dans le cas de scènes complexes, le nombre de seuils détectés est trop important sur chaque composante, pour que seuls ceux nécessaires à la segmentation en grandes régions soient pris en compte.

Pour prendre en compte un critère global et non plus local, nous devons, non seulement repérer les minima mais aussi les maxima de l'histogramme. Les minima indiquaient les seuils de séparations possibles, les maxima indiqueront leur poids.

Prenons l'exemple de la figure IV.20 : l'histogramme $H(c)$ est très bruité mais intuitivement, on voit que deux classes principales se détachent de l'ensemble.

Supposons que l'on partitionne cet histogramme et que l'on conserve, sur chaque partie, uniquement les minima et maxima les plus significatifs. De ce fait, nous obtenons un nouvel

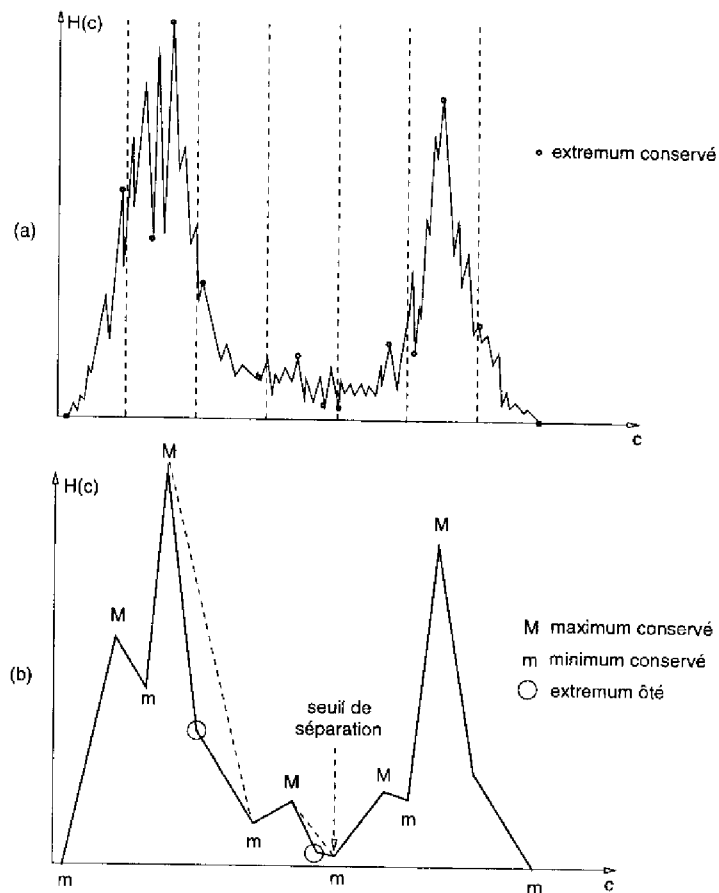


FIG. IV.20 - Génération des seuils de séparation des classes

histogramme, plus grossier, et qui fait ressortir l'allure générale de la courbe (représentée sur la figure IV.20-b). Il est alors facile d'en déduire où devrait se produire la séparation.

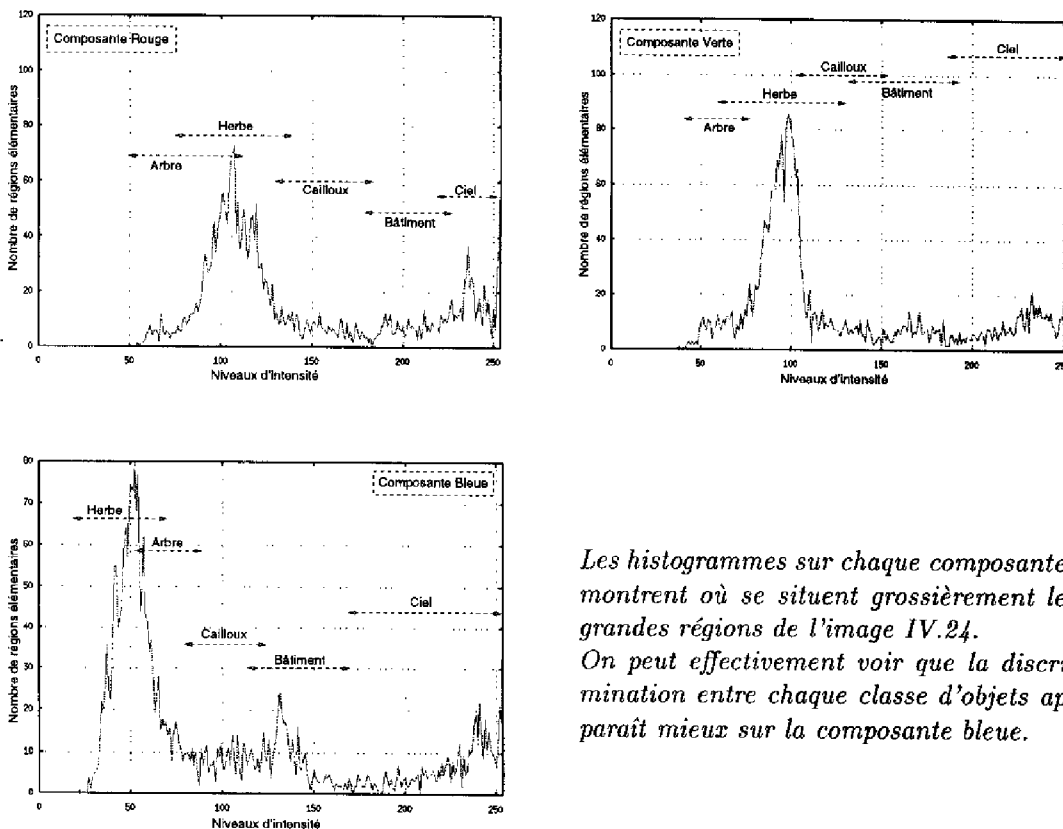
L'obtention de cette allure générale suit l'idée présentée: sur chaque intervalle correspondant à la partition choisie, les minima et maxima «globaux» sont retenus. Lorsque deux extrema de même type (respectivement minimum ou maximum) se suivent, seul est conservé le plus significatif (respectivement le plus faible ou le plus élevé). La courbe obtenue après avoir ôté ces extrema est représentée en pointillés sur la figure IV.20-b.

Le critère de séparation devient alors :

- le minimum doit être compris entre deux maxima,
- l'ordre des valeurs des maxima définit l'ordre de priorité pour séparer les classes. Ainsi, fixer le nombre de maxima d'intérêt, fixe le nombre de classes souhaitées.

Le partitionnement choisi fixe donc la finesse de la segmentation puisqu'il définit le nombre de maxima à détecter c'est-à-dire le nombre de seuils possibles. Un partitionnement assez grand permet donc de prendre uniquement en compte les plus grandes variations dans l'image.

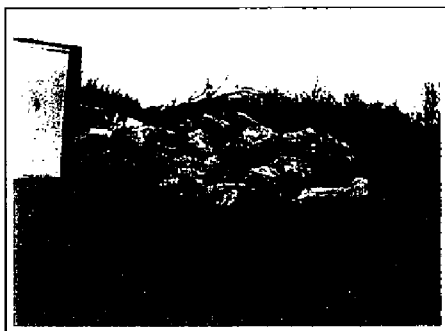
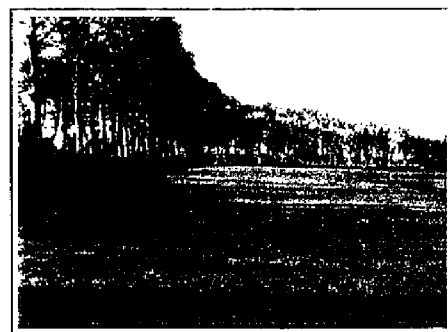
Nous avons appliqué cette technique à l'ensemble des images dont nous disposions. Le partitionnement a été fixé à 16 niveaux d'intensité, ce qui correspond à rechercher dans un premier temps 16 maxima et 16 minima. Le nombre de seuils détectés sur chacune des composantes est expérimentalement pour les scènes traitées de l'ordre d'une dizaine. Après plusieurs essais, il s'avère que pour les images analysées, le nombre de seuils peut être fixé respectivement à 3, 4 et 5 pour les composantes rouge, verte et bleue. Ceci est dû au fait que la composante rouge est très peu discriminante pour les scènes que nous avons à traiter (voir figure IV.21).



Les histogrammes sur chaque composantes montrent où se situent grossièrement les grandes régions de l'image IV.24. On peut effectivement voir que la discrimination entre chaque classe d'objets apparaît mieux sur la composante bleue.

FIG. IV.21 – Histogrammes sur les trois plans couleur de l'image IV.24

Les résultats obtenus correspondent bien à ce qui est attendu même dans le cas d'images complexes. Le nombre de régions obtenues a été réduit de moitié sur les scènes les plus

FIG. IV.22 – *Image 1 couleur*FIG. IV.23 – *Image 2 couleur*FIG. IV.24 – *Image 5 couleur*FIG. IV.25 – *Image 7 couleur*FIG. IV.26 – *Image 8 couleur*FIG. IV.27 – *Image 9 couleur*

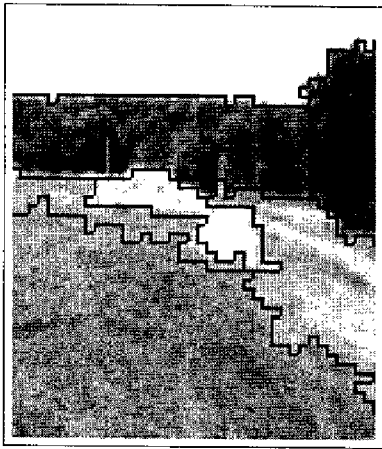


FIG. IV.28 – Image 1 segmentée



FIG. IV.29 – Image 2 segmentée

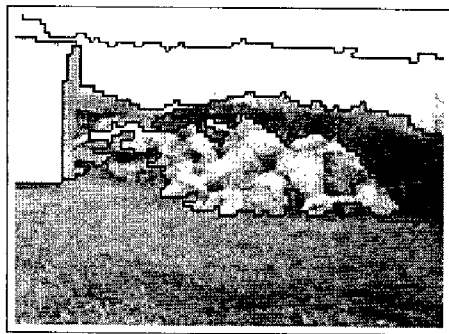


FIG. IV.30 – Image 5 segmentée



FIG. IV.31 – Image 7 segmentée

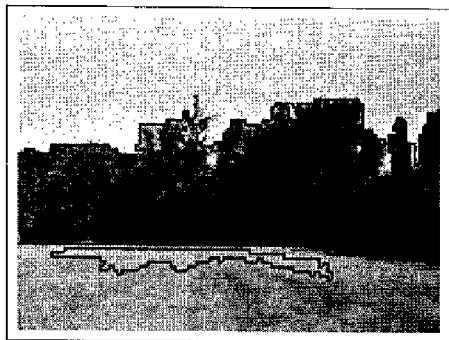


FIG. IV.32 – Image 8 segmentée



FIG. IV.33 – Image 9 segmentée

complexes (passage de 27 à 13 régions maximum).

Les figures IV.22 à IV.33 montrent le résultat de la segmentation pour six types d'images très différentes.

Les trois premières figures IV.28 à IV.30 (résultats des images des figures IV.22 à IV.24) sont à comparer avec le résultat issu de la première segmentation (méthode 1). La différence est importante : on a retrouvé les grandes régions de la scène contrairement au cas de la méthode 1. Ces figures ne présentent que des scènes dans lesquelles les éléments sont à des distances de l'ordre de 15 à 20 mètres au maximum du capteur. Le nombre de régions est faible dans chaque image (de 7 à 12 régions pour les quatre images présentées au lieu de 11 à 22 régions précédemment) et l'essentiel de la scène est retrouvé.

Les figures IV.31 à IV.33 montrent le cas de la segmentation sur des images où beaucoup d'arbres sont présents (figures IV.25 à IV.27). Ceux-ci sont bien regroupés (2 régions produites uniquement pour les images IV.31 et IV.33) et le nombre total de régions est faible (respectivement 5 et 6 régions au total pour ces deux mêmes images). Enfin, l'herbe est bien séparée de la zone terreuse sur l'image IV.32.

Les figures suivantes (IV.25 et IV.31) montrent la segmentation sur une image où la distance peut varier de manière plus importante et où des zones d'ombres apparaissent. On peut remarquer que la segmentation est très efficace : les ombres les moins marquées restent fusionnées à l'herbe ; les arbres sont bien regroupés et le nombre total de régions produites est faible (8 régions).

IV.2.4 Conclusion

Les deux techniques de segmentation que nous avons présentées se basent sur l'étude des histogrammes sur chacune des composantes RVB. Les résultats obtenus par ces deux méthodes sont très différents mais intéressants.

La première méthode (méthode 1), basée sur une étude locale, prend en compte de faibles variations dans l'image. Elle permet d'avoir des régions très homogènes, spécificité utile lorsqu'on souhaite caractériser un type d'objet.

La deuxième (méthode 2), plus globale, conserve uniquement les grandes composantes de l'image. Elle sera particulièrement intéressante pour générer le modèle de la scène où il n'est pas toujours nécessaire de connaître avec précision tout ce qui constitue l'image.

A la fin du procédé de segmentation nous obtenons donc une partition de l'image en régions. L'objectif étant d'identifier la nature des différentes régions, l'étape suivante consistera donc à caractériser ces régions à l'aide d'un vecteur d'attributs. C'est l'objet de la section suivante.

IV.3 La caractérisation des régions

A l'issue du procédé de segmentation, nous avons un certain nombre de régions dont nous voudrions définir les caractéristiques. Ces dernières doivent permettre de différencier, par la suite, chaque classe d'éléments qui constituent l'environnement, à savoir, l'herbe, les roches, la terre, les arbres et le ciel. L'objectif ici sera donc de déterminer quels attributs seront discriminants pour distinguer les différentes classes et préparer ainsi la phase d'apprentissage pour l'identification.

Le choix des attributs est donc important pour la caractérisation des régions. L'ensemble des attributs doit être déterminé en fonction de la possibilité qu'ils offrent de séparer les différentes classes définies.

Nous avons étudié plusieurs possibilités comme en particulier l'utilisation de l'image de luminance pour segmenter les images. Il s'avère que cet attribut est insuffisant pour la segmentation (voir annexe A) : les variations de luminance sur les arbres et sur l'herbe peuvent être trop similaires pour être détectées. Il paraît donc naturel d'utiliser des attributs colorimétriques et de texture lorsqu'on travaille dans un environnement tel que celui présenté. Dans ce cadre, l'utilisation de la luminance pour caractériser la texture d'un objet aura son intérêt.

Trouver des attributs invariants paraît impossible mais la combinaison couleur-texture, dont la variance dépendrait de paramètres différents est envisageable.

Nous allons, à présent, présenter les principales représentations utilisées pour la couleur et montrerons comment caractériser la texture d'un objet. Nous justifierons par la suite les choix faits pour la couleur et la texture à partir d'une étude menée sur un grand nombre de *régions-échantillons* issues de la segmentation couleur (méthode 1 présentée en IV.2.3). Des régions de chaque classe «herbe», «terre», «ciel», «arbre» et «roche» ont été prélevées (221 *régions-échantillons* au total sur plus d'une trentaine d'images). Le critère était que ces régions ne devaient contenir qu'une seule classe.

IV.3.1 Les attributs colorimétriques

Il existe dans la littérature une grande variété de représentations de la couleur pour un point de l'image. Généralement, cette dernière est représentée par trois valeurs correspondant à l'intensité lumineuse rouge (R), verte (V) et bleue (B). Ces trois primaires contiennent à la fois les informations de luminance et de chromaticité. La normalisation de ces trois primaires permet de s'abstraire des effets de luminance. On obtient alors (r), (v) et (b) par :

$$r = \frac{R}{R + V + B} \quad v = \frac{V}{R + V + B} \quad b = \frac{B}{R + V + B}$$

D'autres transformations linéaires et non linéaires se rapprochent plus de la vision humaine en particulier le système ITS (Intensité, Teinte, Saturation) dont les équations de

transformation sont [Cattoen 85]:

$$I = \frac{R + V + B}{3} \quad T = \arctan \frac{\sqrt{3} * (V - B)}{(2 * R - V - B)} \quad S = 1 - 3 * \frac{\min(R, V, B)}{R + V + B}$$

où $I \in [0, 255]$, $T \in [0, 2\pi]$ et $S \in [0, 1]$.

Par un changement d'échelle, il est facile de ramener les intervalles à des variations entre 0 et 255. Notons que la teinte correspondant à un angle, l'origine du repère est fixé sur l'une des teintes. Dans notre cas, le rouge est choisi comme origine.

D'autres systèmes de représentation sont présentés de manière détaillées dans [Ohta 85] tels que XYZ , $I_1 I_2 I_3 \dots$

Certaines de ces représentations (XYZ -norme CIE-, XYZ -norme UER-, UVW -norme UER-) ont été testées et n'ont pas apporté de meilleurs résultats; nous n'irons donc pas plus avant dans leur description.

Afin de valider les résultats de la segmentation, nous avons projeté dans les trois systèmes proposés (RVB, rvb et ITS) la valeur colorimétrique moyenne de chaque type de *régions-échantillons*. L'organisation des différentes classes pour ces trois représentations est présentée dans les figures IV.34, IV.35 et IV.36. Quel que soit l'espace de représentation utilisé, la classe «arbre» est trop dispersée et des classes «roche» et «terre» s'intersectent. Il est donc *a priori* difficile de fixer son choix sur l'une de ces représentations.

Bien que les classes soient mieux délimitées sur rvb , ce système s'est avéré peu robuste dans notre cas. La variation sur la composante v est parfois très faible: trop faible même pour pouvoir obtenir une segmentation correcte. La figure IV.37 montre la variation sur cette composante (15 niveaux d'intensité pour 4 classes); toutes les classes (ou presque) sont mélangées.

Les deux systèmes de représentation adéquats pour toutes les images dont nous disposons sont finalement les systèmes RVB et ITS.

Il est difficile d'évaluer les résultats issus de la segmentation; cette évaluation est visuelle et par là même subjective. Nous avons choisi le système RVB car la description obtenue (surtout au niveau des arbres) correspondait plus à l'idée que l'on pouvait se faire d'un tel procédé de segmentation.

La caractérisation des régions cependant peut être faite à partir de l'une ou l'autre (RVB ou ITS).

L'intérêt du système ITS dans notre cas réside dans la possibilité qu'il offre d'imposer la séparation entre deux classes. En effet, le choix de l'origine du repère teinte pourrait devenir un avantage, au lieu d'être un inconvénient, lorsque celui-ci est placé entre deux

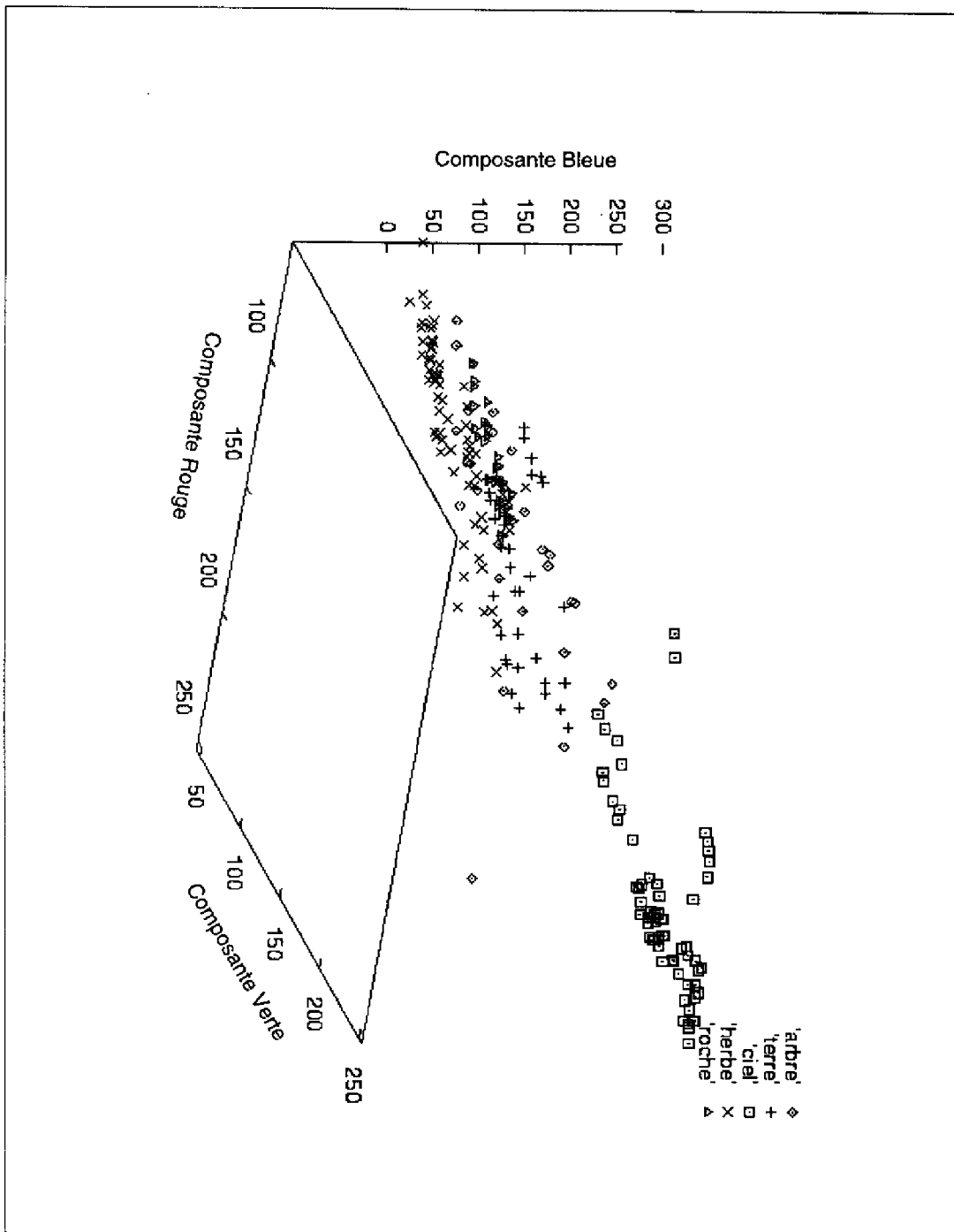
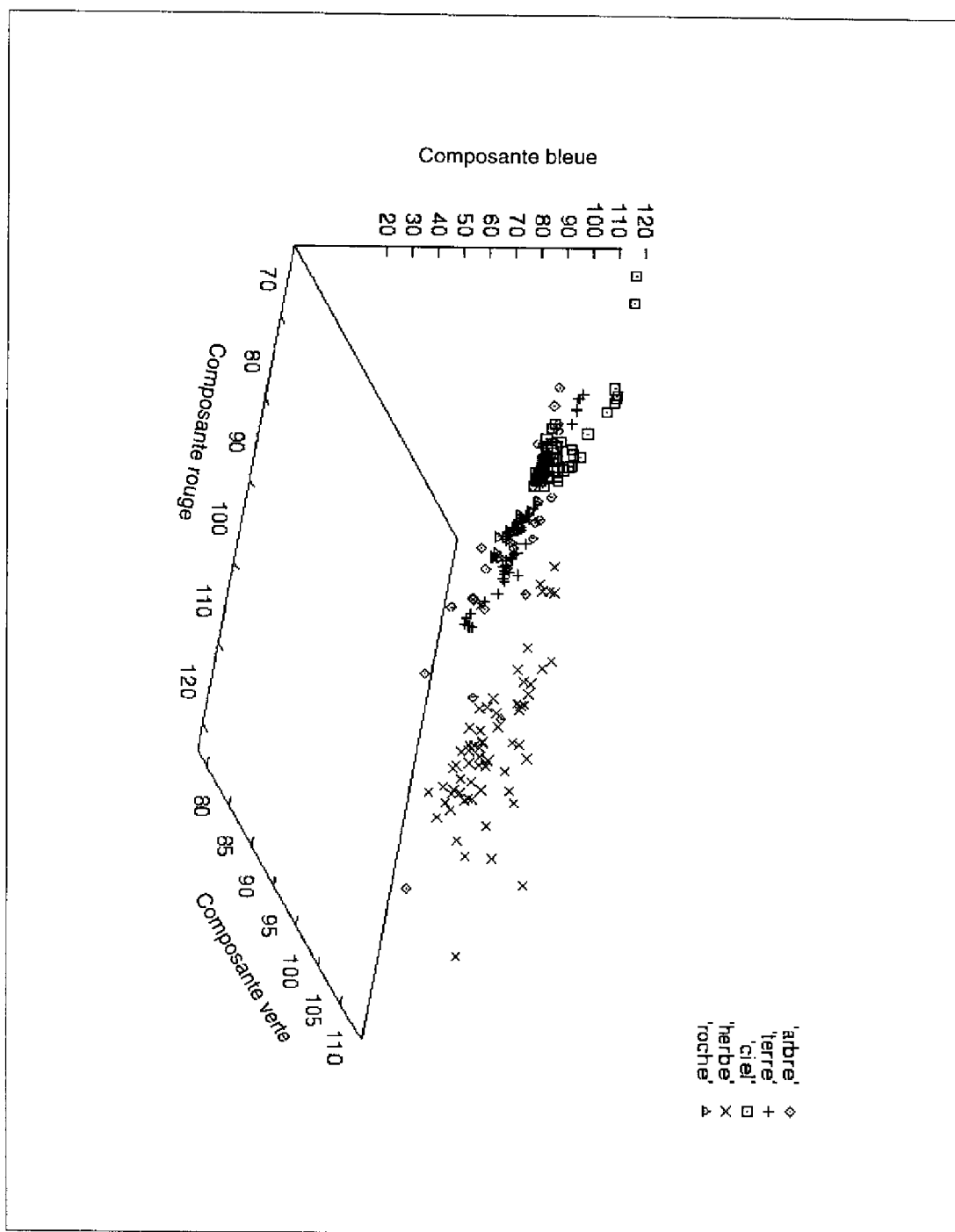


FIG. IV.34 – Représentation des classes dans RVB

FIG. IV.35 – Représentation des classes dans *rvb*

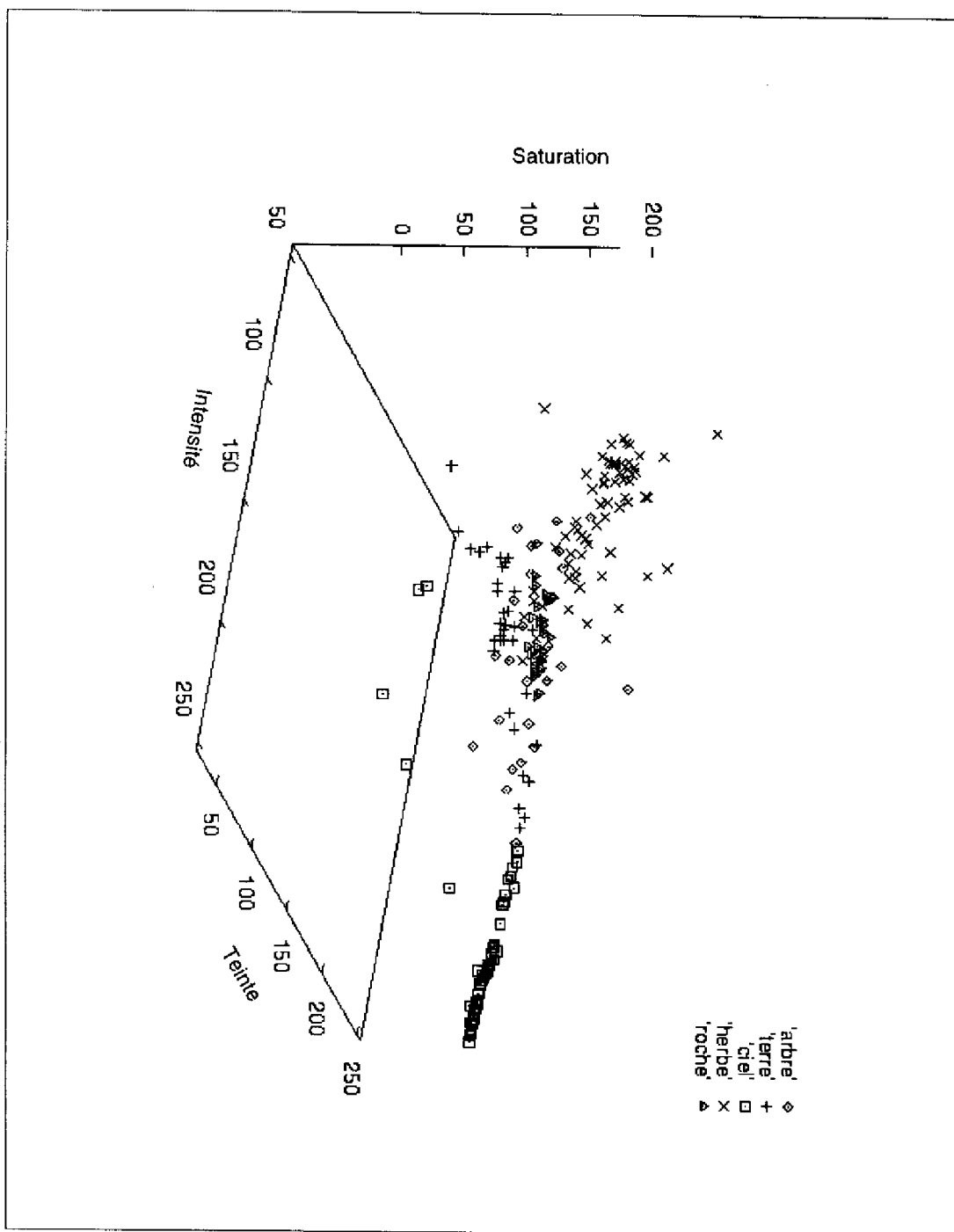
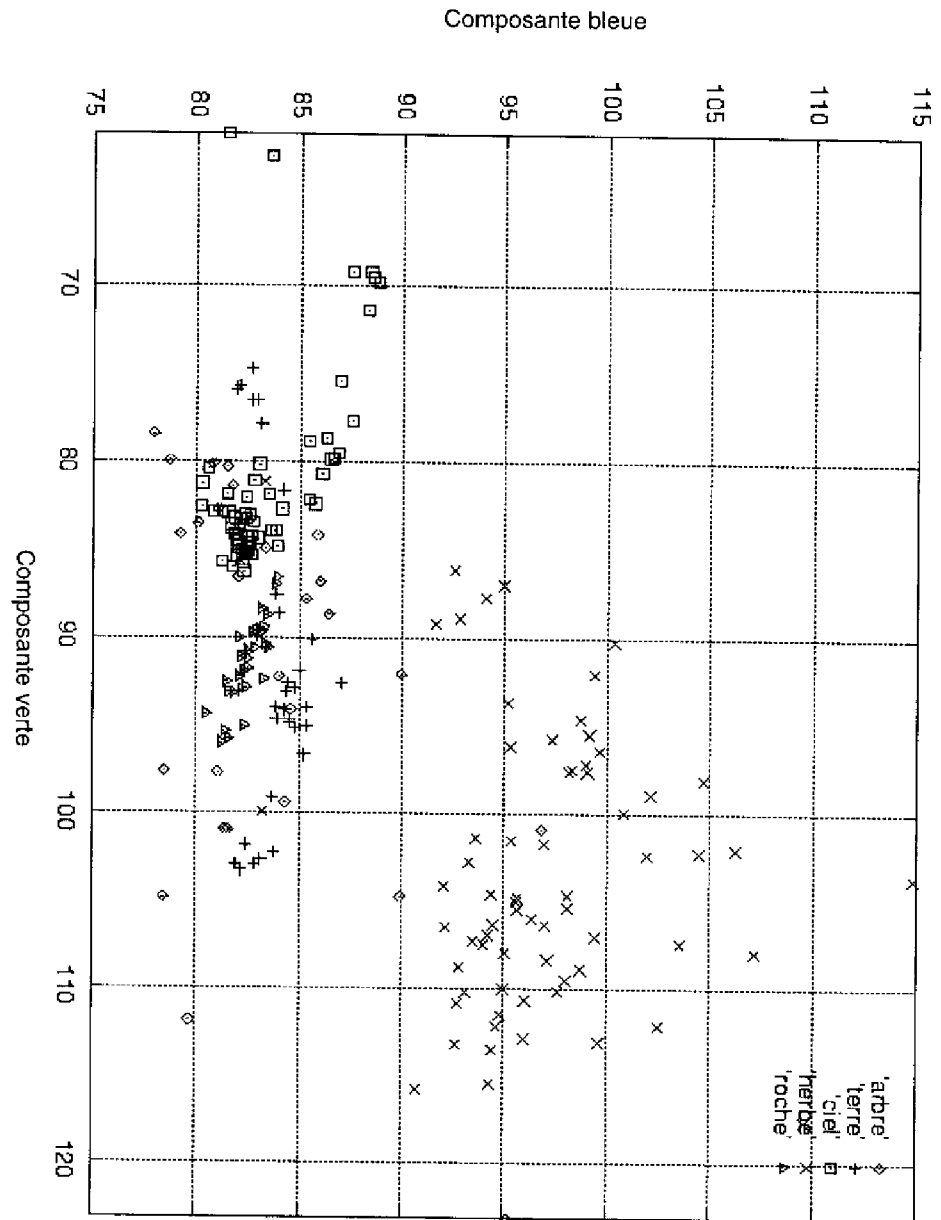


FIG. IV.36 – Représentation des classes dans ITS

FIG. IV.37 - Représentation des classes dans rv

classes : la figure IV.38 montre la projection de trois classes C_1 , C_2 , C_3 dans le repère teinte.

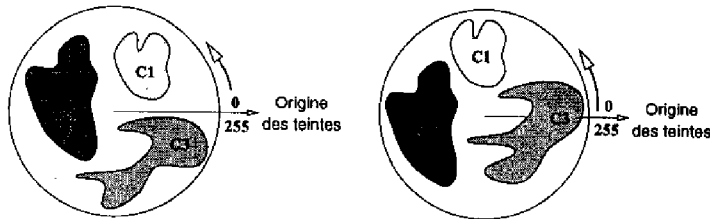


FIG. IV.38 – Représentation de trois classes dans le repère teinte : sur la figure de gauche, la référence est bien choisie tandis que dans le cas de la figure de droite, une même classe va voir ses valeurs de teinte varier autour de 0 et 255 (problème de discontinuité).

Dans la figure de droite, la discontinuité correspondant au passage de 0 à 255 (ou de 0 à 2π) se retrouve dans les valeurs de teintes associées à la classe C_3 . Sur la figure de gauche en revanche, les valeurs de teinte sur chaque classe varient continûment et la discontinuité introduit une séparation entre les classes C_1 et C_2 .

Ceci pourrait être particulièrement intéressant lorsque l'identification de certaines classes risque d'être difficile (comme dans le cas de la classe arbre) : la discontinuité pourrait permettre d'éviter des erreurs d'interprétation. La validité d'une telle approche nécessiterait tout de même une étude plus approfondie.

Mais le système ITS a un inconvénient majeur : les équations de transformation sont non-linéaires et, au voisinage des points achromes et au voisinage du noir, des singularités apparaissent : de faibles variations des valeurs de R, V, B au voisinage de ces points peuvent entraîner de fortes variations de la teinte ou de la saturation.

Nous avons donc choisi de conserver la représentation utilisée pour la segmentation, à savoir RVB, pour la caractérisation des régions.

IV.3.2 Les attributs de texture

La texture est généralement définie comme *traduisant l'organisation spatiale plus ou moins homogène d'éléments simples* [Postaire 87]. Elle fournit des informations qualitatives telles que la finesse ou la grossièreté, la granularité, etc. Il existe de nombreuses méthodes de caractérisation de la texture. Elles sont décomposées en deux grandes approches :

- les approches structurales qui s'intéressent à l'organisation spatiale d'éléments de texture formant l'objet. La texture de l'objet peut être définie à partir de règles. Ce sont des approches complexes et précises de mesure de la texture ;
- les méthodes statistiques : où il s'agit plutôt de déterminer les caractéristiques locales parallèlement en chaque point de l'image. On retrouve dans ce type de techniques

l'utilisation de statistiques de niveau plus ou moins élevé (moyenne, variance, matrices de co-occurrence, transformée de Fourier ...). D'autres techniques utilisent aussi des caractéristiques locales telles que les contours, les pics ou les vallées, ...

Le lecteur pourra trouver dans [Tomita 90] de plus amples informations concernant l'ensemble de ces méthodes.

L'étude de la texture est largement abordée dans la littérature: certains s'intéressent à la segmentation proprement dite [Lee 92], [Chaudhuri 95], [Dunn 94], d'autres cherchent à caractériser l'état d'une surface (recherche de défauts [Delagnes 96, Thomas 94])... mais elle reste encore généralement peu utilisée en robotique mobile d'extérieur.

[Derras 94a, Derras 94b] utilisent des attributs de luminance et de texture pour séparer dans l'image les zones d'herbe coupée de celle qui ne l'est pas. A partir de matrices de co-occurrence, ils définissent deux paramètres de mesure de la texture: l'entropie et l'homogénéité. Le calcul de ces attributs se fait aussi à partir de l'image de luminance.

[Luo 94] classe des images de scènes naturelles pour des textures telles que celles décrites dans [Brodatz 66]. Elles sont modélisées par un processus de mouvement Brownien fractionné. Son but est de segmenter l'image afin de détecter des amers intéressants pour guider le robot sur la route.

[Panjwani 95] segmente des images de scènes naturelles en modélisant la «texture couleur» à partir de chaînes de Markov. Il entend par «texture couleur» le fait qu'il existe une dépendance statistique entre les valeurs des trois plans RVB mesurées en un point et celles de ses pixels voisins.

[Thorpe 88] utilise pour classifier les pixels en route et non route des attributs de couleur et de texture. Le calcul de la texture combine l'utilisation de gradient de bas et haut niveau (changement de la résolution de l'image) pour déduire le nombre de micro-contours détectés à l'intérieur d'une fenêtre de dimension 8x8. Cet attribut lui permet de déterminer où se trouvent, dans l'image, les zones texturées et celles qui ne le sont pas et d'en déduire où se trouve la route.

Notre choix

Nous avons voulu privilégier la rapidité d'exécution plutôt que la finesse du résultat. La technique que nous avons choisie s'inspire de [Casals 92]: les paramètres décrits ont été définis avec l'idée de concilier pouvoir de discrimination et facilité de réalisation d'une carte câblée en vue d'une utilisation en temps réel.

De plus, les textures naturelles présentées dans [Brodatz 66] sont généralement définies à partir de paramètres de finesse, régularité ... qui diffèrent peu des paramètres présentés ici; ils s'expriment généralement en terme d'un ou plusieurs de ces paramètres.

Une région sera caractérisée à partir des six paramètres suivants: densité, granularité,

discontinuité, courbure, rectitude, et netteté. Ces paramètres sont basés sur l'étude des lignes de contraste à l'intérieur de la région : ces dernières sont obtenues à partir du gradient appliqué sur l'image d'intensité.

Chaque paramètre est calculé sur une fenêtre de taille 16x16 appelée par la suite *texel*. Pour calculer les attributs pour l'ensemble d'une région donnée, il suffira de déterminer le nombre de texels contenus dans celle-ci et de calculer les valeurs des paramètres sur chaque texel de la région ; en dernier lieu, les attributs associés à la région seront les moyennes de chacun des paramètres. Ils formeront ainsi le vecteur caractéristique (au niveau texture) de cette région.

Chacun des paramètres est défini de la façon suivante :

1. Densité : elle précise le nombre de lignes de contraste contenues à l'intérieur de la région,

$$T_{DS} = \sum_{i,j} DS_{ij}$$

où DS_{ij} est le nombre de lignes de contraste de longueur supérieure ou égale à 2 pixels et qui appartiennent au texel.

2. Granularité : ce paramètre quantifie la présence d'un grand nombre de points de gradient non concaténés dans une région,

$$T_G = \sum_{i,j} G_{ij}$$

où G_{ij} est le nombre de pixels qui appartiennent aux lignes de contraste de longueur inférieure ou égale à 2.

3. Discontinuité : elle indique le nombre de lignes discontinues dans la région,

$$T_D = \sum_{i,j} D_{ij}$$

où D_{ij} est le nombre d'extrémités des lignes de contraste de longueur supérieure à 2, internes au texel.

4. Courbure : elle mesure la densité des lignes de contraste non rectilignes dans la région,

$$T_C = \sum_{i,j} (C_{ij} - S_{ij})$$

où C_{ij} est le nombre de pixels qui appartiennent aux lignes de contraste de longueur supérieure à 3 et S_{ij} est le nombre de pixels qui sont alignés sur ces lignes de contraste.

5. Rectitude : elle fournit le nombre de lignes droites dans la région,

$$T_{ST} = \sum_{i,j} ST_{ij}$$

où ST_{ij} est le nombre de pixels alignés appartenant aux lignes de contraste de longueur comprise entre 5 et 8.

6. Netteté : une valeur élevée de ce paramètre indique que la région présente des variations lentes d'intensité dans toutes les directions. Au contraire, une valeur faible indique que la région présente des variations brusques d'intensité.

$$T_B = \frac{(A - D) + (C - B)}{1 + \sum_{i,j} P_{i,j}}$$

où A , B , C et D sont les moyennes du niveau de gris correspondant aux 4 sous-fenêtre carrées (de taille 8x8) contenues dans un texel sur l'image originale, et $P_{i,j}$ est le nombre de pixels appartenant aux lignes de contraste contenues dans le texel.

Il faut noter que l'obtention des différents attributs de texture est basée sur l'étude des lignes de contraste dans l'image. La qualité des attributs de texture dépend donc fortement de cette image : les contours doivent être différents sur chacune des classes définies.

Dans notre cas, le gradient de l'image est obtenu par application d'un masque de Sobel. Les lignes de gradient sont ensuite amincies ; et finalement, un seuillage par hystérésis permet de conserver uniquement les lignes de contraste d'amplitude suffisamment importante.

Le choix des différents seuils utilisés pour l'obtention de l'image de contraste conditionne le nombre de contours et leurs formes dans chacune des classes. Le lecteur pourra trouver dans [Murrieta Cid 95] une étude complète sur le choix des seuils adéquats. En pratique, l'image de lignes de contraste dépend uniquement de la valeur du seuil d'amincissement des lignes de gradient (la valeur du module du gradient doit être supérieure à 20).

D'autre part, une étude menée sur l'ensemble des attributs de texture est présentée en annexe B. Le lecteur pourra en particulier vérifier que les attributs choisis sont discriminants pour certaines classes mais pas pour toutes. La caractérisation à partir de tous ces attributs est donc justifiée.

Le graphe de la figure IV.39 montre qu'effectivement les nuages des classes sont bien définis. Il reste cependant un problème : les classes « roche » et « arbre » s'intersectent et risquent d'entraîner des erreurs.

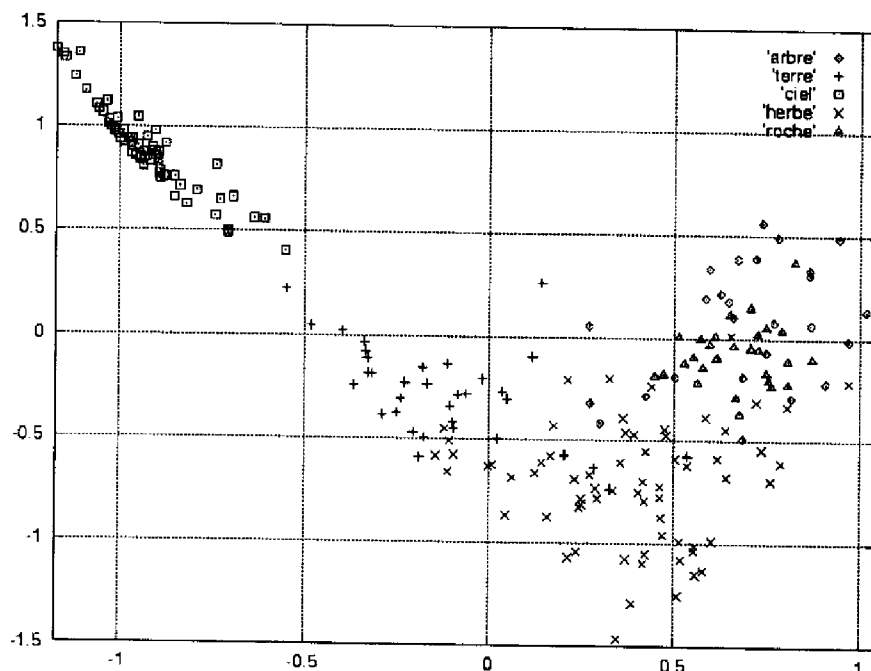


FIG. IV.39 – Représentation après analyse factorielle discriminante sur l'ensemble des attributs de texture

IV.3.3 Conclusion

Finalement, toute région de l'image pourra être caractérisée à partir d'un vecteur contenant à la fois des informations sur sa couleur et sa texture. Ce vecteur contient les moyennes statistiques des six paramètres de texture définis précédemment, et les moyennes colorimétriques rouge, verte et bleue :

$$\langle m_{\text{densité}}, m_{\text{granularité}}, m_{\text{discontinuité}}, m_{\text{courbure}}, m_{\text{rectitude}}, m_{\text{netteté}}, \\ m_{\text{Rouge}}, m_{\text{Vert}}, m_{\text{Bleu}} \rangle$$

Les deux graphiques IV.40 et IV.41 présentent les différents nuages des classes lorsque les attributs de texture et de couleur sont utilisés (ces graphes ont été obtenu à partir de l'analyse factorielle discriminante sur les neufs attributs. Cette méthode sera détaillée dans le chapitre suivant en section V.2.1).

Ils montrent qu'il est effectivement possible, à partir de ces neufs attributs, de séparer relativement bien les nuages des différentes classes.

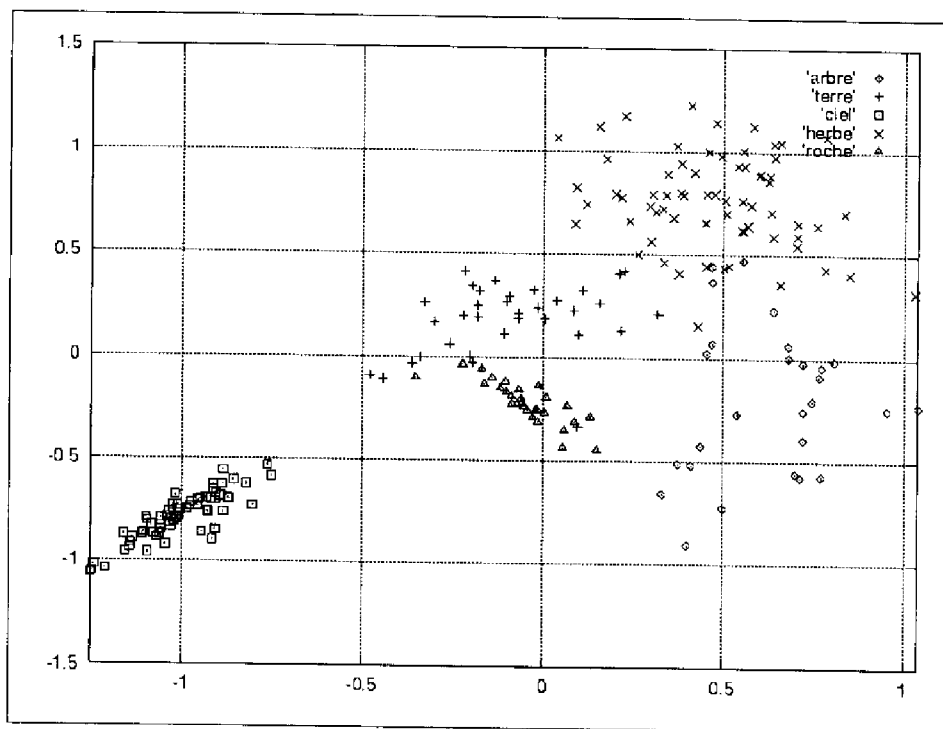


FIG. IV.40 – Représentation des classes après analyse factorielle discriminante sur RVB et texture (2 axes discriminants conservés)

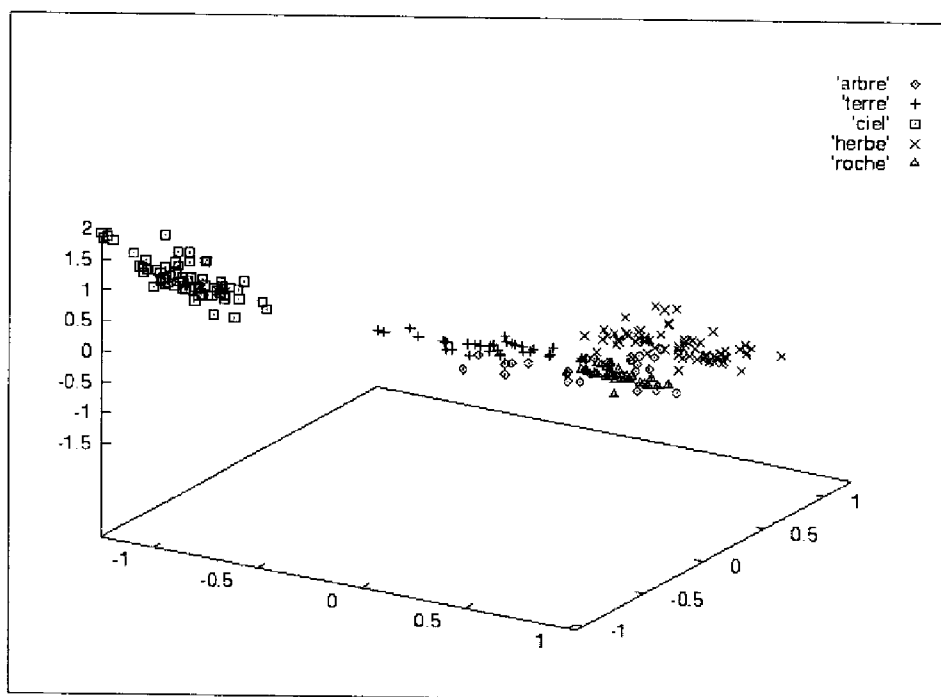


FIG. IV.41 – Représentation des classes après analyse factorielle discriminante sur RVB et texture (3 axes discriminants conservés)

IV.4 Conclusion

Au cours de ce chapitre, nous avons présenté l'outil de base pour la construction d'un modèle de l'environnement. Nous avons tout d'abord introduit les différentes fonctions nécessaires à la réalisation de cet outil ainsi que la manière de les intégrer aux fonctions déjà existantes. Nous avons montré ensuite que le capteur caméra peut apporter des informations pertinentes pour la modélisation de deux manières :

- la segmentation couleur peut fournir une description plus ou moins synthétique de l'image : elle peut produire soit des régions colorimétriquement très homogènes soit les principales composantes de l'image ; dans ce second cas, en association avec la segmentation 3D, une image de régions 3D pourrait être construite par fusion des deux informations : la segmentation couleur fournirait les contours, tandis que la segmentation 3D nous préciserait le nombre d'objets différents (même s'ils sont homogènes colorimétriquement) ;
- la caractérisation fournit des informations complémentaires pour déterminer la nature d'une région : autant pour la localisation que pour la navigation du robot, cette information serait très enrichissante. Les informations de texture et de couleur telles quelles ont été calculées pourront facilement être intégrées à la fonction de perception sur le robot (les algorithmes ne nécessitent pas de gros moyens de calcul).

Finalement, nous avons à ce stade une image de régions caractérisées par leur couleur et leur texture. Il semble naturel de vouloir extraire une information plus riche qu'une simple étiquette : associer à chaque région de l'image, la nature de l'objet qu'elle représente. Nous avons donc défini 5 classes correspondant aux principaux éléments de la scène : herbe, roche, terre, arbre et ciel. Le chapitre suivant sera consacré à l'identification de ces régions à partir de leurs caractéristiques.

Chapitre V

Identification des régions

L'identification consiste à donner un nom aux différentes régions de la scène. Elle se fait généralement à partir d'une base de connaissances établie pendant une phase d'apprentissage.

Suivant la tâche à effectuer par le robot, navigation ou localisation, cette identification peut être de nature différente : la localisation s'intéresse à la nature des objets constituant la scène (arbre, rochers...); la navigation, quant à elle, préfère connaître la «praticabilité» du terrain.

Cette base de connaissances peut cependant être commune aux deux tâches en considérant que l'on fournira la nature des éléments constituant la scène. Il suffira alors de définir la «praticabilité» à l'aide de règles adaptées aux capacités du robot utilisé.

Les caractéristiques des régions étant connues, il s'agira dans ce chapitre de présenter la base de connaissances que nous avons développée ainsi que la technique employée pour identifier les régions de la scène.

Nous montrerons ensuite que l'analyse des résultats peut permettre de détecter certains des problèmes dus à des erreurs au moment de la reconnaissance ou de la segmentation. L'utilisation du contexte applicatif peut aussi guider l'identification et améliorer ainsi la qualité du modèle fourni.

V.1 Objectifs et principe

La segmentation produit des régions colorimétriquement homogènes mais ne fournit qu'un étiquetage des régions.

Donner un nom, et donc identifier la nature de chacune des régions, fournit une information

plus riche et plus exploitable par le niveau décisionnel du robot.

L'identification des régions demande de classifier chacune des régions de l'image en l'affectant à l'une des classes pré-définies. Afin de conserver un caractère général à celle-ci, nous avons défini des classes présentes dans la plupart des scènes naturelles : terre, herbe, roche, arbres et ciel.

L'affectation de chacune des régions à une classe est réalisée en comparant les données perceptuelles à une base de connaissances fournie par l'utilisateur, qui définit, pour chacune des classes, ses principales caractéristiques. Il s'agit alors d'affecter la région à la classe qui a les caractéristiques les plus proches suivant une fonction de ressemblance choisie.

La caractérisation des différentes classes, et par la suite, des régions à identifier, a été détaillée dans le chapitre précédent (cf. section IV.3). Elle est réalisée à partir des neuf attributs (texture et couleur).

Le principe de fonctionnement pour identifier les régions est montré sur la figure V.1 et suit le déroulement suivant :

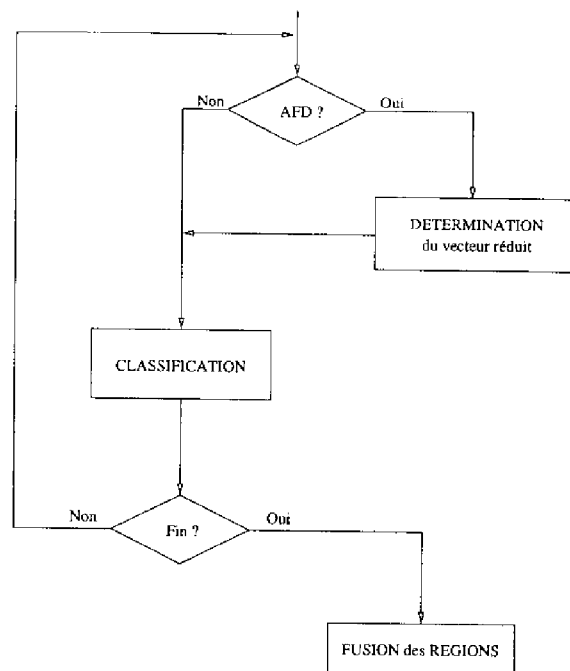


FIG. V.1 - Déroulement de la phase d'identification

La *classification* peut être envisagée soit en utilisant les neuf attributs précédemment définis, soit en utilisant un vecteur réduit, issu de l'analyse factorielle discriminante (notée AFD sur le schéma), suivant le choix de l'utilisateur.

Le résultat obtenu est ensuite validé ou rejeté en vérifiant l'homogénéité sur la région. Si

la région n'est pas suffisamment homogène, c'est-à-dire que la région aurait pu être affectée à plusieurs classes, une nouvelle étape de segmentation est relancée et les nouvelles régions issues de cette dernière sont à nouveau classifiées.

Lorsque cette étape est terminée, les régions de même nature et connexes sont fusionnées pour dégager les grandes régions. On obtiendra ainsi, le modèle nominatif de régions. De simples règles issues du contexte applicatif permettront d'infirmier ou de confirmer les résultats obtenus.

Nous verrons, dans un premier temps, comment nous avons créé la base de connaissances du système de classification. Nous présenterons aussi la méthode de réduction du vecteur d'attributs. La deuxième partie de ce chapitre sera consacrée à la classification. La méthode employée sera tout d'abord détaillée. Viendront ensuite les résultats obtenus sur les régions issues de la segmentation. Finalement, nous préciserons comment est réalisée la fusion des différentes régions afin d'obtenir le modèle final.

V.2 Génération de la base de connaissances

La première étape pour identifier les régions est une phase d'apprentissage des différentes classes qui permet de définir la base de connaissances.

L'étude présentée ci-après a été réalisée à partir de régions issues de la segmentation. Des régions de chaque classe «herbe», «terre», «ciel», «arbre» et «roche» ont été prélevées. Chacune de ces régions a été caractérisée à l'aide des attributs définis précédemment.

La base de connaissances doit correctement caractériser chaque classe. Les régions qui serviront de base à la reconnaissance ont donc été sélectionnées uniquement si elles contenaient des éléments d'une seule classe. Ceci explique en partie le choix de la première technique de segmentation proposée (méthode 1, section IV.2.3). L'autre raison est que les tailles des régions obtenues par cette technique sont du même ordre de grandeur que celles issues de la technique de discrétisation en cellules proposée par S. Lacroix.

Les 221 *régions-échantillons* prélevées se répartissent de la manière suivante : 27 régions d'arbres, 34 régions de terre, 64 régions de ciel, 67 régions d'herbe, 29 régions de roche.

Nous nous sommes imposés d'avoir un nombre d'échantillons identique sur chacune des classes. Malgré la variété d'images dont nous disposons, nous avons été limités par les régions de la classe «arbre» (seulement 27 *échantillons caractéristiques* disponibles). Nous avons donc limité le nombre d'échantillons dans les autres classes.

Bien que ce choix ait été peu judicieux, nous verrons qu'il n'a pas altéré la qualité du résultat dans la mesure où les problèmes rencontrés sont essentiellement liés à la classe «arbre».

Le choix des échantillons représentatifs étant important, nous avons vérifié la validité

des critères visuels par une méthode plus théorique. Nous avons choisi l'analyse factorielle discriminante, notée AFD, car elle permet de réduire le vecteur caractéristique initial en recherchant les axes de projection les plus discriminants. Il sera ainsi plus aisé de visualiser les nuages de chaque classe.

De plus, au delà de cette vérification, il sera aussi particulièrement intéressant de comparer les résultats de l'identification à partir des neuf attributs ou à partir du nouveau vecteur fourni par l'AFD. En effet, la propriété de l'AFD est d'agencer les nuages des différentes classes de façon à les séparer au mieux dans le nouveau repère.

V.2.1 L'analyse factorielle discriminante

Le problème est de déterminer les meilleures combinaisons linéaires des attributs permettant de bien séparer les différentes classes, c'est-à-dire rechercher les meilleurs axes de projection u .

Une bonne discrimination doit condenser les classes, donc minimiser la variance intra-classe, et séparer les classes, donc maximiser la variance interclasse.

Le critère à maximiser est de la forme :

$$J(u) = \frac{u^t(Bu)}{u^tWu}$$

et peut se ramener à :

$$J(u) = \frac{u^t(Bu)}{u^tTu}$$

où : W est la variance intraclasse,

B est la variance interclasse,

T est l'estimation de la matrice de covariance totale; cette matrice est généralement inversible et on peut écrire $T^{-1}Bu = \lambda u$.

Les axes discriminants s'obtiennent alors en déterminant les vecteurs propres de la matrice $T^{-1}B$. Le critère devient finalement :

$$J(u) = \frac{u^t(Bu)}{u^tTu} = \frac{u^t(\lambda Tu)}{u^tTu} = \lambda$$

où la valeur propre λ_i représente la capacité discriminante de l'axe i : plus cette valeur propre est grande, plus l'axe est discriminant.

A partir de la partition initiale de l'ensemble des échantillons, l'AFD fournit les vecteurs propres et valeurs propres du nouvel espace de classification.

L'AFD appliquée aux données indique que trois axes de projection contiennent la quasi-totalité de l'information.

Les figures V.2 et V.3 montrent l'organisation des différentes classes dans le nouveau repère fourni par l'AFD (27 régions-échantillons par classe).

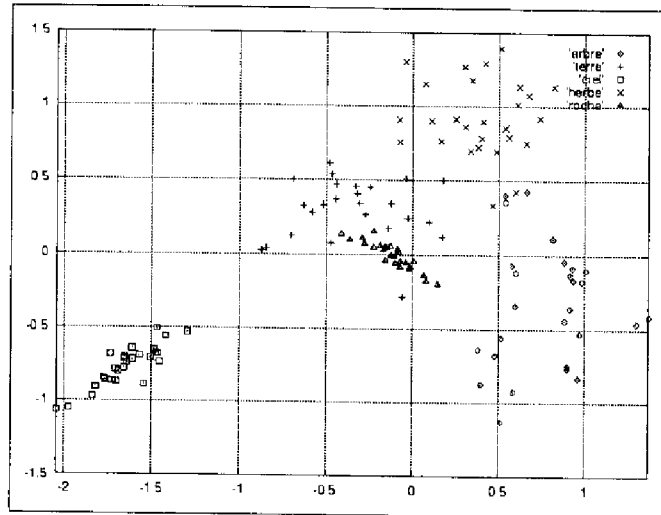


FIG. V.2 - Représentation des classes après AFD sur l'ensemble des attributs (2 axes discriminants conservés)

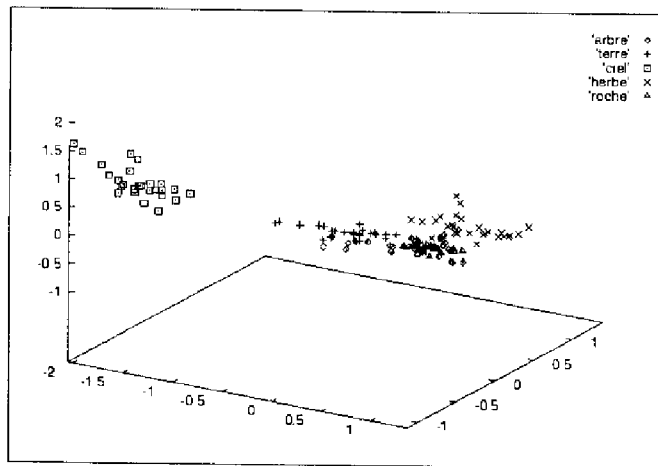


FIG. V.3 - Représentation des classes après AFD sur l'ensemble des attributs (3 axes discriminants conservés)

Le résultat de la classification est intéressant : les classes sont assez bien séparées mais la classe «arbre» reste toujours la plus dispersée.

Nous avons donc une base de connaissances contenant, pour chaque classe, 27 régions-échantillons caractéristiques. Cette base de données servira pour l'identification future de régions issues d'autres images.

V.3 Classification des régions

La classification consiste à affecter une région à l'une des classes prédéfinies. L'utilisation de méthodes statistiques permet d'évaluer la probabilité d'appartenance de la région à l'une ou l'autre des classes. La région sera naturellement affectée à la classe pour laquelle sa probabilité d'appartenance est la plus élevée.

Nous présentons donc, tout d'abord, la technique de classification utilisée à savoir, la classification bayésienne. Nous verrons ensuite les résultats obtenus sur des régions de tailles différentes (plus ou moins grandes en fonction du processus de segmentation utilisé, ou de l'élément à segmenter).

V.3.1 La classification bayésienne

La classification bayésienne est une approche statistique puissante qui permet de calculer la probabilité d'appartenance. Cette classification est basée sur (i) les attributs mesurés, (ii) les probabilités *a priori* sur la répartition des classes, et (iii) sur les fonctions de densité de probabilité conditionnelle des attributs [Andrews 72]. La formule de Bayes est définie par :

$$P(C_i | X) = \frac{P(X | C_i)P(C_i)}{\sum_{i=1}^n P(X | C_i)P(C_i)}$$

- $P(C_i)$ est la probabilité *a priori* sur la répartition des classes,
- $P(X | C_i)$ désigne la fonction de densité de probabilité sous-jacente à la distribution des observations provenant de la classe C_i .
- $P(C_i | X)$ représente la probabilité conditionnelle *a posteriori* pour que l'élément associé à X appartienne à la classe C_i .

Nous avons considéré que la probabilité *a priori* $P(C_i)$ était identique pour chacune des classes. Dans ce cas, le calcul de la probabilité *a posteriori* $P(C_i | X)$ est simplifié et se ramène uniquement au calcul de la probabilité conditionnelle $P(X | C_i)$.

$P(X | C_i)$ peut être estimée par des méthodes paramétriques ou non paramétriques. Nous avons choisi une méthode non paramétrique : la méthode du $k^{\text{ième}}$ plus proche voisin. Cette méthode est adaptée lorsque les classes ne sont pas bien séparées dans l'espace de classification. D'autre part, elle ne nécessite pas obligatoirement de connaître la loi de la densité de probabilité suivie par les classes.

Estimation locale non paramétrique

Les méthodes du noyau et du $k^{\text{ième}}$ plus proche voisin sont des procédures classiques d'estimation locale non paramétrique de la densité de probabilité.

Toutes deux s'appuient sur le fait que la densité de probabilité p en un point x peut

être estimée en comptabilisant le nombre d'observations dans un petit domaine \mathcal{D} autour de ce point.

Soient V le volume du domaine considéré,
 k le nombre d'échantillons à l'intérieur de ce volume,
 et n le nombre d'échantillons pour une classe donnée.

Sous certaines conditions de convergence [Briot 77], l'estimation de p se déduit de la formule suivante :

$$p(x) = \frac{k}{n.V}$$

La méthode du noyau travaille à volume constant V et assigne une observation à la classe qui a le plus d'éléments k dans ce volume tandis que la méthode du $k^{\text{ième}}$ plus proche voisin travaille à k constant et assigne l'observation à la classe dont le volume est le plus petit.

Nous avons choisi d'estimer la probabilité $P(X | C_i)$ par la méthode du $k^{\text{ième}}$ plus proche voisin. Le calcul de $P(X | C_i)$ se ramène alors simplement au calcul du volume V . Ce volume peut être approché par une hypersphère de dimension égale au nombre d'attributs servant à caractériser la classe. La seule variable intervenant dans le calcul de ce volume est la distance du point X au $k^{\text{ième}}$ voisin choisi (X correspondant à la région à identifier et dont les coordonnées sont fournies par le vecteur d'attributs). Il suffit donc de calculer cette distance pour évaluer les probabilités $P(X | C_i)$ et $P(C_i | X)$. Finalement, l'observation X (soit la région) sera affectée à la classe C_i dont la distance au $k^{\text{ième}}$ voisin sera la plus petite.

La classification sera donc finalement effectuée en calculant la distance euclidienne entre l'observation et le $k^{\text{ième}}$ voisin. Notons que les distances seront déterminées après avoir normalisé l'ensemble des attributs, ceci afin d'avoir un poids identique sur chacune des composantes du vecteur.

V.3.2 Validation expérimentale

Une étude préliminaire menée par R. Murrieta a permis de montrer que la classification à partir des attributs de texture proposés était envisageable. Cette étude s'appuyait sur une base de données construite à partir de pavés caractéristiques prélevés dans l'image et correspondant à chacune des classes. Elle a néanmoins fait ressortir la nécessité de combiner couleur et texture pour la caractérisation afin d'améliorer l'identification.

Notre objectif a donc été d'une part, de réaliser l'identification sur un plus grand nombre d'images et d'ajouter à la texture les informations colorimétriques ; d'autre part de bâtir la base de connaissances sur des régions issues de la segmentation et non des pavés prélevés dans l'image. Nous avons souhaité introduire cette contrainte car il est préférable

d'imaginer que, à long terme, l'apprentissage de nouvelles classes se fasse «en ligne», c'est-à-dire que la création d'une nouvelle classe puisse s'effectuer au fur et à mesure qu'elle apparaît dans l'environnement (par télé-programmation par exemple).

L'identification des régions a été réalisée sur des régions de tailles différentes :

- sur des régions dont la taille est du même ordre de grandeur que celle des régions utilisées pour l'apprentissage ; il s'agit dans ce cas d'identifier des régions dans des images sur lesquelles la première technique de segmentation a été appliquée (méthode 1, en section IV.2.3 page 58) ;
- sur des régions de taille inférieure : lorsque l'analyse de l'identification justifie la nécessité de relancer le processus de segmentation, une nouvelle phase d'identification des nouvelles régions obtenues est nécessaire ;
- sur des régions de taille plus importante ; la deuxième technique de segmentation (méthode 2, en section IV.2.3 page 63) permet d'obtenir des régions correspondant aux principales composantes de l'image. Les régions sont généralement plus grandes et par là même parfois moins homogènes.

Après avoir présenté les différents résultats obtenus, nous les comparerons afin de préciser la qualité de l'identification et la robustesse des algorithmes.

Identification de régions obtenues par la méthode 1

L'identification des régions a été réalisée sur neuf images différentes non utilisées pour la génération de la base de données.

Chaque image comporte en moyenne 21 régions colorimétriquement homogènes. Pour analyser le résultat de la reconnaissance, nous avons fixé des critères permettant de prédire si le logiciel sera capable d'identifier correctement les régions. Ces critères sont bien entendu visuels : nous avons ainsi séparé les régions en *régions identifiables* et *régions non identifiables*.

Une région est dite *identifiable* si elle satisfait les conditions suivantes :

1. elle est **homogène** : elle contient essentiellement des éléments d'une même classe ;
2. elle est **reconnaissable** : la **classe** à laquelle elle devrait appartenir est **définie** dans la base de données ;
3. elle est correctement **texturée** : sa texture est du même type que celle définie pour les régions de la même classe dans la base de connaissances ;

Si une région ne satisfait pas ces critères, nous la considérons comme *non identifiable*.

Ces définitions nous ont permis d'évaluer le résultat de l'identification mais aussi de la

segmentation. Les pourcentages affichés dans le tableau V.1 ont été obtenus de la manière suivante :

- Pourcentage de réussite de l'identification :

1. pour une classe C_i :

$$\mathcal{R}_{Identification}(C_i) = \frac{N_{reconnues}(C_i)}{N_{reg}(C_i)} * 100$$

2. au total :

$$\mathcal{R}(Total) = \frac{N_{reconnues}(Total)}{N_{reg}(Total)} * 100$$

Il faut remarquer que N_{reg} correspond au nombre de régions identifiables et non à l'ensemble des régions de l'image (voir tableau V.1). En effet, nous souhaitons connaître quand le procédé échoue alors qu'il ne le devrait pas (les régions que nous considérons comme *non identifiables* correspondent en effet aux cas où le procédé ne peut *a priori* qu'échouer).

- Pourcentage de réussite pour la segmentation :

$$\mathcal{R}_{Segmentation}(Total) = \frac{N_{reg}(Total) - N_{nonhomogènes}(Total)}{N_{reg}(Total)} * 100$$

En effet, une mauvaise segmentation correspond au fait que le critère d'homogénéité n'est pas vérifié.

Dans ce cadre, une région qui ne satisferait pas plusieurs critères dont celui d'homogénéité, sera dite non identifiable car non homogène.

L'identification a été réalisée pour différentes valeurs de k . La valeur de k peut être obtenue par l'équation $k = k_0\sqrt{N}$ où N est le nombre d'échantillons par classe et k_0 est fixé par l'utilisateur.

Expérimentalement, nous avons testé pour des valeurs de k égales à 4, 5 et 6. Les résultats obtenus varient peu (une ou deux régions sont différentes sur les neuf images testées).

Les résultats de cette analyse sont regroupés dans le tableau suivant (valeur de k égale à 4) :

		Nombre de régions N_{reg}	Nombre de régions reconnues $N_{Reconnues}$		pourcentage de réussite \mathcal{R}	
			sans AFD	avec AFD	sans AFD	avec AFD
Régions identifiables	Arbre	41	17	16	41%	39%
	Terre	16	16	16	100%	100%
	Ciel	48	46	47	96%	98%
	Herbe	33	27	31	82%	94%
	Roche	6	6	6	100%	100%
	Total	144	112	116	78%	81%
Régions non identifiables	Non homogènes	24	14	15	58%	62%
	Mal texturées	29				
	Classes non définies	6				
Segmentation	Total	197	173		88%	

TAB. V.1 – Résultats de l'identification et de la segmentation (méthode 1)

Finalement, on peut voir que les pourcentages de réussite pour la segmentation et la reconnaissance sont relativement bons (respectivement 88% et 78%).

Le résultat de l'identification doit être modulé par le fait que certaines régions sont difficiles à reconnaître : la classe «arbre» n'est reconnue que dans 41% des cas. Cet état de fait est compréhensible :

- la texture varie beaucoup en fonction de la distance,
- il y a beaucoup de classes d'arbres possibles. Malheureusement, nous ne disposons pas d'une base de données assez importante.
- la couleur des feuilles (lorsqu'il y en a) varie avec les saisons. La base de connaissances a été fortement limitée au niveau des arbres pour cette raison.

Dans notre cas, la principale gêne est la variation de distance et la variété des arbres dans nos images (du pin au platane). Dans la plupart des cas nous retrouvons donc des erreurs du type :

- «arbre» pris pour «roche» (cas des arbres éloignés, peu de couleur, texture proche de celle des roches),
- «arbre» pris pour «terre» (peu de texture car pas de feuillage ou éloigné dans le cas de pins).

On peut aussi noter que 58% des régions définies comme non homogènes ont pu cependant être correctement identifiées.

Les autres erreurs commises lors de l'identification des régions sont essentiellement liées à une mauvaise reconnaissance de la classe «herbe». Elle est parfois identifiée comme «terre» lorsque la région n'est pas assez texturée (problèmes de distance, d'ombres) ou comme «arbre» lorsqu'elle est trop proche.

Ces erreurs pourraient être réduites en ajoutant des attributs liés à l'information tridimensionnelle. Nous n'avons pas introduit, dans un premier temps, la relation entre texture et distance. En revanche, nous considérons que cette relation ne doit pas être négligée. Il serait intéressant de l'introduire et d'étudier son effet sur la reconnaissance.

La figure V.5 montre le résultat de l'identification sur l'image résultant de la segmentation de l'image V.4.

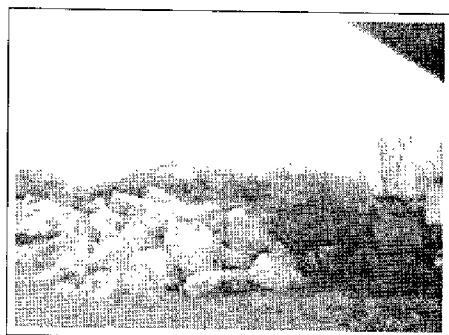


FIG. V.4 - Image d'intensité de l'image IV.11

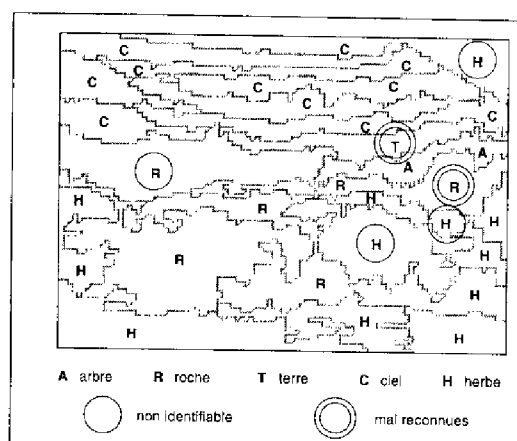


FIG. V.5 - Résultat de l'identification de l'image V.4 (segmentation méthode 1)

Les symboles *A*, *C*, *T*, *H* et *R* représentent la première lettre de la classe à laquelle la région a été affectée. Sont précisées sur la figure V.5 les régions que nous considérons comme *non identifiables* (entourées d'un cercle) et les régions mal reconnues lors du processus de reconnaissance (doublement encerclées).

Parmi les 18 régions considérées comme *identifiables*, 16 régions ont pu être correctement identifiées (1 région «arbre» sur 2, 8 régions de «ciel» sur 9, 5 régions «herbe» sur 5 et 2 régions «roche» sur 2).

De plus, deux régions sur quatre que nous considérons non identifiables ont été affectées

tées à la bonne classe («herbe»).

L'une des deux dernières régions non reconnues était indéfinie (bord de l'image) : elle correspond à une partie d'un panneau solaire ; l'autre (reconnue comme de la roche) est en réalité un mélange de plusieurs classes (mélange d'arbres et de ciel). Dans ce dernier cas, il paraît souhaitable de re-segmenter la région.

Nous présenterons, un peu plus loin, le résultat de l'identification sur les régions re-segmentées (voir V.3.2).

De manière similaire, l'étude a été menée sur le vecteur d'attributs réduit obtenu par AFD. Ce nouveau vecteur n'a plus que trois composantes (au lieu de neuf). Les résultats de l'identification montrent que le pourcentage de réussite est légèrement supérieur (81%).

Les classes qui étaient généralement bien identifiées le sont toujours, parfois mieux comme dans le cas de la classe «herbe» ; en revanche, le problème de reconnaissance de la classe «arbre» s'est amplifié.

L'AFD, en recherchant les meilleurs axes de projection, assure une bonne séparation entre les classes bien définies mais ne résout pas le problème de dispersion d'une classe : si elle est dispersée, elle le reste.

Elle garantit donc la meilleure séparation possible de *l'ensemble des classes* mais pas obligatoirement celle *d'une classe donnée par rapport aux autres*.

Le résultat de l'identification était dans le cas de l'ADF de 17 régions reconnues sur 18 (la seconde région d'«arbre» a été correctement identifiée) ; les régions que l'on considérait comme non identifiables ne l'ont effectivement pas été (1 reconnue sur 3).

Il faut aussi noter que cette technique indique les probabilités d'appartenance de la région à chacune des classes pré-définies. Il faut donc ensuite déterminer à quelle classe l'affecter.

Une classe de rejet dans laquelle seraient contenues les régions mal identifiées, c'est-à-dire dont la probabilité d'appartenance serait inférieure à un seuil de rejet, doit être introduite. Cette partie sera présentée en section V.4.

Identification de régions obtenues par la méthode 2

La même étude a été reprise après segmentation à partir de la seconde méthode.

Le nombre de régions par image est en moyenne de 11. La valeur de k est toujours égale à 4.

Les résultats sont regroupés dans le tableau V.2.

		Nombre de régions N_{reg}	Nombre de régions reconnues $N_{Reconnues}$	pourcentage de réussite \mathcal{R}
Régions identifiables	Arbre	30	20	67%
	Terre	4	3	75%
	Ciel	14	13	93%
	Herbe	15	12	86%
	Roche	4	4	100%
	Total	67	52	78%
Régions non identifiables	Non homogènes	16	9	56%
	Mal texturées	15		
	Classes non définies	3		
Segmentation	Total	103	87	84%

TAB. V.2 – Résultats de l'identification et de la segmentation (méthode 2)

Les résultats obtenus sont du même ordre de grandeur que précédemment (respectivement 78% et 84% pour l'identification et la segmentation).

Notons toutefois que les résultats obtenus sur chacune des classes ont une variance plus faible. La classe «arbre» est toujours la moins bien reconnue mais les résultats sont tout de même améliorés.

La figure V.6 montre le résultat de la reconnaissance après segmentation de l'image V.4 présentée précédemment (reprise ci-dessous).

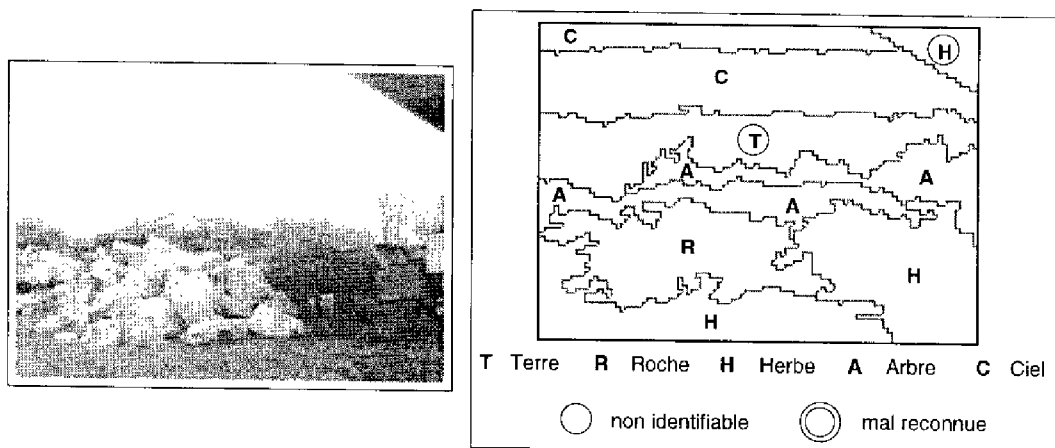


FIG. V.6 – Résultat de l'identification de l'image V.4 (segmentation méthode 2)

Les régions considérées comme *identifiables* l'ont été (7 régions sur 9 au total dans l'image). Les deux régions *non identifiables* (encadrées sur la figure V.6) étant comme précédemment la région comprenant un mélange d'arbres et de ciel (reconnue comme de la terre), et la région du bord de l'image (définie comme de l'herbe). Il est donc normal que le processus échoue dans ces deux cas.

Plusieurs remarques importantes peuvent finalement être déduites :

- Le procédé de reconnaissance est robuste : les régions à reconnaître sont de taille très différente de celles contenues dans la base de données. La base de données est donc bien caractéristique et le moyennage sur chaque attribut reflète bien son comportement général sur la région.
- La non-homogénéité d'une région n'est pas toujours gênante : lorsque l'une des classes prédomine dans la région, cette dernière est reconnue. Ceci est particulièrement intéressant car cela permet d'envisager sagement l'identification sur de grandes régions.
- Il est préférable de reconnaître des grandes régions : le nombre de régions constituant ainsi la scène est réduit et chaque région bien identifiée fournit une information très importante sur la scène finale.

Identification de petites régions

Un certain nombre de régions issues de la segmentation ne sont pas identifiables à cause de leur non-homogénéité (plusieurs classes sont incluses dans une même région). Dans ce cas, un nouveau processus de segmentation (méthode 1) est relancé.

Nous détaillerons par la suite comment déterminer algorithmiquement quelles régions doivent ou non être re-segmentées (cf section suivante V.4) et intéressons nous plutôt au résultat de l'identification sur celles-ci.

Reprenons l'exemple précédent concernant l'image de la figure V.4. La figure V.7 indique les régions qui sont détectées comme à re-segmenter (régions encadrées).

La segmentation est relancée sur les quatre régions précisées : les régions connexes sont segmentées pendant la même phase de segmentation. La segmentation sera donc dans cet exemple relancée trois fois.

La figure V.8 montre le résultat de la reconnaissance sur les régions obtenues après segmentation des deux régions connexes de l'image V.7 (représentées en gris).

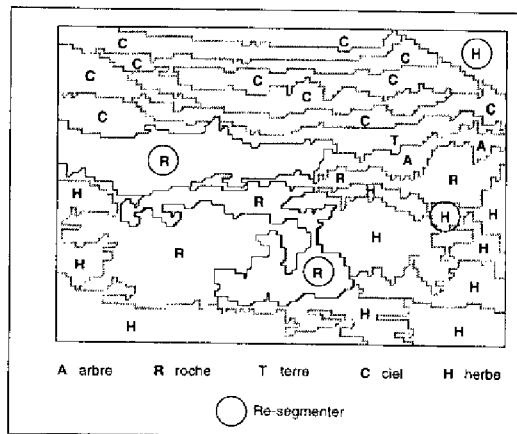


FIG. V.7 – Evaluation de la segmentation (méthode 1) de l'image V.4

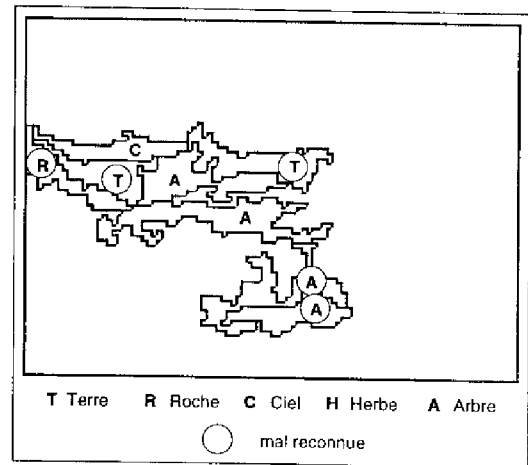


FIG. V.8 – Identification des régions re-segmentées

Sur 8 régions, 5 sont mal reconnues. Ces résultats montrent que les régions issues de la re-segmentation sont trop petites pour être correctement identifiées; la moyenne des différents paramètres, dans ce cas, ne caractérise plus vraiment la région. Il n'est donc pas obligatoirement profitable de relancer tout le processus. Celui-ci doit être conditionné par la taille des régions obtenues lors de la seconde phase de segmentation.

Il faut noter que ce résultat est obtenu pour des régions issues de la première méthode de segmentation. Dans le cas de la méthode 2, la remarque précédente demeure: si les régions sont généralement de tailles supérieures, il n'en reste pas moins que certaines régions n'ont pas des tailles suffisantes pour être reconnues. En revanche, affiner des grandes régions devient possible. Les deux exemples des figures V.9 et V.10 reprennent ces différents aspects.

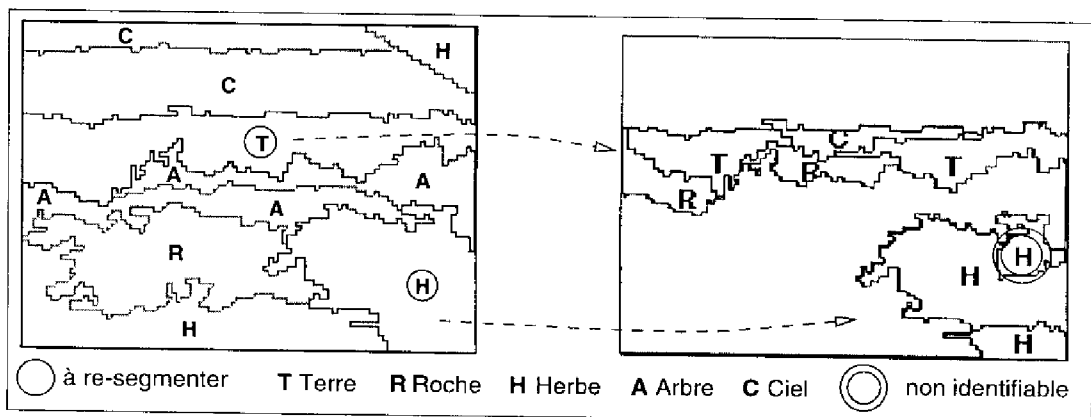


FIG. V.9 – Evaluation de la segmentation (méthode 2) de l'image V.4 et identification des régions re-segmentées

Dans la figure V.9, on peut remarquer que l'affinement de la zone d'herbe s'effectue correctement : les deux régions d'herbe sont correctement identifiées. La troisième (encadrée sur la figure) n'est pas identifiable (accessoire posé sur le sol et entouré d'herbe). En revanche, la zone supérieure qui a été resegmentée n'améliore pas le résultat : au lieu d'une grande région de terre, on se retrouve avec plusieurs petites régions, mal identifiées (terre, roche). Seule une des régions est correctement reconnue comme du ciel. La taille des régions résultantes intervient bien entendu dans ce résultat ; ce phénomène se confirme dans le second exemple (figure V.10).

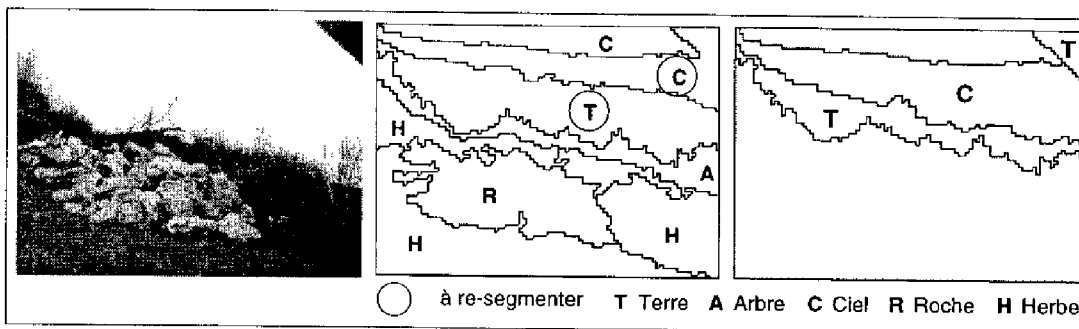


FIG. V.10 – Evaluation de la segmentation (méthode 2) et identification sur une image du même type

Ici, la segmentation permet en effet d'affiner la frontière entre ciel et arbre même si la région de terre subsiste. Re-segmenter dans ce cas est intéressant.

Finalement, on peut remarquer que la re-segmentation des images est à manipuler avec précaution. Elle permet d'affiner le résultat de la segmentation lorsque les régions à re-segmenter sont suffisamment grandes. Dans le cas contraire, ce type de traitement risque de nuire au résultat final. Notons toutefois qu'il serait intéressant de reprendre cette étude en augmentant le nombre d'images à reconnaître et en introduisant la notion de distance pour la texture lors de l'identification. Il serait alors possible de mieux évaluer l'intérêt d'une telle approche avec ces attributs.

V.3.3 Conclusion

Nous venons de présenter les résultats de l'identification sur des régions variées, issues de procédés de segmentation différents.

On peut noter que la segmentation en grandes régions (méthode 2) est beaucoup plus intéressante que la première technique employée (méthode 1). D'une part, le nombre de régions à identifier est beaucoup plus faible ; d'autre part, traiter de grandes régions permet de minimiser les erreurs d'identification (une classe ayant de fortes chances de prédominer dans la région). Ceci est en partie dû aux attributs de texture choisis : l'homogénéité en texture n'est assurée que dans la mesure où la région est assez importante. Il arrive

fréquemment que sur une région d'herbe par exemple, une petite zone soit peu texturée alors que dans sa globalité, la région est homogène en texture.

Les attributs choisis sont pertinents et la base de connaissances caractérise bien les principales composantes de l'image.

On peut cependant remarquer que l'identification est satisfaisante tant que les régions à identifier ont une taille supérieure à celles des régions ayant servi pour l'apprentissage. Des régions trop petites ne permettent pas en effet d'avoir des valeurs moyennes proches de celles calculées pour la base de connaissances (surtout pour les attributs de texture). Ce type de problème pourrait être résolu en permettant de relancer l'identification sur certains attributs uniquement, ou si le besoin s'en fait sentir, en relançant le processus de segmentation, les régions étant assez grandes pour le permettre.

Mais cette dernière idée repose sur le fait que l'on sait évaluer le résultat de la segmentation et de la reconnaissance. C'est l'objet de la section suivante.

V.4 Evaluation de l'identification et de la segmentation

Notre but n'est pas seulement d'indiquer la nature de chaque région segmentée mais aussi de connaître, dans la mesure du possible, le degré de confiance que l'on peut accorder au résultat fourni.

Nous voulons finalement quantifier, grâce à des critères, la qualité de la segmentation et de l'identification : ceci permettra de définir si une région doit être re-segmentée ou non, et si la reconnaissance est fiable ou pas [Charroux 96]. Par la suite, on peut faire coopérer segmentation et identification afin d'obtenir une meilleure description de la scène.

Nous avons défini deux critères pour évaluer la qualité de la segmentation et de l'identification :

- la probabilité $P(C_i | X)$: définie précédemment (cf équation V.3.1). Une trop faible probabilité, sur la classe correspondant à l'affectation de la région, ne permet pas de garantir l'appartenance à cette classe ;
- l'entropie : elle mesure l'ambiguïté existant entre les différentes classes. Une entropie forte correspond à des probabilités fortes sur plusieurs classes tandis qu'elle sera faible lorsque seule une classe aura une probabilité forte.

$$Entropie = - \sum_{i \in Classe} P(C_i | X) \log P(C_i | X)$$

Elle permet de vérifier si une région est homogène : si elle dépasse un certain seuil, nous considérerons que la région doit être re-segmentée.

Ces critères ont été calculés dans un premier temps sur les régions segmentées par la méthode 1. Par la suite, nous avons recherché les valeurs de seuils adaptés à la seconde méthode.

Pour chaque classe, des seuils d'entropie et de probabilité ont été fixés. Ces seuils ont été déterminés expérimentalement par comparaison des valeurs d'entropie et de probabilité suivant les critères d'homogénéité et d'identification fixés visuellement : en comparant leurs valeurs et le résultat (valide ou non) de l'identification, nous avons fixé les limites (supérieure et inférieure respectivement pour l'entropie et la probabilité) au-delà desquelles il ne faut plus accorder crédit au résultat. Les seuils résultants sont fournis dans le tableau suivant :

Classes reconnues	limite supérieure de l'entropie		limite inférieure de la probabilité	
	méthode 1	méthode 2	méthode 1	méthode 2
Arbre	0.2	0.2	0.50	0.38
Terre	0.3	indéterminé	0.975	0.9967
Ciel	0.035	0.1	0.95	0.59
Roche	0.3	indéterminé	0.95	0.992
Herbe	0.15	0.034	0.6	0.95

TAB. V.3 – Limites sur l'entropie et la probabilité pour chaque classe

Plusieurs remarques peuvent être faites sur l'obtention de certains seuils d'entropie et de probabilité :

- Dans le cas de la méthode 1 :
 - Le seuil d'entropie pour la classe «terre» a été fixé arbitrairement à la valeur d'entropie la plus grande sur cette classe ; en effet, par rapport aux résultats que nous avons obtenus, les régions déclarées à tort de ce type ont généralement une entropie trop faible pour entraîner une remise en cause.
 - Dans le même ordre d'idée, la classe «arbre» peut avoir des seuils de probabilité faible. Nous l'avons fixé ainsi car nous ne nous sommes jamais trouvés dans le cas où une région était identifiée par erreur comme «arbre».
 - D'un autre côté, la variation de probabilité sur l'herbe est plus importante. Cependant, par rapport aux erreurs d'identification commises, il est possible d'accepter des seuils faibles. Ce n'est en revanche pas le cas des classes «terre» et «roche». En effet, ces deux dernières classes sont le plus souvent très bien reconnues avec des probabilités de l'ordre de 0.99. Une forte variation est alors synonyme d'erreur d'identification.

La figure V.11 montre le résultat de l'évaluation de l'identification et de la segmentation sur l'image de la figure V.4 étudiée précédemment. On peut noter qu'effectivement les régions problématiques sont détectées.

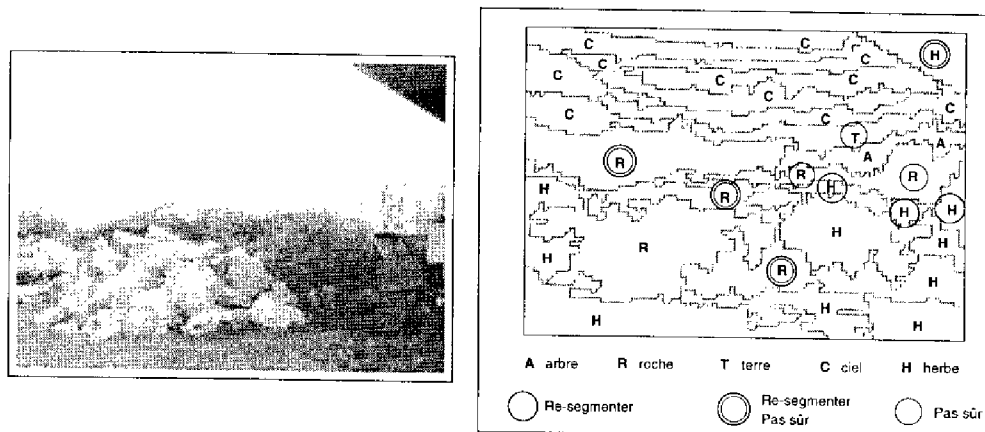


FIG. V.11 - Evaluation de la segmentation et de l'identification de l'image V.4

L'aide de ces critères permet de rejeter correctement les régions mal identifiées ou à re-segmenter. Cependant, si ces seuils sont valables sur nos images, ils devront être affinés en réalisant une nouvelle étude sur un plus grand nombre d'images.

Remarquons aussi que ces seuils ont été déterminés à partir de l'étude sur le vecteur d'attributs complet (neuf attributs). Le même type de raisonnement peut bien sûr être appliqué pour l'identification à partir du vecteur d'attributs fourni par l'AFD.

- Dans le cas de la méthode 2: le tableau V.3 fait apparaître que, si les seuils de probabilité peuvent être adaptés à cette méthode de segmentation, il n'en est pas de même pour l'entropie; en effet, l'intervalle de variation de l'entropie pour les régions à re-segmenter est inclu dans celui des autres régions (cas des classes «roche» et «terre»).

On se retrouve finalement devant deux possibilités non satisfaisantes: choisir le seuil pour segmenter toutes les régions qui le nécessitent (ce qui impliquera que des régions qui n'étaient pas à re-segmenter vont l'être) ou choisir de conserver les régions qui ne sont pas à re-segmenter (et dans ce cas les régions qui doivent l'être ne seront pas détectées).

Ce problème montre que le seuil d'entropie n'est pas suffisant pour détecter les régions à re-segmenter dans le cas de la méthode 2. La raison principale semble résider dans le fait que, pour ce type de segmentation ne s'intéressant pas aux détails, des régions considérées comme bien segmentées auront une entropie forte. Précédemment, ce cas n'arrivait pas, ou rarement. Un compromis était donc possible entre les régions de faible entropie à re-segmenter et les régions de forte entropie qui ne

devaient pas l'être; ce n'est plus le cas maintenant.

Pour pouvoir réaliser le même type de raisonnement que précédemment, il sera donc nécessaire de modifier le critère d'évaluation de la segmentation. Une possibilité serait que le critère ne dépende pas seulement de l'entropie mais aussi de la probabilité d'appartenance de la région à la classe.

Nous avons vu que les résultats obtenus en relançant le processus de segmentation sur des régions (issues de la méthode 1) ne sont pas intéressants: les régions obtenues n'ont pas des tailles assez grandes pour fournir une bonne identification de leur nature. En revanche, pour certaines régions issues du deuxième procédé de segmentation, il est possible d'affiner les frontières tout en conservant une reconnaissance correcte par la suite. Bien entendu, cela suppose une étude poussée pour déterminer efficacement les régions à re-segmenter.

Bien que nous n'ayons pas déterminé un critère adapté, nous pensons qu'il ne faut pas rejeter l'approche par coopération segmentation/identification. Elle ne doit seulement être envisagée que lorsque des régions très grandes doivent être re-segmentées (d'où nécessité d'utiliser la seconde technique de segmentation), et lorsque le contexte ne permet pas de lever les ambiguïtés dues à une mauvaise reconnaissance.

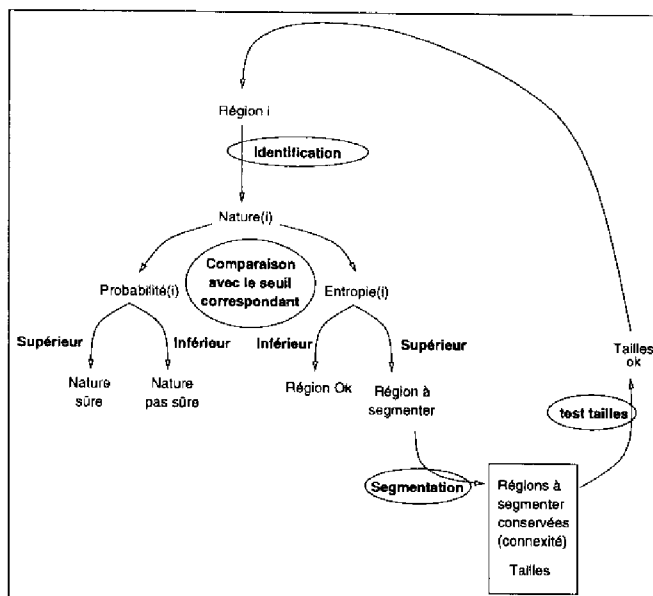


FIG. V.12 – Schéma de principe de fonctionnement du procédé d'évaluation

Le comportement général de l'algorithme devra suivre le schéma suivant (voir figure V.12) :

- ▷ calcul de l'entropie et de la probabilité sur chaque région,

▷ si $Entropie(C_i) \geq S_{entropie}(C_i)$ alors relancer la segmentation.

Bien entendu, la condition pour tenir compte du résultat de cette nouvelle segmentation sera un seuil sur la taille des régions obtenues ;

▷ si $Probabilite(C_i) \leq S_{probabilite}(C_i)$ alors prévenir que l'identification est peu sûre.

Finalement, à terme, nous serons capables de connaître, pour chaque région, sa nature et la confiance accordée à ce résultat. La construction du modèle final sera alors obtenue par fusion des différentes régions de même nature et dont le coefficient de confiance est suffisant.

V.5 Fusion des régions et construction du modèle

La construction du modèle nominatif de régions ne dépend plus à présent que de deux étapes :

1. la fusion des régions connexes de même classe,
2. la vérification de la cohérence du résultat.

Nous allons à présent détailler les méthodes employées et montrer le modèle obtenu suivant les deux méthodes de segmentation employées. Nous concluons finalement sur les suites possibles de ce travail.

V.5.1 Fusion des régions connexes

Chaque région de l'image est à ce stade assortie de sa nature et de la confiance accordée à cette identification. Mais le nombre de régions obtenues ne correspond pas aux grandes composantes de la scène. L'étape de fusion est donc nécessaire pour regrouper toutes les régions de même nature et connexes.

Cependant, ce critère de fusion n'est pas suffisant. Une région dont la nature n'est pas sûre ne doit pas être fusionnée. En effet, il peut s'agir d'une erreur de reconnaissance qui pourrait être détectée lors de la vérification de la cohérence.

Nous nous sommes donc limités pour la fusion de façon à ce que :

- les régions fusionnées soient adjacentes,
- les régions à fusionner sont de même nature.

Nous avons appliqué ce raisonnement sur les images issues des deux types de segmentation (méthodes 1 et 2). Les résultats de la fusion sur les images V.5 et V.6 qui ont servi d'exemple jusqu'ici, sont montrés respectivement sur les figures V.13 et V.14.

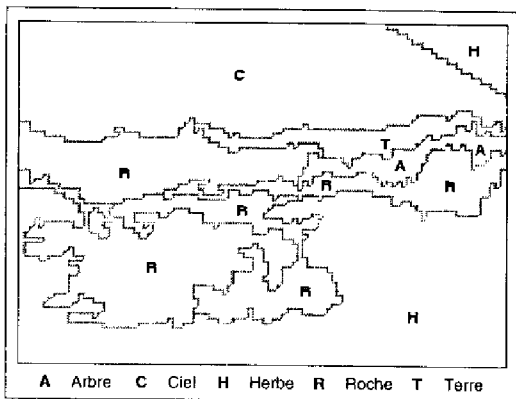


FIG. V.13 - Fusion des régions de l'image V.5

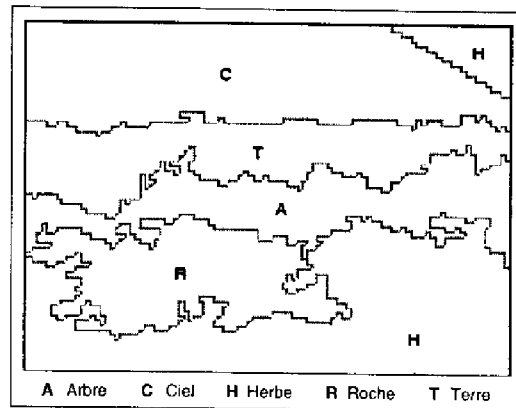


FIG. V.14 - Fusion des régions de l'image V.6

Sur l'image V.13, les régions peu sûres comme les régions de roche n'ont pas été fusionnées. Par contre les régions de ciel et d'herbe sont bien regroupées. De même sur la figure V.14, toutes les régions de même nature sont fusionnées.

Notons que les seuils sont restés les mêmes pour les deux types de segmentation et correspondent à ceux obtenus à partir de la première technique employée (méthode 1).

Toutefois, un certain nombre d'erreurs subsistent. Ces erreurs dues à une mauvaise identification ne peuvent pas toujours être éliminées, comme par exemple les régions reconnues comme étant de la roche dans l'image V.13. En revanche, certaines d'entre elles peuvent être détectées en vérifiant la cohérence du résultat global sur l'image et le contexte : par exemple les régions dont la nature est dite «terre» sur les deux images sont mal identifiées mais la détection de l'erreur est possible (elles sont comprises entre du ciel et de l'arbre).

V.5.2 Vérification de la cohérence et construction du modèle

La cohérence de l'image peut être étudiée de manière simple en utilisant :

- d'une part la position relative des régions : les unes par rapport aux autres ou par rapport au bord de l'image,
- d'autre part le contexte applicatif : l'environnement dans lequel évolue le robot a ses spécificités.

Les principales erreurs détectables dans une image sont présentées sur le schéma de la figure V.15.

Bien entendu, les remarques qui vont suivre se limitent à notre contexte (le ciel et le sol doivent obligatoirement apparaître sur les images et le sol peut être en pente douce). Si nous traitons des images en pleine montagne, ceci ne serait plus justifié.

Les régions entourées sur ce schéma ont été affectées à tort à une classe :

1. une région de «terre» ou «roche» ne peut pas être située entre du ciel et de l'arbre ; leur nature doit alors être remplacée respectivement par «ciel» et «arbre» ;

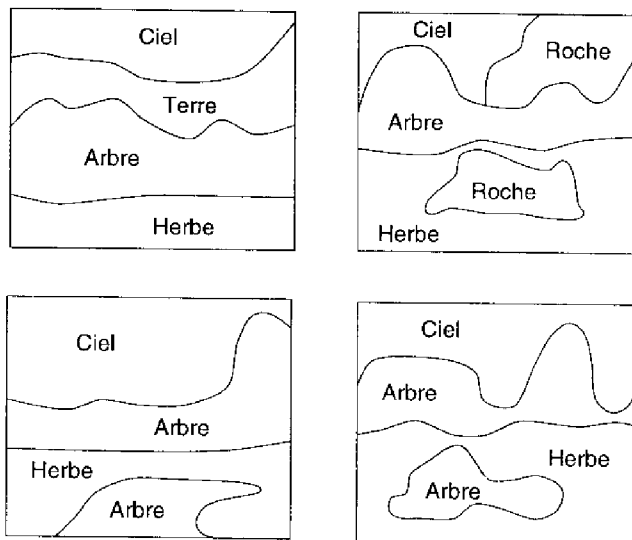


FIG. V.15 – Principales erreurs détectables par le contexte et la position dans l'image

2. une région «arbre» ne peut pas être entourée d'herbe ou toucher le bord inférieur de l'image ; elle sera considérée finalement comme de l'«herbe».

La détection de ces erreurs nécessite de connaître uniquement des relations du type *au-dessus de* ou *au-dessous de*. Ces relations sont obtenues en observant la position horizontale minimum et maximum des régions dans l'image.

Le schéma V.16 montre le raisonnement employé pour les déterminer.

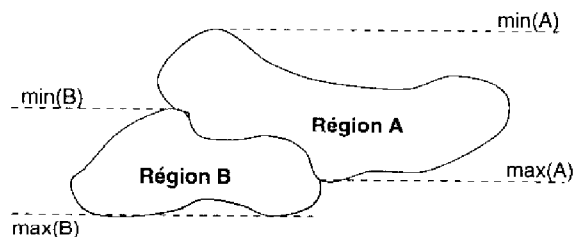


FIG. V.16 · Principe du relevé de la position d'une région

Soient $\min(R)$ et $\max(R)$ les positions minimales et maximales de la région R .

- La région A sera dite *au-dessus de* la région B si :

$$\min(A) > \min(B) \text{ et } \max(A) > \max(B)$$

- Réciproquement, la région B sera dite *au-dessous de* la région A si :

$$\min(B) < \min(A) \text{ et } \max(B) < \max(A)$$

Finalement, la nature de chaque région pourra être validée si elle vérifie les conditions suivantes :

- elle ne se trouve pas dans un des cas cités précédemment,
- et que les régions au-dessus et au-dessous sont sûres.

Toute modification dans la nature d'une région entraînera la remise en cause des régions qui l'entourent. Le processus s'arrête une fois qu'il n'y a plus d'erreurs détectées.

Les figures V.17 et V.18 montrent le résultat final obtenu à partir des deux techniques de segmentation.

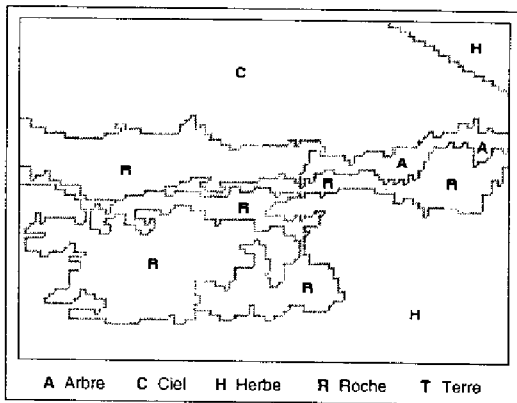


FIG. V.17 - Fusion des régions de l'image V.13

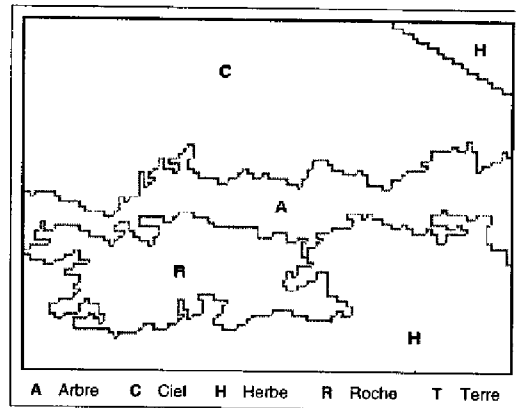


FIG. V.18 - Fusion des régions de l'image V.14

Comparons les performances de l'identification sur cette scène traitée par les deux segmentations (méthodes 1 et 2).

Dans le cas de la figure V.17 il reste des erreurs indétectables: la situation n'est pas invraisemblable et le contexte est insuffisant pour relever les erreurs commises dans la reconnaissance des arbres.

En revanche, l'image V.18 montre que le modèle construit est valide et correspond bien à la

réalité. Les erreurs ont été détectées et corrigées par le contexte (sauf bien sûr la région au coin supérieur gauche de l'image, mais qui ne correspond pas à une classe apprise puisqu'il s'agit du panneau solaire).

D'autres relations du même type doivent être mises en œuvre telles que *à côté de*, *autour de* ... Ces relations sont très intéressantes car elles sont très simples d'utilisation et apportent des informations très importantes pour identifier les erreurs. La forme des régions varie en effet énormément et les relations présentées ne suffisent pas toujours. En particulier, lorsque les régions sont grandes, la relation *autour de* s'avère utile.

La figure V.20 montre le type d'erreurs que permettrait de détecter la relation *autour de*. Les régions «roche» (cerclees de gris) sont entourées par une grande région étiquetée «arbre». On peut considérer qu'une région «roche» ne peut être en grande partie incluse dans une région «arbre» car elle doit en effet être en contact avec le sol (au moins la partie inférieure).



FIG. V.19 – Image intensité d'une scène

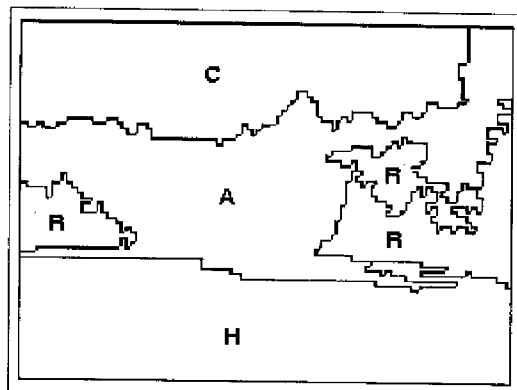


FIG. V.20 – Résultat pour l'image V.19

D'autres techniques utilisant le centre de gravité de la région, ou des pseudo-centres de gravité (liés à la forme de la région) ont été envisagées. Cependant, la variété des formes des régions rend la tâche difficile : ces centres ne se trouvaient pas obligatoirement à l'intérieur de la région. Toute relation de position par rapport à ceux-ci pouvait s'avérer fautive dans ces cas-là. C'est pourquoi nous préférons ce type de relations, très simples mais efficaces. Nous sommes persuadés qu'elles sont suffisantes pour résoudre la plupart des cas rencontrés.

La figure V.22 montre le modèle nominatif de régions obtenu pour l'image IV.27 présentée en page 66. Le modèle fourni est cohérent : on retrouve bien les régions principales de la scène. La région détectée «terre» correspond dans la réalité une zone où l'herbe est peu dense et où la terre apparaît.

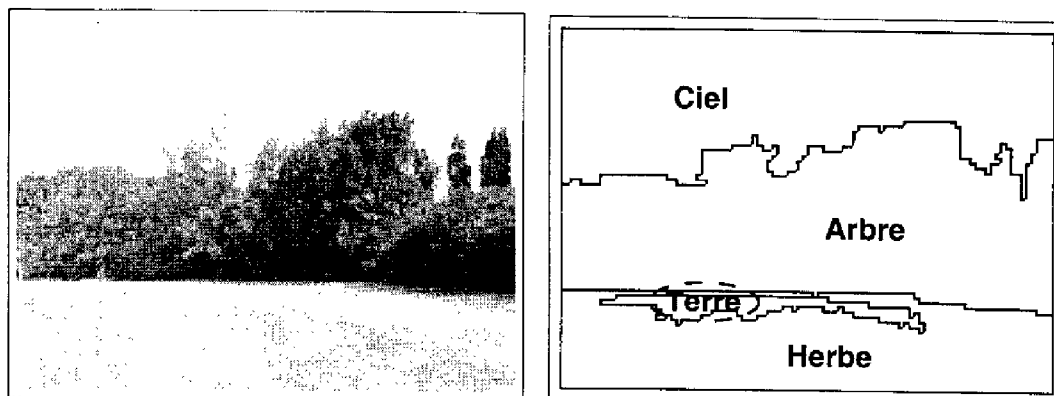


FIG. V.21 – Image intensité de l'image IV.27 FIG. V.22 – Résultat pour l'image V.21

Outre ces relations, les différents seuils déterminés (entropie et probabilité) permettraient certainement de définir s'il est utile ou non de relancer le processus d'identification en utilisant uniquement certains attributs. L'ensemble du processus gagnerait ainsi en fiabilité. L'apport d'autres informations comme la profondeur, lorsqu'elle est connue serait aussi d'une aide précieuse pour l'identification.

V.6 Conclusion

Dans ce chapitre nous avons présenté la construction finale du modèle nominatif de régions. Ce modèle fournit une description très synthétique de la scène perçue et affecte à chaque région la nature qui lui correspond.

A partir d'une technique de classification classique, et d'attributs de texture et de couleur, chaque région de l'image segmentée est identifiée.

Nous avons éprouvé l'algorithme sur des régions de tailles différentes et de nature très variée. Les résultats montrent qu'il est préférable de travailler sur des régions assez grandes pour limiter les erreurs lors de l'étape de reconnaissance.

Certaines erreurs d'identification ont été détectées en utilisant des règles simples basées sur le contexte et la position relative des régions les unes par rapport aux autres.

La plus grosse difficulté réside dans la reconnaissance des régions «arbre». Une seule classe est insuffisante pour représenter l'ensemble de ces éléments.

D'autre part, l'information tridimensionnelle devrait être prise en compte lors de la caractérisation. Elle permettrait d'introduire une notion de distance indispensable pour caractériser au mieux la texture d'un objet.

La deuxième partie du travail a consisté à étudier la possibilité de faire coopérer segmentation et identification pour améliorer la qualité de la représentation fournie. L'évaluation de ces deux procédés a été mise en œuvre. En revanche, la coopération de ces deux

derniers est limitée par la taille des régions obtenues par la segmentation.

Notons enfin que l'ensemble des modèles géométriques, topologiques et nominatif permettra d'interpréter la scène et de générer si nécessaire un modèle sémantique de l'environnement.

Conclusion générale

La perception de l'environnement pour un robot mobile en environnement naturel reste un domaine ouvert à ce jour. Si l'on est à même de faire déplacer un robot mobile dans de tels environnements, celui-ci est encore loin d'avoir acquis une réelle autonomie. Les problèmes liés à l'environnement naturel sont nombreux et variés : variété du type de sol, des objets, variation au cours du temps de l'environnement. . .

L'utilisation de la vision par ordinateur, bien que limitée à l'heure actuelle, doit avoir sa place dans le système de perception. Malgré la faible rapidité des traitements en imagerie, la richesse des informations fournies par ce capteur vidéo, fait de celui-ci un atout non négligeable pour améliorer la connaissance de l'environnement. En outre, les progrès techniques évoluent très vite en électronique et en informatique et les coûts calculatoires associés aux traitements devraient diminuer dans les années à venir.

Au cours de ce manuscrit, nous avons tout d'abord fait ressortir les limites actuelles de la modélisation de l'environnement et montré l'intérêt du capteur caméra vidéo. Pour acquérir une autonomie décisionnelle suffisante, deux besoins essentiels ont été énumérés :

1. augmenter la rapidité d'exécution des actions en diminuant le temps d'acquisition des données en particulier,
2. améliorer la connaissance de l'environnement par des informations autres que géométriques ou topologiques.

Nous avons répondu à ces deux problèmes en proposant, en premier lieu, un algorithme de stéréovision par corrélation. Celui-ci fournit une carte dense de points 3D en un temps inférieur à la caméra laser placée sur ADAM. Nous avons montré que l'un des critères exploités dans la phase de corrélation peut être calculé en une seule passe, réduisant ainsi de moitié le temps de calcul nécessaire. De plus, le repère rectifié que nous avons choisi permet de minimiser les déformations des images rectifiées afin d'améliorer l'appariement. Conserver ce repère lors de la phase de reconstruction 3D diminue aussi le temps de calcul.

La deuxième contribution, plus originale, concerne la création d'un nouveau modèle de l'environnement : le modèle nominatif de régions. A l'aide d'informations telles que la couleur et la texture, nous associons à chaque région de l'image, sa nature. Cette information est nécessaire aussi bien pour la navigation du robot que pour caractériser les

différents objets contenus dans la scène. C'est pourquoi nous avons tenu compte de cette spécificité en rendant le modèle « malléable » dans le sens où les régions peuvent provenir de différentes segmentations (3D ou couleur).

L'étape de segmentation couleur est en elle-même intéressante car elle est prévue pour s'effectuer sur une partie ou sur la totalité de l'image sans modifier le résultat sur la partie commune. Un autre intérêt est que nous pouvons dégager uniquement les grandes régions de la scène. Associer à ces grandes régions leur nature permet d'envisager par la suite d'avoir une connaissance approximative de ce qui se trouve au loin dans la scène. Ceci ne peut être apporté que par le capteur vidéo.

Bien entendu, ces travaux ne sont que les bases d'un système de vision qui pourrait être beaucoup plus complexe. Il pourrait déjà être amélioré de deux façons :

- au niveau de l'identification, une première étape serait de compléter la base de connaissances, trop limitée actuellement, en particulier pour la classe « arbre ». Il est préférable d'envisager plusieurs classes d'arbres au lieu d'une seule comme nous l'avons fait. De plus, l'introduction de l'information de profondeur lorsqu'elle existe devrait aussi permettre une meilleure identification de la nature des régions ;
- au niveau du modèle : dans le même ordre d'idée, il serait aussi intéressant de l'améliorer en fournissant des relations topologiques plus complètes entre les régions. On pourrait ainsi interpréter la scène et envisager la construction d'un modèle sémantique de la scène plus exploitable par le niveau décisionnel.

Notre objectif n'étant que la mise en place du modèle, nous avons préféré privilégier l'approche régions. Cependant, d'autres techniques de segmentation pourraient être employées. En particulier, les zones éloignées de l'image pourraient être traitées à partir d'une image de ligne de contraste. L'intérêt de cette méthode serait de définir de manière plus précise la ligne d'horizon. Celle-ci pourrait être utilisée pour détecter des amers lointains (un arbre qui se dégagerait à l'horizon...).

D'autre part, il faut noter que segmenter l'image entière n'est nécessaire que dans la phase d'initialisation du modèle. Par la suite, des techniques de suivi de contours [Delagnes 94] pourraient suffire pour compléter le modèle.

Nous sommes en mesure de fournir l'information tridimensionnelle lorsqu'elle existe par stéréocorrélation, mais il serait aussi possible de la déterminer pour les régions plus lointaines : si la stéréocorrélation est limitée par l'homogénéité de l'image pour des objets situés au-delà d'une douzaine de mètres, d'autres techniques (stéréovision basée sur des primitives de haut niveau telles que les contours et les régions) peuvent prendre le relais sur des distances supérieures. Ces techniques pourraient permettre de déterminer l'information de profondeur tant que la disparité pourra être calculée. L'identification de la nature des régions pourrait, dans ce cas, guider l'appariement.

Ce modèle ouvre aussi des perspectives intéressantes pour le suivi d'amers. D'une part, les régions étant mieux caractérisées, il sera possible de dégager les amers les plus

discriminants dans la scène. D'autre part, le contour de la région pourra servir de contour initial au suivi. De même, le ciel étant identifié, il sera possible de dégager rapidement la zone correspondant à l'horizon et de déterminer plus précisément où il se trouve. La détection des amers lointains sur la ligne d'horizon devrait ainsi être accélérée.

Il faudrait enfin étudier l'aspect incrémental de la construction du modèle, aspect indispensable lors de déplacement du robot. En particulier, il faudra réfléchir aux informations indispensables à conserver au cours du temps, à la manière de les intégrer pour limiter la place mémoire utilisée. . .

Bien entendu, les temps de calcul actuels ne sont pas satisfaisants. L'utilisation de cartes câblées ou l'étude de la parallélisation des algorithmes devront être envisagées pour accélérer les procédés. Cette étape sera cruciale pour que la vision prenne toute son ampleur et son intérêt dans la perception de l'environnement du robot.



Annexe A

Segmentation d'images intensité

Sur des scènes naturelles comme celles que nous avons à traiter, l'information de luminosité varie beaucoup. Il paraît donc difficile *a priori* de segmenter l'image uniquement avec cet attribut. Nous avons cependant vérifié cette idée en segmentant différentes images issues de scènes d'extérieur et en utilisant les deux méthodes de segmentation présentées.

Les résultats sont peu probants quelle que soit la méthode employée : avec la méthode 1 les images sont parfois bien segmentées comme dans le cas des images A.1 et A.3, même si le nombre de régions est important ; mais il arrive que l'intensité ne suffise pas comme le montre la figure A.5 (l'herbe et les arbres en fond sont fusionnés).

Avec la deuxième méthode (méthode 2), où l'on peut espérer détecter les meilleurs seuils, le problème s'amplifie (voir figures A.2, A.4 et A.6) : les valeurs de partitionnement de l'histogramme peuvent énormément varier d'une image à l'autre. Il faut parfois travailler pratiquement sur l'histogramme initial comme dans le cas de l'image A.2 pour obtenir, malgré tout, une segmentation insuffisante. Là aussi, la séparation entre terre et arbres éloignés devient indétectable sur l'intensité seule. Les variations d'intensité sont trop faibles pour que la méthode puisse correctement fonctionner.

La comparaison entre ces résultats et ceux obtenus par la segmentation couleur permet de justifier notre choix. Certes, la segmentation couleur est loin d'être parfaite mais les grandes régions sont bien détectées même si les limites sont à affiner.



FIG. A.1 - Image 2 (méthode 1)

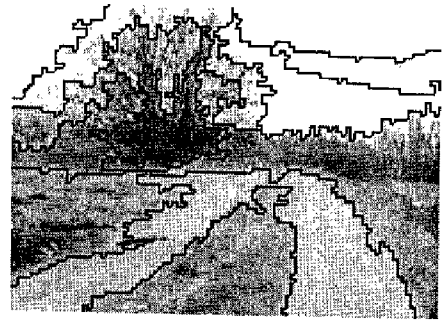


FIG. A.2 - Image 2 (méthode 2)

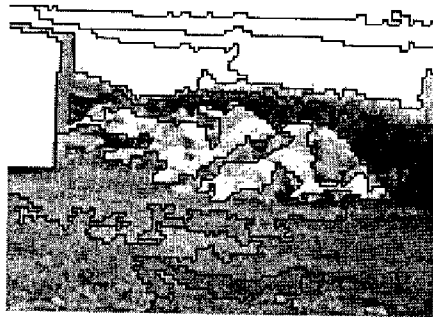


FIG. A.3 - Image 5 (méthode 1)



FIG. A.4 - Image 5 (méthode 2)



FIG. A.5 - Image 7 (méthode 1)



FIG. A.6 - Image 7 (méthode 2)

Annexe B

Etude sur les attributs de texture

Pour avoir une idée de l'intérêt de chacun des attributs, nous les avons comparés deux à deux. Les différents graphiques présentés (figures B.1 à B.15) montrent la répartition des différentes classes selon les différentes possibilités. Certains de ces attributs sont corrélés mais il est difficile d'évaluer a priori l'intérêt de supprimer l'un ou l'autre.

Nous avons regroupé dans le tableau suivant les résultats de chacun de ces graphes.

	Arbre	Roche	Terre	Ciel	Herbe
Densité/Granularité	NS	NS	PS	S	NS
Densité/Discontinuité	MS	MS	PS	S	MS
Densité/Courbure	MS	S	PS	S	MS
Densité/Rectitude	MS	NS	PS	S	NS
Densité/Netteté	MS	S	PS	S	MS
Granularité/Discontinuité	MS	MS	PS	S	MS
Granularité/Courbure	MS	S	PS	S	MS
Granularité/Rectitude	MS	NS	PS	S	NS
Granularité/Netteté	MS	S	PS	S	MS
Discontinuité/Courbure	MS	S	PS	S	MS
Discontinuité/Rectitude	NS	NS	PS	S	NS
Discontinuité/Netteté	MS	S	PS	S	MS
Courbure/Rectitude	MS	S	PS	S	MS
Courbure/Netteté	MS	S	PS	S	MS
Rectitude/Netteté	MS	S	PS	S	MS

TAB. B.1 – Dispersion des classes selon les attributs de texture

Pour chaque classe dans chaque graphe, nous avons noté lesquelles étaient, bien séparées (S), presque séparées (PS) - *quelques échantillons sont dispersés* -, mal séparées (MS) - *il est possible de trouver un axe correspondant à une combinaison linéaire des deux axes* - ou non séparées (NS) - *les classes sont mélangées*.

On peut immédiatement remarquer que seule la classe «ciel» est séparée quel que soit l'attribut utilisé. Pour les autres classes, il est nécessaire de faire coopérer plusieurs attributs. Les classes «roche» et «terre» peuvent être bien séparées. En revanche, les classes «arbre» et «herbe» le sont généralement mal. Il est donc intéressant d'exploiter l'ensemble des attributs pour séparer au maximum les différentes classes.

Notons toutefois qu'une analyse plus détaillée devrait être effectuée, en utilisant l'analyse en composantes principales, pour déterminer l'apport de chaque attribut.

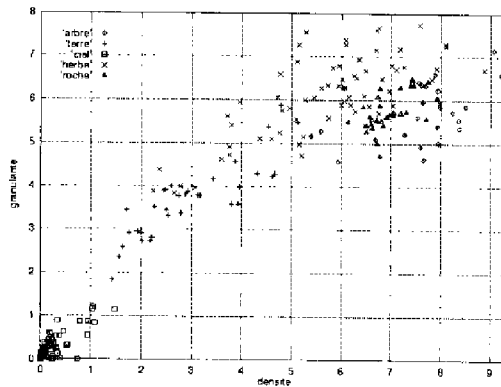


FIG. B.1 - Représentation des classes selon densité et granularité

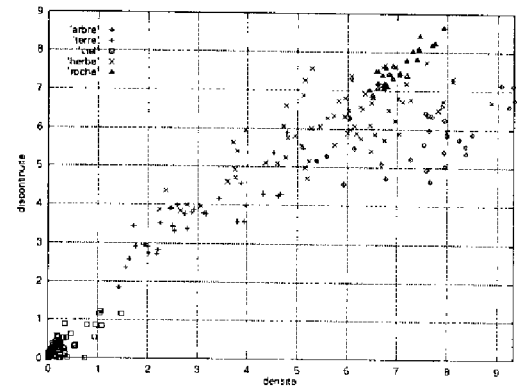


FIG. B.2 - Représentation des classes selon densité et discontinuité

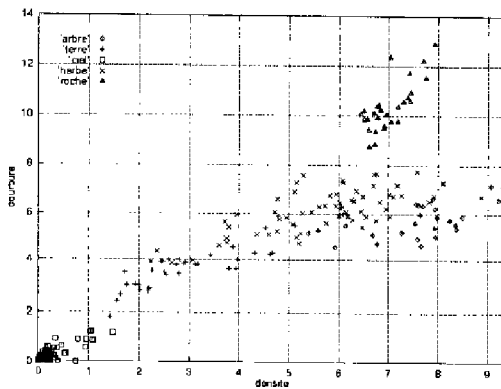


FIG. B.3 - Représentation des classes selon densité et courbure

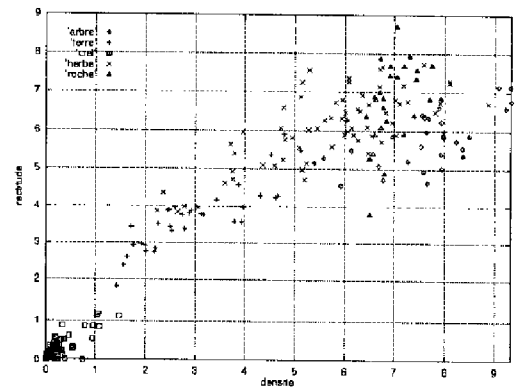


FIG. B.4 - Représentation des classes selon densité et rectitude

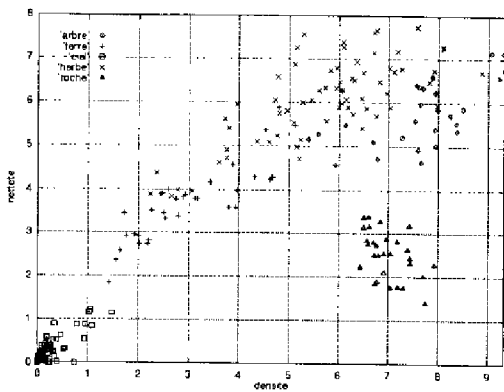


FIG. B.5 – Représentation des classes selon densité et netteté

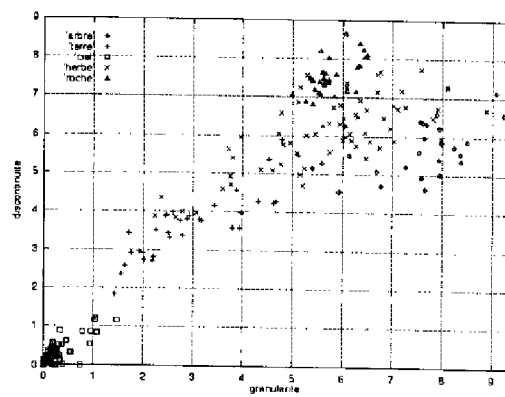


FIG. B.6 – Représentation des classes selon granularité et discontinuité

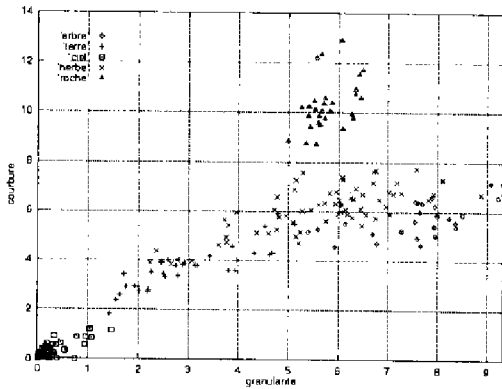


FIG. B.7 – Représentation des classes selon granularité et courbure

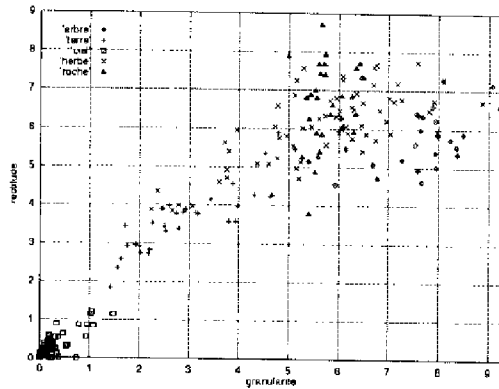


FIG. B.8 – Représentation des classes selon granularité et rectitude

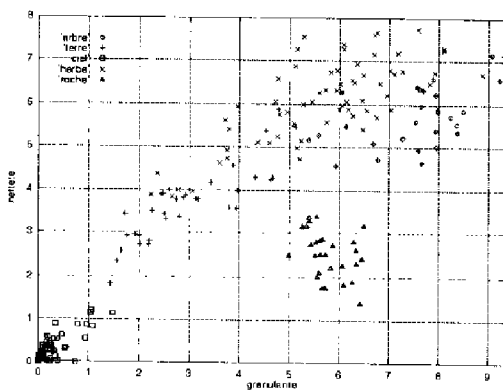


FIG. B.9 – Représentation des classes selon granularité et netteté

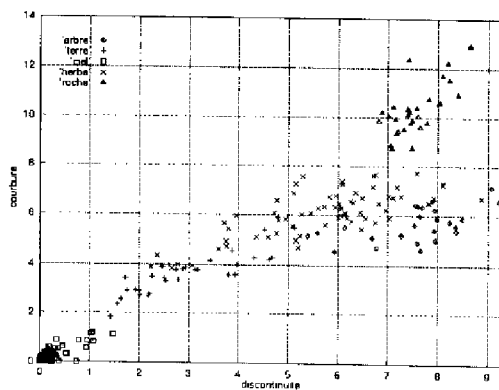


FIG. B.10 – Représentation des classes selon discontinuité et courbure

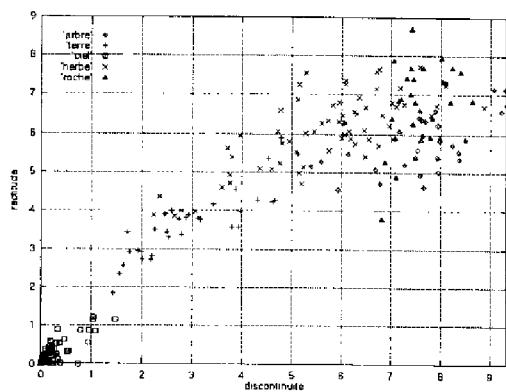


FIG. B.11 – Représentation des classes selon discontinuité et rectitude

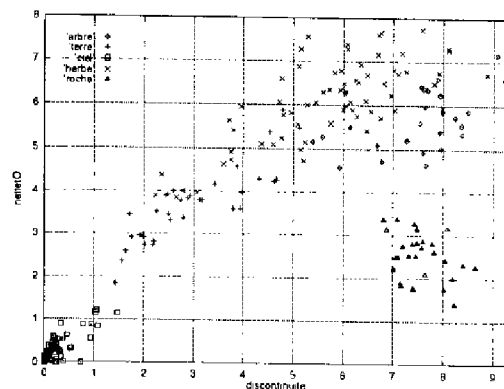


FIG. B.12 – Représentation des classes selon discontinuité et netteté

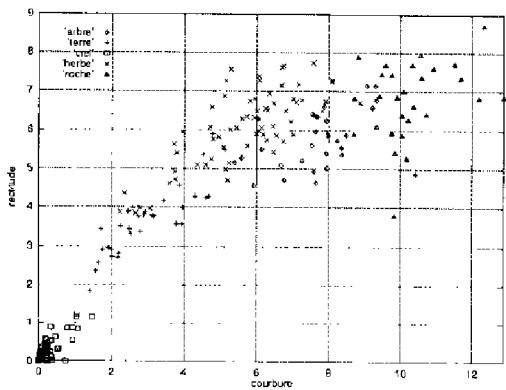


FIG. B.13 – Représentation des classes selon courbure et rectitude

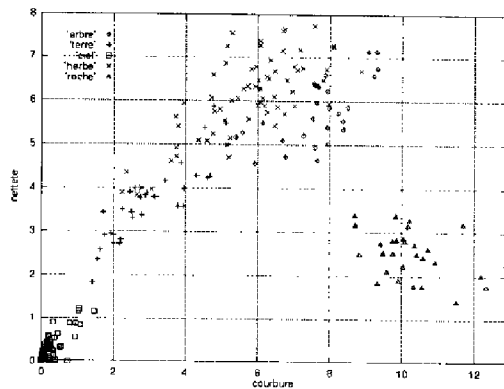


FIG. B.14 – Représentation des classes selon courbure et netteté

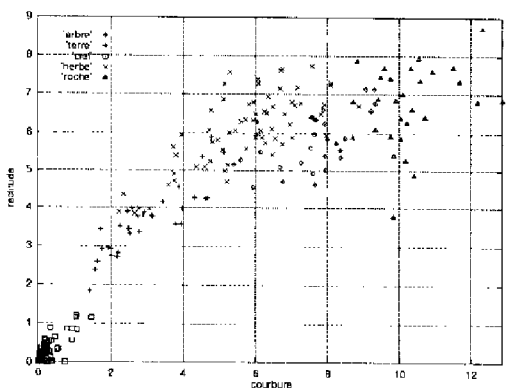


FIG. B.15 – Représentation des classes selon courbure et rectitude

Références bibliographiques

- [Andrews 72] H.C. Andrews. *Mathematical techniques in pattern recognition*. Wiley-Interscience, 1972.
- [Asada 88] M. Asada. *Building A 3-D World Model For A Mobile Robot From Sensory Data*. I.E.E.E Transactions on Robotics and Automation, Philadelphia (USA), Avril 1988.
- [Asada 90] M. Asada. *Map Building for a Mobile Robot from Sensory Data*. I.E.E.E Transactions on Systems, Man, and Cybernetics, vol. 37, no. 6, pages 1326–1336, 1990.
- [Ayache 88] N. Ayache & C. Hansen. *Rectification of images for binocular and trinocular stereovision*. In 9th IEEE International Conference on Pattern Recognition, Rome (Italie), pages 11–16, Novembre 1988.
- [Ayache 89] N. Ayache. *Vision stéréoscopique et perception multisensorielle. Applications à la robotique mobile*. InterEditions, Paris (France), 1989.
- [Baghdassarian 95] C. Baghdassarian. *Classification chromatique pour la reconnaissance d'objets: application aux scènes routières*. Thèse de troisième cycle, Ecole des Mines de Paris, Octobre 1995.
- [Baron 91] T. Baron. *De la perception à la modélisation par la vision achrome et trichrome en robotique*. Thèse de l'Université Paul Sabatier, Toulouse (France), Laboratoire d'Automatique et d'Analyse des Systèmes (C.N.R.S.), Décembre 1991.
- [Betge-Brezetz 95] S. Betge-Brezetz, R. Chatila & M. Devy. *Object-based Modelling and Localization in Natural Environment*. In I.E.E.E International Conference on Robotics and Automation, Nagoya (Japon), 1995.
- [Betge-Brezetz 96] S. Betge-Brezetz. *Modélisation incrémentale et localisation par amers pour la navigation d'un robot mobile autonome en environnement naturel*. Thèse de l'Université Paul Sabatier, Toulouse (France), Laboratoire d'Analyse et d'Architecture des Systèmes (C.N.R.S.), Février 1996.
- [Boissier 91] L. Boissier & G. Giralt. *Autonomous Planetary Rover (VAP): the robotics concepts*. In I.A.R.P. 91 on Robotics in Space, Pisa (Italie), 1991.

- [Boissier 92] L. Boissier, B. Hotz, C. Proy, O. Faugeras & P. Fua. *Autonomous Planetary Rover (V.A.P.): On-Board Perception System Concept and Stereovision by Correlation Approach*. In I.E.E.E International Conference on Robotics and Automation, Nice (France), pages 181–186, 1992.
- [Bracco 92] C. Bracco. *Stéréovision par corrélation pour le MARSOKHOD*. Rapport de stage de 3ème cycle D.E.S.S Automatique Industrielle et Robotique (Nice-Sophia Antipolis), Centre National d'Etudes Spatiales, 1992.
- [Briot 77] M. Briot. *La stéréognosie en robotique - Application au tri de solides*. Thèse d'état, Laboratoire d'Automatique et d'Analyse des Systèmes (C.N.R.S.), 1977.
- [Brodatz 66] P. Brodatz. *Textures: A photographic album for artists and designers*. New York: Dover Publications, 1966.
- [Casals 92] A. Casals, J. Amat & A. Grau. *Texture Parametrization Method for Image Segmentation*. In Second European Conference on Computer Vision, Santa Margherita Ligure (Italie), 1992.
- [Cattoen 85] M. Cattoen. *Conception et réalisation d'une machine de vision achrome et trichrome pour la robotique*. Thèse d'état, Institut National Polytechnique de Toulouse, Février 1985.
- [Charroux 96] B. Charroux, S. Philipp & J.P. Cocquerez. *Système de vision mettant en œuvre une coopération d'opérateurs de segmentation guidée par l'interprétation*. In 10ème Congrès Reconnaissance des Formes et Intelligence Artificielle (RFIA'96), Rennes (France), 1996.
- [Chatila 93] R. Chatila, R. Alami, S. Lacroix, J. Perret & C. Proust. *Planet exploration by robots: from mission planning to autonomous navigation*. In '93 International Conference on Advanced Robotics (ICAR), Tokyo (Japon), Novembre 1993.
- [Chaudhuri 95] B.B. Chaudhuri & N. Sarkar. *Texture Segmentation Using Fractal Dimension*. I.E.E.E Transactions on Pattern Analysis and Machine Intelligence, vol. 17, no. 1, pages 72–77, Janvier 1995.
- [Coutance 91] V. Coutance. *La couleur en vision par ordinateur - Application à la robotique*. Thèse de l'Université Paul Sabatier, Toulouse (France), Laboratoire d'Automatique et d'Analyse des Systèmes (C.N.R.S.), Janvier 1991.
- [Dacre-Wright 93] B. Dacre-Wright. *Planification de trajectoire pour un robot mobile sur terrain accidenté*. Thèse de l'Université Paul Sabatier, Toulouse (France), Laboratoire d'Analyse et d'Architecture des Systèmes (C.N.R.S.), Octobre 1993.
- [Daily 88] M. Daily, J. Harris, D. Keirse, K. Olin, D. Payton, K. Reiser, J. Rosenblatt, D. Tseng & V. Wong. *Autonomous cross-country navigation with*

- the ALV*. I.E.E.E Transactions on Robotics and Automation, Philadelphie (USA), Avril 1988.
- [Delagnes 94] P. Delagnes, J. Benois & D. Barba. "Adjustable Polygons": A Novel Active Contour Model for Objects Tracking on Complex Background. *Journal on Communications*, vol. 45, 1994.
- [Delagnes 96] P. Delagnes & D. Barba. *Détection et extraction de défauts dans les images texturées par modélisation markovienne définie sur des graphes irréguliers - Application à l'analyse d'images de chaussées*. In 10ème Congrès Reconnaissance des Formes et Intelligence Artificielle (RFIA'96), Rennes (France), 1996.
- [Derras 94a] M. Derras, M. Berducat, P. Bonton, J. Gallice & S. Naudet. *Segmentation Texturale Originale. Analyse et Parallélisation Temps Réel*. 9ème Congrès AFCET Reconnaissance des Formes et Intelligence Artificielle, Paris (France), 1994.
- [Derras 94b] M. Derras, C. Debain, M. Berducat, P. Bonton & J. Gallice. *Unsupervised Regions Segmentation: Real Time Control of an Unkeep Machine of Natural Spaces*. Third European Conference on Computer Vision, Stockholm (Suède), vol. 2, pages 207-212, 1994.
- [Duda 73] R.O. Duda & P.E. Hart. *Pattern Classification and Scene Analysis*. Wiley & sons, 1973.
- [Dunn 94] D. Dunn, W.E. Higgins & J. Wakeley. *Texture Segmentation Using 2-D Gabor Elementary Functions*. I.E.E.E Transactions on Pattern Analysis and Machine Intelligence, vol. 16, no. 2, pages 130-149, Février 1994.
- [Eastman 87] R.D. Eastman & M. Waxman. *Using Disparity Functionals for Stereo Correspondance and Surface Reconstruction*. *Computer Vision, Graphics, and Image Processing*, 1987.
- [Fillatreau 93] P. Fillatreau, M. Devy & R. Prajoux. *Modelling of Unstructured Terrain and Feature Extraction using B-spline Surfaces*. In '93 International Conference on Advanced Robotics (ICAR), Tokyo (Japon), Novembre 1993.
- [Fleury 96] S. Fleury. *Architecture de contrôle distribuée pour robot mobile autonome : principes, conception et applications*. Thèse de l'Université Paul Sabatier, Toulouse (France), Laboratoire d'Analyse et d'Architecture des Systèmes (C.N.R.S.), Février 1996.
- [Fouchet 94] X. Fouchet. *Module de vision pour la modélisation d'un environnement naturel*. Rapport de stage de D.E.A Signaux et Images (E.N.S.E.E.I.H.T), Laboratoire d'Analyse et d'Architecture des Systèmes (C.N.R.S.), 1994.
- [Fua 93] P. Fua. *A parallel stereo algorithm that produces dense depth maps and preserves image features*. *Machine Vision and Applications*, vol. 6, no. 1, pages 35-49, 1993.

- [Garric 96] V. Garric. *Vision pour la robotique de manipulation : calibration, localisation et saisie d'objets*. Thèse de docteur ingénieur de l'Institut National des Sciences Appliquées de Toulouse, Laboratoire d'Analyse et d'Architecture des Systèmes (C.N.R.S.), Mai 1996.
- [Gavin 94] A.S. Gavin. *Low computation vision-based navigation for a martian rover*. Conference on Intelligent Robotics in Field, Factory, Service and Space, Houston (Texas), 1994.
- [Giralt 92] G. Giralt & L. Boissier. *THE FRENCH PLANETARY ROVER VAP: Concept and Current Developments*. In IEEE International Workshop on Intelligent Robots and Systems (IROS '92), Raleigh (North Carolina, USA), pages 1391–1398, Juillet 1992.
- [Grandjean 93] P. Grandjean & L. Matthies. *Perception Control for Obstacle Detection by a Cross-country Rover*. I.E.E.E Transactions on Robotics and Automation, 1993.
- [Halme 93] A. Halme & K. Koskinen. *An Approach to Automate Mobility of Working Machines in Outdoor Environment - The PANORAMA Project*. I.E.E.E International Conference on Intelligent Robots and Systems, Yokohama (Japon), pages 598–603, Juillet 1993.
- [Hotz 91] B. Hotz. *Etude de techniques de stéréo vision par corrélation. Application au programme Véhicule Autonome Planétaire (V.A.P)*. Rapport de stage de 3ème année de l'Ecole Supérieure en Sciences Informatiques et du D.E.A de Robotique et Vision Artificielle (INRIA de Sophia Antipolis), Centre National d'Etudes Spatiale, 1991.
- [Kanade 94] T. Kanade & M. Okutomi. *A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment*. I.E.E.E Transactions on Pattern Analysis and Machine Intelligence, vol. 16, no. 9, Septembre 1994.
- [Kemurdjian 92] A. Kemurdjian, V. Gromov, V. Mishkinyuk, V. Kucherenko & P. Sologub. *Small Marsokhod Configuration*. I.E.E.E Transactions on Robotics and Automation, Nice (France), pages 165–168, Mai 1992.
- [Kemurdjian 95] A.L. Kemurdjian, E.K. Potiemkin & V.K. Mishkinyuk. *Planet rover today*. International Conference on Advanced Robotics, Sant Feliu de Guixols (Espagne), vol. 1, pages 293–299, Septembre 1995.
- [Krotkov 94] E. Krotkov & R. Hoffman. *Terrain Mapping for a Walking Planetary Rover*. I.E.E.E Transactions on Robotics and Automation, vol. 10, no. 6, pages 728–739, Décembre 1994.
- [Kweon 92] I.S. Kweon & T. Kanade. *High-Resolution Terrain Map from Multiple Sensor Data*. I.E.E.E Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pages 278–292, Février 1992.

- [Lacombe 91] J. L. Lacombe & T. Blais. *PROGRAMME AMR DE ROBOTS MOBILES AVANCES. Solutions retenues pour le développement des modules de perception et d'interface homme-machine du démonstrateur ADAM*. In 4th International Symposium on Offshore, Robotics and Artificial Intelligence (ORIA), Marseille (France), pages 223–230, Décembre 1991.
- [Lacroix 94] S. Lacroix, R. Chatila, S. Fleury, M. Herrb & T. Simeon. *Autonomous Navigation in Outdoor Environment: Adaptive Approach and Experiment*. In I.F.E.E International Conference on Robotics and Automation, San Diego, California (USA), 1994.
- [Lacroix 95] S. Lacroix. *Stratégies de perception et de déplacement pour la navigation d'un robot mobile autonome en environnement naturel*. Thèse de l'Université Paul Sabatier, Toulouse (France), Laboratoire d'Analyse et d'Architecture des Systèmes (C.N.R.S.), Avril 1995.
- [Langer 94] D. Langer, J.K. Rosenblatt & M. Hebert. *A Behavior-Based System for Off-Road Navigation*. I.F.E.E Transactions on Robotics and Automation, vol. 10, no. 6, pages 776–783, Décembre 1994.
- [Langer 95] D. Langer & C. Thorpe. *Range Sensor Based Outdoor Vehicle Navigation, Collision Avoidance and Parallel Parking*. Autonomous Robots, pages 147–161, 1995.
- [Lapierre 94] I. Lapierre & C. Laugeau. *Modélisation, Simulation et compréhension de scènes d'extérieur. Application aux scènes routières*. In 9ème Congrès AF-CET Reconnaissance des Formes et Intelligence Artificielle, Paris (France), 1994.
- [Lasserre 95] P. Lasserre & P. Grandjean. *Stereovision Improvements*. In International Conference on Advanced Robotics, San Felu de Guixol (Espagne), 1995.
- [Lasserre 96] P. Lasserre & M. Briot. *Segmentation couleur d'images de scènes naturelles pour la robotique mobile d'extérieur*. In Automatique - Génie informatique - Image, AGI'96, Tours (France), 1996.
- [Lee 92] T.S. Lee, D. Mumford & A. Yuille. *Texture Segmentation by Minimizing Vector-Valued Energy Functionals: The Coupled-Membrane Model*. In Second European Conference on Computer Vision, Santa Margherita Ligure (Italie), 1992.
- [Lin 91] X. Lin & S. Chen. *Color Image Segmentation Using Modified HSI System for Road Following*. In IEEE International Conference on Robotics and Automation, Sacramento (USA), Avril 1991.
- [Luo 94] R.C. Luo & H. Potlapalli. *Fractal Based Outdoor Landmark Recognition System for the Navigation of a Mobil Robot*. In I.E.E.E International Conference on Robotics and Automation, San Diego, California (USA), 1994.

- [Maître 92] H. Maître & W. Luo. *Using Models to Improve Stereo Reconstruction*. I.E.E.E Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pages 269-277, Février 1992.
- [Malafeew 94] E. Malafeew & W. Kaliardos. *The MITy micro-rover: sensing, control and operation*. Conference on Intelligent Robotics in Field, Factory, Service and Space, Houston (Texas), 1994.
- [Marapane 94] S.B. Marapane & M.M. Trivedi. *Multi-Primitive Hierarchical (MPH) Stereo Analysis*. I.E.E.E Transactions on Pattern Analysis and Machine Intelligence, vol. 16, no. 3, pages 227-240, Mars 1994.
- [Matthies 92] L. Matthies. *Stereo Vision for Planetary Rovers: Stochastic Modeling to Near Real-Time Implementation*. International Journal of Computer Vision, vol. 8, no. 1, pages 71-91, 1992.
- [Murrieta Cid 95] R. Murrieta Cid. *Vision pour la robotique mobile d'extérieur: identification de régions par des critères de texture*. Rapport de stage de D.E.A Signaux et Images (E.N.S.E.E.I.H.T), Laboratoire d'Analyse et d'Architecture des Systèmes (C.N.R.S.), 1995.
- [Nalwa 93] V.S. Nalwa. *A Guided Tour of Computer Vision*. Addison Wesley, 1993.
- [Nashashibi 93] F. Nashashibi. *Perception et Modélisation de l'Environnement Tridimensionnel pour la Navigation Autonome d'un Robot Mobile*. Thèse de l'Université Paul Sabatier, Toulouse (France), Laboratoire d'Automatique et d'Analyse des Systèmes (C.N.R.S.), Février 1993.
- [Ohta 85] Y. Ohta. *Knowledge-based interpretation of outdoor natural color scenes*. Morgan Kaufman Publishers, Inc., Palo Alto, CA, 1985.
- [Orteu 89] J.J. Orteu. *Segmentation d'image en régions par la méthode de Séparation-Fusion (Split and Merge)*. Rapport de recherche 89127, Laboratoire d'Automatique et d'Analyse des Systèmes (C.N.R.S.), Avril 1989.
- [Panjwani 95] D. Panjwani & G. Healy. *Markov Random Field Models for Unsupervised Segmentation of Textured Color Images*. I.E.E.E Transactions on Pattern Analysis and Machine Intelligence, vol. 17, no. 10, pages 939-954, Octobre 1995.
- [Parra 95] C. Parra. *Module de stéréovision pour des environnements naturels*. Rapport de stage de D.E.A Automatique et Informatique Industrielle, Laboratoire d'Analyse et d'Architecture des Systèmes (C.N.R.S.), 1995.
- [Postaire 87] J. G. Postaire. *De l'image à la décision - analyse des images numériques et théorie de la décision*. DUNOD informatique, 1987.
- [Pratt 91] W.K. Pratt. *Digital Image Processing (Second edition)*. J. Wiley and Sons, 1991.

- [Proy 93] C. Proy, M. Lamboley & T.N. Guyen. *A vision system to improve the autonomy of Marsokhod 96*. Workshop on Computer Vision for Space Applications, Antibes (France), pages 123–131, Septembre 1993.
- [Pujas 95] P. Pujas & M.J. Aldon. *Robust Colour Image Segmentation*. In 7th International Conference on Advanced Robotics, ICAR, San Feliu de Guixols (Espagne), 1995.
- [Robert de Saint Vincent 86] A. Robert de Saint Vincent. *Perception et modélisation de l'environnement d'un robot mobile: une approche stéréovision*. Thèse de l'Université Paul Sabatier, Toulouse (France), Laboratoire d'Analyse et d'Architecture des Systèmes (C.N.R.S.), Novembre 1986.
- [Simmons 92] R. Simmons, E. Krotov, W. Whittaker, B. Albrecht, J. Bares, C. Fedor, R. Hoffman, H. Pangels & D. Wettergreen. *Progress Towards Robotics Exploration of Extreme Terrain*. Journal of Applied Intelligence, pages 163–180, 1992.
- [Solder 93] U. Solder & V. Graefe. *Visual Detection of Distant Objects*. In I.E.E.E International Conference on Intelligent Robots and Systems, IROS'93, Yokohama (Japon), pages 1042–1049, Juillet 1993.
- [Stentz 95] A. Stentz & M. Hebert. *A Complet Navigation System for Goal Acquisition in Unknown Environments*. Autonomous Robots, pages 127–145, 1995.
- [Stone 93] H.W. Stone. *Design and control of the mesur/pathfinder microrover*. International Conference on Advanced Robotics, pages 263–270, 1993.
- [Thomas 94] T. Thomas & M. Cattoen. *Image Processing System for Detection of Weaving Defects*. Journal on Communications, vol. 45, 1994.
- [Thorpe 88] C. Thorpe, M. Herbert, T. Kanade & S.A. Shafer. *Vision and Navigation for the Carnegie-Mellon Navlab*. I.E.E.E Transactions on Pattern Analysis and Machine Intelligence, vol. 10, no. 3, pages 362–373, Mai 1988.
- [Thorpe 91a] C. Thorpe, M. Hebert, T. Kanade & S. Shafer. *Towards Autonomous Driving: The CMU Navlab. Part I: Perception*. I.E.E.E Expert, vol. 6, no. 4, pages 31–42, Août 1991.
- [Thorpe 91b] C. Thorpe, M. Hebert, T. Kanade & S. Shafer. *Towards Autonomous Driving: The CMU Navlab. Part II: Architecture and Systems*. I.E.E.E Expert, vol. 6, no. 4, pages 45–52, Août 1991.
- [Tomita 90] F. Tomita & S. Tsuji. *Computer Analysis of Visual Textures*. Kluwer Academic Publisher, 1990.
- [Vandapel 96] N. Vandapel. *Contrôle visuel du mouvement d'un robot mobile d'exploration planétaire*. Rapport de stage de fin d'études d'ingénieur (ENSEEIH), Laboratoire d'Analyse et d'Architecture des Systèmes (C.N.R.S.), 1996.

- [Wilcox 92] B. Wilcox, L. Matthies, D. Gennery, B. Cooper, T. Nguyen, T. Litwin, A. Mishkin & H. Stone. *Robotics Vehicles for Planetary Exploration*. I.E.E.E Transactions on Robotics and Automation, Nice (France), pages 175–180, Mai 1992.

Thèse de Patricia LASSERRE

Vision pour la robotique mobile en environnement naturel

Cette thèse se situe dans le cadre de l'augmentation des capacités de perception des robots mobiles destinés à évoluer en environnement naturel.

Les travaux présentés dans le mémoire portent sur le développement de l'utilisation du capteur caméra vidéo pour apporter des informations utiles à la localisation et à la navigation du robot dans son environnement. Les besoins actuels pour réaliser ces deux tâches ont permis de déterminer deux directions à développer. L'une concerne l'obtention de l'information tridimensionnelle par un système stéréoscopique ; l'autre consiste à enrichir la connaissance de l'environnement par identification de la nature des objets contenus dans la scène. La stéréovision est un moyen répandu pour obtenir l'information tridimensionnelle. Nous avons développé un algorithme de stéréocorrélation, technique qui est adaptée pour des scènes texturées. L'algorithme de base est amélioré pour prendre en compte à la fois les contraintes de rapidité d'exécution et de qualité de la corrélation.

Lorsque le robot évolue dans un environnement extérieur, la connaissance du terrain ou des objets situés dans cet environnement peut être améliorée en ajoutant des informations autres que géométriques ou topologiques, comme la couleur ou la texture. Nous avons proposé comme représentation complémentaire, le modèle nominatif de régions qui indique la nature de chaque région d'une image.

Tout d'abord, un algorithme de segmentation donne une description très synthétique de la scène : la méthode combine à la fois les techniques de grossissement de régions et de classification (basée sur l'étude de l'allure générale des histogrammes sur chaque composante colorimétrique).

Les régions issues de cette segmentation sont ensuite caractérisées puis identifiées afin d'obtenir leur nature (herbe, roche, terre ...). Leur caractérisation est obtenue à partir d'informations sur leur couleur et leur texture. Les opérateurs de texture développés sont basés sur l'étude des lignes de contraste dans la région considérée (densité, courbure, ...).

Une technique probabiliste est finalement utilisée pour déterminer la nature des éléments présents dans l'environnement.

Mots clés : *Robots Mobiles, Vision, Stéréovision, Segmentation, Couleur, Texture, Identification.*

Vision for mobile robots in natural environments

This thesis deals with the development of the use of a video camera in order to add useful information for both localization and navigation of a mobile robot roving in natural environments.

The current needs for the realization of these two tasks offers two directions. The first concerns the three-dimensional information obtaining using a stereoscopic system ; the second is the addition of knowledge about the environment by object nature identification.

Stereovision is a well-known method to obtain depth information. We have implemented a stereocorrelation algorithm, a well-adapted technique for textured scenes. The basic algorithm was improved so as to take into account both the execution speed constraints and the correlation quality.

When the robot executes a mission in an outdoor environment, the ground knowledge may be improved on adding information other than geometrical or topological, such as color and texture. We have proposed and implemented as a complementary representation, the nominally model of regions which indicates each region nature in an image.

Firstly, a segmentation algorithm provides a synthetic description of the scene : the method combines both region growing and clustering (based on general histograms shapes of each colorimetric component) techniques.

Regions issued from the segmentation stage are then characterized and afterwards identified in order to obtain their nature (grass, rocks, ground ...). Their characterization is obtained from their color and texture. The developed texture operators are based on contrast lines in the considered region (density, curve ...).

Finally, a probabilistic method is used to determine the nature of current elements in the environment.

Keywords : *Mobile Robots, Vision, Stereovision, Segmentation, Color, Texture, Identification.*