N° d'ordre : 220

50376

# THESE

#### présentée à

L'UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE

FLANDRES ARTOIS

pour l'obtention du titre de

DOCTEUR DE L'UNIVERSITE



50376

47

par

**Christophe VIEREN** 

# SEGMENTATION DE SCENES DYNAMIQUES EN TEMPS REEL. APPLICATION AU TRAITEMENT DE SEQUENCES D'IMAGES POUR LA SURVEILLANCE DE CARREFOURS ROUTIERS

# SOUTENUE LE 18 AVRIL 1988 DEVANT LA COMMISSION D'EXAMEN :

MM

VIDAL P. DAVID Y. POVY L. POSTAIRE J.G. BONNET P. MELIN C. BLOSSEVILLE J.M.

Président Rapporteur Rapporteur Directeur de Recherche Examinateur Examinateur Invité



# SOMMAIRE

# PRESENTATION GENERALE

# I- TRANSPORTS GUIDES ET VISION ARTIFICIELLE

I.1- AUTOMATISATION DES SYSTEMES DE TRANSPORTS GUIDES	I.1
I.1.1- Nécessité de l'automatisation	I.1
I.1.2- Marchés potentiels	I.2
I.1.3- Programme de recherche	I.2
1.2- DETECTION D'OBSTACLES ET VISION ARTIFICIELLE	I.3
I.3- LE TRAITEMENT D'IMAGES APPLIQUE AUX TRANSPORTS	I.4
I.3.1- Poursuite de cibles	I.5
I.3.1.1- Généralités	I.5
I.3.1.2- Etat de l'art	I.5
Gilbert	I.5
Nègre	I.6
Samy	I.7
Gambotto	I.7
I.3.1.3- Conclusion	I.8
I.3.2- Conduite automatique de véhicules autonomes	1.9
I.3.2.1- Généralités	I.9
I.3.2.2- Etat de l'art	I.9
Robot Rover	I.9
Tsugawa	I.10
Inigo	I.10
I.3.2.3-Conclusion	I.10
I.3.3- Contrôle du trafic routier	I.11
I.3.3.1- Généralités	I.11
I.3.3.2- Etat de l'art	I.12
Projet du JPL	I.12
Projet de l'UMIST	I.13
Projet du SRIT	I.13
Projet de l'UT	I.14
Système opérationnel d'OOYAMA-SHIGETA	I.14
Autres systèmes	I.15
I.3.3.3- Aspect temps réel	I.15
I.3.3.4- Conclusion	I.16
I.3.4- Conclusion : spécificité du projet	· I.16

I.4- CHOIX D'UNE STRATEGIE DE PRISE DE VUE	I.17
I.4.1- Choix de la caméra	I.17
I.4.2- Prise de vue	I.18
I.4.3- Analyse du mouvement	I.19
II- ANALYSE DU MOUVEMENT DANS LES SEQUENCES D'IMAGES	II.1
II.1- TECHNIQUES DE TRAITEMENT DE SEQUENCES D'IMAGES	II.1
II.1.1- Analyse structurelle	II.3
II.1.1.1 Extraction d'éléments structurels	II.3
II.1.1.2- Mise en correspondance	II.4
II.1.2- Etude des variations locales de l'intensité	II.6
II.1.2.1- Différences d'images	II.6
II.1.2.2- Corrélation	II.7
II.1.2.3- Gradient spatio-temporel	II.8
II.1.2.4- Méthode de Fourier	II.10
II.1.3- Comparaison des différentes approches	II.10
II.1.4- Flux optique et image différence	II.11
II.1.4.1- Le flux optique	II.11
II.1.4.2- L'image de la différence	II.14
II.2- ANALYSE DE SEQUENCES D'IMAGES	II.14
II.2.1- Détection du mouvement	II.16
II.2.2- Segmentation	II.17
II.2.3- Estimation du mouvement	II.19
II.2.3.1- Estimation du mouvement 2-D	II.19
II.2.3.2- Estimation du mouvement 3-D	II.20
II.2.3.3- Navigation passive	II.23
II.2.4- Reconstitution de formes 3-D	II.24
II.2.4.1- Profondeur de l'objet	II.24
II.2.4.2- Structure de l'objet	II.24
II.2.5- Reconnaissance d'objets non-rigides	II.25
II.3- CONCLUSION ET CHOIX D'UNE METHODE	II.25

S-2

# **III- DETECTION DES CONTOURS D'OBJETS EN MOUVEMENT SUR FOND NON UNIFORME ET EN ECLAIRAGE NON CONTROLE**

III.1- EXTRACTION DES REGIONS EN MOUVEMENT	III.1
III.1.1- Avec image de référence.	III.2
III.1.1.1- Eclairement non constant.	III.3
III.1.1.2- Impossibilité de disposer d'une image de référence exempte d'éléments	
mobiles	III.4
Dreschler et Nagel	III.4
Tang et Snyder	III.4
Jain	III.5

III.1.2- Sans image de référence	III.6
III.1.2.1- Deux images successives	III.6
III.1.2.2- Plus de deux images successives	III.9
III.1.3- Limites de l'approche par analyse de régions	III.10
<b>III.2- EXTRACTION DES CONTOURS EN MOUVEMENT</b>	III.10
III.2.1- Opérateurs existants	III.10
MARR et ULLMAN	III.11
HAYNES et JAIN	III.12
STELMASZYK et Coll.	III.14
III.2.2- Critique des opérateurs existants	III.17
III.2.3- Opérateur proposé	III.20
III.3.3.1- Description de l'algorithme	III.20
III.3.3.2- Comparaisons et résultats	III.22
III.3.3.3- Limites de l'opérateur	III.22
III.3- CONCLUSION	III.26

# **IV- MISE EN OEUVRE ORIENTEE TEMPS REEL**

IV.1- DESCRIPTION MATERIELLE	IV.1
IV.1.1- Acquisition des images	IV.2
IV.1.2- Description du système de traitement d'images	IV.2
IV.1.2.1- Carte iSBC 286/12	IV.3
IV.1.2.2- Cartes IP-512.	IV.3
IV.1.2.3- Transfert des données entre calculateur et mémoires d'images	IV.5
IV.2- OPERATEUR DE DETECTION DE CONTOURS	IV.5
IV.2.1- Détection des contours	IV.5
IV.2.2- Opérateurs différentiels de type gradient	IV.6
IV.2.3- Opérateurs différentiels de type laplacien	IV.9
IV.2.4- Choix du détecteur de contours	IV.10
<b>IV.3- OPERATEUR DE DETECTION DES CONTOURS EN MOUVEMENT</b>	IV.13
IV.3.1- Extraction des contours en mouvement	IV.13
IV.3.2- Filtrage et mise en forme	IV.15
IV.3.3- Conclusion	IV.16
IV.4- MODELISATION DES OBJETS EN MOUVEMENT	IV.18
IV.4.1- Descriptions géométriques	IV.18
IV.4.2- Modélisation du contour extérieur par un polygone	IV.20
IV.4.2.1- Modèle polygonal convexe de périmètre minimal	IV.20
IV.4.2.2- Modèle rectangulaire	IV.21
IV.4.2.3- Modèle octogonal	IV.22
IV.4.3- Implantation	IV.25
IV.5- PERFORMANCES TEMPORELLES	IV.25

<u>S-3</u>

# V- APPLICATION A LA SURVEILLANCE DU TRAFIC URBAIN EN VUE DE LA CONDUITE AUTOMATIQUE D'UN TRAMWAY

V.1- ANALYSE DE DIVERSES SCENES REELLES	<b>V.1</b>
V.1.1- Séquence 1	<b>V.1</b>
V.1.2- Séquence 2	<b>V.4</b>
V.1.3- Séquence 3	<b>V.</b> 6
V.2- EFFETS DES CONDITIONS D'ECLAIRAGE	<b>V.</b> 6
V.3- ACCELERATION DU PROCESSUS DE TRAITEMENT DES SEQUENCES	V.9
V.4- CONCLUSION	V.12

#### CONCLUSION

#### **BIBLIOGRAPHIE**

## - CHAPITRE I -

## **TRANSPORTS GUIDES ET VISION ARTIFICIELLE**

#### **I.1- AUTOMATISATION DES SYSTEMES DE TRANSPORT GUIDE**

#### I.1.1- Nécessité de l'automatisation

L'automatisation s'est beaucoup développée dans les métros ces dernières années. C'est le cas notamment des métros de Lille, Kobe, Osaka, Vancouver et plus récemment, de Londres, qui sont entièrement automatisés. Cette automatisation concerne essentiellement des nouvelles lignes de métros, dits légers, qui ne nécessitent pas de trop gros travaux de génie civil (tunnels de faibles sections, longueurs de quais réduites, viaducs aériens). Cette automatisation a peu intéressé jusqu'à présent les transports de surface guidés tels que les tramways, à cause de leur interaction avec le trafic routier, bien que celle-ci soit souvent réduite.

La conception des tramways évolue actuellement, et l'on voit apparaître des "métros légers" dont la particularité est de rouler en surface comme les tramways, mais sur des voies construites en site protégé sur la majeure partie de leur parcours. La raison pour laquelle ce genre de moyen de transports se développe est principalement d'ordre économique car ils ne nécessitent que peu de travaux de génie civil.

Les conditions de circulation de ces systèmes se rapprochent par conséquent de celles des métros, sauf bien entendu dans les zones banalisées où ils se trouvent mêlés à la circulation routière. Il serait très intéressant de doter également de tels systèmes de transport guidés d'automatismes de conduite afin :

- d'accroître leur sécurité grâce à des dispositifs de détection automatique d'obstacles;

- de moduler facilement le flux des rames en fonction des besoins ;

# **PRESENTATION GENERALE**

Dans le premier chapitre, nous présentons tout d'abord le cadre dans lequel ont été effectués les travaux qui font l'objet de cette thèse.

Après avoir exposé l'état de l'art du traitement de séquences d'images en temps réel dans les domaines de la poursuite de cibles, de la conduite de véhicules autonomes et de la surveillance du trafic routier, nous mettons en évidence la spécificité de notre application. Nous définissons alors une stratégie de prise de vue et mettons en évidence la nécessité d'une technique d'analyse de mouvement adaptée.

Le deuxième chapitre décrit les techniques classiques de traitement de séquences d'images ainsi que leurs possibilités d'exploitation. Cet exposé bibliographique permet de mettre en évidence l'intérêt de la technique basée sur la différence d'images.

Dans le troisième chapitre, aprés avoir exposé les bases des méthodes de détection de mouvement par différence d'images, nous montrons les limites de toutes les méthodes existantes basées sur cette approche. Nous proposons alors une méthode originale utilisant trois images successives afin de s'affranchir de la nature du fond sur lequel évoluent les objets en mouvement.

Le quatrième chapitre est consacré à l'implantation de cette méthode sur machine orientée temps réel. Il expose également les différents traitements complémentaires nécessaires pour aboutir à une modélisation correcte des contours extérieurs des objets en mouvement par des octogones exinscrits.

Les résultats obtenus par l'application de la procédure décrite dans le chapitre IV à des images de scènes réelles enregistrées au niveau de carrefours urbains sont présentés dans le chapitre V. Ces essais montrent la robustesse, la fiabilité mais également les limites de la procédure.

Enfin, en conclusion, nous indiquons les perspectives à envisager afin d'apporter une réponse aux divers problèmes liés à une implantation permanente sur site réel. - de mieux synchroniser les mouvements des rames avec les commandes de feux de circulation ;

- de mettre de très nombreuses rames à la disposition des voyageurs.

## **I.1.2-** Marchés potentiels

De par le monde, 48 villes possèdent déjà un réseau de transports guidés de surface dont la plupart sont en extension et 40 villes s'équipent ou prévoient de s'équiper d'un tel réseau. Ceci représente pour ces 88 villes, un total de 1500 km de lignes soit en construction (400 km), soit en projet et de 1800 km de lignes à l'étude [Bou87].

Les exploitants d'autres systèmes existants, tels que les métros de Lyon et Paris, qui ne possèdent pourtant que de faibles zones non-protégées (stations uniquement), s'intéressent également à une automatisation ne nécessitant pas les infrastructures lourdes que représenteraient l'installation de portes palières afin de surveiller les interfaces quais-voies.

Pour tenter de répondre aux besoins grandissants d'automatisation de tout ces systèmes de transports guidés (métros souterrains, métros de surface, tramways), le Groupement Régional Nord/Pas-de-Calais pour la Recherche dans les Transports (G.R.R.T) a entrepris, en 1983, un vaste programme de recherche sur ce thème.

#### **I.1.3- Programme de recherche**

Le programme se décompose de la façon suivante :

1. Analyse sur un site réel des conditions de circulation d'une ligne de tramway existante ayant les caractéristiques d'un métro léger.

2. Etude de la configuration et des conditions générales d'exploitation de lignes de transport de surfaces, existantes ou en projet, en France ou à l'étranger, afin d'identifier celles qui pourraient se prêter le mieux à une automatisation partielle ou intégrale et d'évaluer les débouchés potentiels de ces automatismes.

3. Etude sur modèle de simulation des stratégies d'exploitation d'une ligne de métro léger automatique en tenant compte de ses interfaces avec la circulation.

4. Spécification et étude des automatismes à développer pour ces transports, notamment en matière :

- de pilotage automatique ;

- de détection d'obstacles ;

- de régulation des véhicules.

La détection d'obstacles évoquée dans cette quatrième partie du programme fut confiée au Centre d'Automatique de Lille dès 1983 dans l'optique d'une approche intégrant la vision artificielle. Cette approche a fait l'objet d'une thèse en 1985 [Ste85] et les travaux poursuivis depuis au sein de ce même laboratoire font l'objet de cette étude.

#### **I.2- DETECTION D'OBSTACLES ET VISION ARTIFICIELLE**

Les lignes de tramway reliant Lille, Roubaix et Tourcoing constituent un site d'expérimentation privilégié. En effet, ce système de transport guidé, appelé Mongy (du nom de son créateur) répond de manière satisfaisante à la définition d'un métro léger, c'est-à-dire un système de transport guidé circulant en site protégé sur la majeure partie de son parcours. Les rames de ce système de transport ne sont en fait mêlées à la circulation routière qu'au niveau des carrefours.

Une étude précédente [Dep84] montre qu'à ces carrefours, régis par des feux tricolores donnant la priorité absolue aux tramways, se produisent fréquemment des situations de conflits, entraînant parfois des incidents, sinon des accidents.

L'étude montre que certains conflits et incidents peuvent être facilement évités. Par contre, d'autres conflits, dûs au non respect des feux et règles de circulation par les usagers, automobilistes ou piétons, se traduisent parfois par des accidents malgré la vigilance des conducteurs.

Il est évident qu'une circulation en site propre au niveau des carrefours grâce à des tunnels ou des viaducs ou par des portes palières au niveau des stations empêcherait tout incident. Cependant, cette solution nécessite des travaux de génie civil trop importants, trop longs et par conséquent trop coûteux pour rendre l'automatisation intéressante.

C'est la raison pour laquelle on envisage actuellement de conserver les lignes en site semi-propre. Il apparaît donc que le pilotage des tramways dans les zones non protégées ne peut se faire sans prise en compte de leur environnement.

La perception de l'environnement du tramway au niveau des carrefours peut être obtenue par l'utilisation de capteurs classiques tels que radars, capteurs tactiles ou de proximités, boucles d'induction, etc ...

Il apparait cependant qu'une prise en compte globale de l'environnement au niveau d'un carrefour constitue le meilleur atout pour prédire avec précision l'évolution des mobiles à proximité du tramway et le piloter ainsi avec le maximum de fiabilité.

L'utilisation des capteurs ponctuels évoqués ci-dessus est donc à proscrire car elle nécessiterait de très nombreux points de mesures difficiles à installer et à exploiter.

Dans ces conditions, la vision artificielle parait très bien adaptée à cette situation puisqu'elle permet de prendre en compte l'environnement grâce à une analyse fine de la scène observée.

Toutes ces considérations ont conduit à un cahier des charges pouvant se résumer de la manière suivante : extraire, à partir d'images du carrefour, les informations indispensables à la traversée de celui-ci par le tramway sans risque de collision avec un obstacle quelconque, tout en ayant des performances de sécurité, de rapidité et de confort pour les usagers comparables, sinon supérieures, à celles obtenues actuellement par un conducteur.

L'analyse du mouvement par traitement d'images ayant déjà fait l'objet de nombreux travaux, une étude préalable de l'état de l'art s'avère nécessaire.

#### **I.3- LE TRAITEMENT D'IMAGES APPLIQUE AUX TRANSPORTS.**

Le traitement d'images appliqué aux transports a fait l'objet de nombreux travaux de recherche dès le début des années 70, autant dans le domaine civil que militaire.

Les résultats obtenus concernent essentiellement le contrôle du trafic routier (essentiellement autoroutier) [Hil80, Wat84, Abr84, Tak82, Tak84, Hou87]. A ce domaine d'application spécifique aux transports on peut rattacher deux autres domaines voisins. Il s'agit de la poursuite de cibles [Gil80, Hua81, Nèg83, Gam84, Gil85, Sam86] et de la conduite automatique de véhicules autonomes [Tsu79, Mor83, Cou84, Ini84, Oza84, Dra85, McV86]. A côté de ces trois domaines qui drainent la majorité des recherches, le projet lancé par le G.R.R.T. nécessite le développement d'un nouvel axe de recherche : la détection d'obstacles aux transports guidés de surface en vue d'une conduite automatique.

Bien que l'objectif recherché ait des similarités importantes avec les trois domaines cités, l'état de l'art analysé dans les trois sections suivantes permettra de mettre en évidence la spécificité de notre projet.

#### **I.3.1-** Poursuite de cibles

#### I.3.1.1- Généralités

Dans ce domaine, le temps réel est requis par définition. Par conséquent, nous ne présenterons que des applications satisfaisant ce critère. Depuis le premier système mis au point par Gilbert en 1980 [Gil80], peu de systèmes temps réel ont vu le jour et la description de ceux-ci, se différenciant autant par l'approche adoptée que par le type de cibles poursuivies, suffit à se rendre compte des limites actuelles de la poursuite de cibles. On trouvera dans [Hua81] une revue plus détaillée de l'état de l'art en 1980.

#### I.3.1.2- Etat de l'art

#### Gilbert [Gil80, Gil85]

Le système de Gilbert se propose de conserver l'image d'un missile dans le champ de la caméra. Le missile se déplace sur un fond composé du ciel, de nuages ou de montagnes. L'orientation et le zoom de la caméra sont contrôlés par trois signaux issus du système de traitement des images, afin de centrer le missile dans le champ de vision. Un quatrième signal, issu également de ce système, permet une rotation de la caméra afin de conserver une image verticale du missile en maintenant la traînée constituée de la flamme et de la fumée vers le bas de l'image.

A l'époque, les simulations effectuées à partir d'images enregistrées ont donné entière satisfaction. Grâce à une stratégie intelligente de poursuite, le système autorise même des pertes momentanées de l'image du missile (occlusion par des nuages). L'hypothèse effectuée par l'auteur est que l'image du missile et de la traînée qui l'accompagne, contiennent des niveaux de gris que l'on ne peut trouver dans le fond voisin. Cette hypothèse étant toujours vérifiée grâce à la présence de la flamme, une étude statistique de la distribution des intensités permet une discrimination aisée des points appartenant soit au fond, soit au missile, soit à la traînée, à l'intérieur d'une fenêtre de calcul centrée sur le missile.

Quoique l'auteur ne mentionne pas d'intervention de l'opérateur au début de la poursuite, il est probable que celle-ci soit nécessaire pour le positionnement initial de la fenêtre de calcul.

#### Nègre [Nèg83]

Plus récemment, Nègre [Nèg83] propose deux systèmes de poursuite, l'un à base de microprocesseurs en tranches, l'autre à base d'opérateurs câblés associés à un microprocesseur monolithique classique. Deux méthodes y sont respectivement associées selon la surface de l'image de la cible : l'une reprend l'étude statistique de la distribution des niveaux de gris utilisée par Gilbert (grande surface), l'autre est basée sur la recherche du maximum de contraste sur une fenêtre réduite (surface réduite). Malheureusement, l'article ne précise ni la nature de la cible poursuivie, ni les résultats obtenus.

En ce qui concerne le premier système, trois opérateurs réalisent en temps réel : le seuillage (selon un critère Bayésien), le calcul d'histogrammes et le gradient (Prewitt) sur une fenêtre de 4000 points (64x64).

Pour le deuxième système, trois opérateurs câblés travaillent en parallèle et possèdent les performances suivantes :

- opérateur de gradient : 4 ms sur une fenêtre de 4000 points (64x64) ;

- opérateur de corrélation : 6,4 ms pour la corrélation d'une fenêtre 16 x 16 avec une fenêtre 32 x 32 ;

- opérateur d'histogrammes : 6 ms pour une fenêtre 100 x 100, soit 10.000 points.

L'auteur prétend résoudre ainsi le problème de poursuite de cibles.

Quoique l'article ne l'exprime pas explicitement, les cibles poursuivies semblent se déplacer sur un fond uniforme (ciel bleu) ou ne présentant pas de fort contraste (nuages). Sinon, les cibles doivent avoir des distributions en intensité différentes de celles du fond, comme dans le système de Gilbert.

#### Samy [Sam85, Sam86]

Dans [Sam86], l'auteur propose deux approches de segmentation de scènes dynamiques texturées. La première est destinée à extraire les objets mobiles d'images fournies par une caméra fixe. La deuxième s'applique à la détection d'une cible se déplaçant dans un paysage texturé, afin d'effectuer une poursuite en temps réel. L'approche nécessite une prédiction plus ou moins précise de la position des objets.

#### Gambotto [Gam84a, Gam84b]

Gambotto propose un algorithme de poursuite multicible à partir d'images infrarouges de basse résolution. Les cibles sont détectées par leurs "points chauds". Une estimation de la vitesse des différentes régions représentant un point chaud permet de reconstituer chaque cible qui est alors modélisée par un ensemble de régions et quelques autres paramètres non précisés.

Les cibles se déplaçant approximativement dans l'axe de la caméra, leur vitesse dans le plan de l'image est faible. Ainsi, malgré la faible cadence de 3 images/s, les variations de positions absolues et relatives des différentes régions correspondant aux points chauds restent également faibles. Ceci permet l'utilisation d'une technique de correspondance globale au niveau des cibles modélisées.

D'autre part, la faible résolution des images (64 x 64) rendant possible le calcul d'histogrammes en temps réel et les variations d'intensité des objets en imagerie infrarouge étant lentes, il est par conséquent possible d'envisager une approche par l'analyse de régions pour segmenter la scène. Celle-ci est ensuite modélisée sous forme d'un ensemble de régions d'intensités homogènes. L'analyse de correspondance locale au niveau des régions entre deux images consécutives est alors effectuée très rapidement en utilisant les deux seuls paramètres suivants : la surface et l'intensité.

Cette double modélisation, des cibles d'une part, des régions d'autre part, permet de détecter d'éventuelles mises en correspondance erronées.

#### I.3.1.3- Conclusion

Les seuls systèmes performants de poursuite temps réel possèdent les caractéristiques suivantes :

- cible unique quelconque se déplaçant sur fond uniforme ou quasi-uniforme (ciel, nuages, montagnes lointaines) [Gil80, Gil85];

- cible unique possédant des intensités différentes de celles du fond (flamme du missile, points chauds en imagerie infrarouge, etc...) sur fond quelconque [Gam84a, Gam84b, Gil80].

Ces caractéristiques permettent par conséquent une détection aisée de la cible. En revanche, les solutions proposées sont toutes mises en défaut en présence de cibles à plusieurs niveaux de gris se déplaçant sur un fond non uniforme.

D'autre part, si la cible ne représente qu'une faible fraction de l'image, une localisation manuelle est souvent requise pour l'initialisation de la poursuite [Gil80].

Etant donné la quantité énorme de données contenues dans une séquence d'images vidéo (numérisée sur 256 x 256 éléments x 8 bits, cela représente 64 Koctets toutes les 40 ms, soit  $1,6.10^6$  octets par seconde) par rapport à la puissance de calcul actuelle des ordinateurs, la poursuite en temps réel nécessite une réduction de données. C'est pourquoi, la plupart des systèmes proposés limitent la surface de traitement à une fenêtre représentant une faible partie de l'image vidéo (cas de petite cible).

Cependant, si le manque de connaissances concernant la ou les cibles en mouvement dans l'image ne permet pas de déterminer la ou les fenêtres de calcul (position et taille), il est possible de parvenir à une réduction des données non plus au niveau de l'image elle-même, mais au niveau de la séquence. La solution consiste alors à réduire la cadence d'acquisition des images selon un échantillonnage de période correspondant au temps effectif de traitement d'une image.

Cela nécessite alors de tenir compte de la sensibilité des algorithmes de poursuite envisagés aux variations de position, d'orientation, de forme et du facteur d'échelle des objets. En effet, si les mouvements de ces derniers entre deux images successives sont trop importants les algotithmes risquent d'échouer.

Enfin, la réduction de données peut être envisagée par l'emploi d'un capteur de faible résolution lorsque les objets à poursuivre sont relativement grands [Gam84a, Gam84b].

#### **I.3.2-** Conduite automatique de véhicules autonomes

#### I.3.2.1- Généralités

De nombreux systèmes expérimentaux ont été développés pour la conduite de véhicules autonomes. La plupart concernent le déplacement de robots mobiles à l'intérieur d'une usine, c'est à dire en univers contrôlé. Ces systèmes ne font généralement appel à la vision que pour repérer des marqueurs divers utilisés pour leur guidage (ligne au sol généralement) [Dra85, McV86, Cou84].

Plus récemment, les systèmes expérimentaux ont été appliqués à des véhicules circulant dans un univers inconnu sans repère spécifique. Par conséquent, le véhicule doit non seulement être capable de suivre un itinéraire (généralement une route) en se servant des informations disponibles (signalisation horizontale) [Oza86], mais également de détecter les obstacles mobiles ou non.

Ces informations doivent évidemment être fournies en temps réel pour permettre le contrôle du véhicule.

Quoique de nombreuses techniques (infrarouges, vision, ultrason, radar, etc...) aient été employées pour le guidage ou la détection et l'évitement d'obstacles, aucune n'a encore affirmé sa supériorité, chacune ayant des avantages spécifiques selon l'environnement considéré. Cependant, un système de vision a l'avantage de pouvoir résoudre à la fois la tâche de navigation et les tâches de détection et d'évitement d'obstacles dans certains environnements banalisés. C'est pourquoi de nombreuses recherches sont effectuées sur la vision artificielle appliquée aux véhicules autonomes.

Quelques systèmes expérimentaux utilisant la vision artificielle sont décrits cidessous, en commençant par le plus ancien.

#### I.3.2.1- Etat de l'art

#### Le Robot Rover [Mor79, Mor83]

Ce "Robot Rover" utilise la stéréovision pour identifier les obstacles sur le chemin du robot. Une technique de génération d'itinéraire utilise une modélisation des obstacles sous forme de sphères et cherche le meilleur chemin pour les éviter. Malheureusement le système ne peut opérer en temps réel et le robot se déplace par bonds de 1 m toutes les 10 mn.

#### TSUGAWA et coll. (1979) [Tsu79, Yat78]

Le système de Tsugawa, sur lequel nous disposons de peu d'information, est embarqué à bord d'une automobile standard se déplaçant sur une route standard, à ceci près qu'elle fut interdite à la circulation durant l'essai. L'automobile se déplaçait alors à la vitesse maximale de 32 km/h.

#### **INIGO** (1984) [Ini84]

Inigo propose un ensemble d'algorithmes implantables en temps réel pour l'identification et la localisation des bords du chemin ainsi que pour la détection d'obstacles.

Le processus de localisation du chemin a été testé sur une scène réelle en lumière naturelle non variable. Le chemin était constitué d'une allée bordée de colonnes, cellesci constituant les seuls obstacles potentiels.

Le processus de détection d'obstacles testé sur une simple scène constituée d'un camion miniature, a permis d'identifier celui-ci comme un obstacle et de mesurer la portion du chemin qu'il occupait.

Les différents algorithmes mis en oeuvre dans chacun des deux processus étant indépendants, ceux-ci peuvent être effectués en parallèle. Par conséquent, la fréquence d'acquisition des images est déterminée par l'algorithme le plus lent, qui est la transformation de Hough. En utilisant une image de 100 x 100 éléments, une fréquence de 10 images/s reste compatible avec la puissance de calcul des microprocesseurs classiques.

#### I.3.2.3- Conclusion

Comme on peut le constater, la conduite automatique de véhicules en est à ses balbutiements et, du Robot Rover en 1979 au système de Inigo en 1984, les quelques avancées dans le domaine semblent bien faibles comparées au chemin qu'il reste à parcourir.

La nécessité d'analyser la scène à partir de caméras embarquées afin de rendre le véhicule totalement autonome, constitue un frein au développement de ce type de véhicules. En effet, sans connaissance *a priori* sur les dimensions des obstacles rencontrés, il est impossible d'apprécier la dimension horizontale sans avoir recours à la

stéréovision. L'appréciation de cette troisième dimension se fait alors au prix de calculs coûteux. La discrimination entre les obstacles et les pseudo-obstacles (ombres, reliefs du sol) nécessite également la stéréovision.

Aucun des systèmes proposés n'évoque le problème des variations d'illumination et la détection d'obstacles se limite aux cas d'objets stationnaires. Malgré ces restrictions importantes, on ne parvient pas encore à des vitesses satisfaisantes, ni pour un robot circulant dans un univers connu et encore moins pour un moyen de transport circulant en environnement inconnu.

La contrainte temps réel freine considérablement l'utilisation d'algorithmes complexes. En [Ini82], il est montré qu'une période d'échantillonnage de 10 ms est nécessaire pour guider un véhicule circulant à 95 km/h. Cette cadence contraint par conséquent à l'emploi d'algorithmes facilement implantables sur circuit VLSI.

De nombreux autres systèmes et techniques peuvent être trouvés dans la littérature [Gen79, Gir79, Sch71, Wil80]. Cependant, une seule référence fait allusion à un contrôle temps réel du véhicule [Sch71].

#### I.3.3- Contrôle du trafic routier

#### I.3.3.1- Généralités

Une gamme de données de plus en plus étendue (volume de trafic, vitesses, espace inter-véhicules, occupation des voies, classification des véhicules, provenances et destinations aux intersections) sur le trafic routier fait l'objet d'une demande sans cesse accrue de la part d'autorités locales ou nationales en vue d'exploitations diverses (modifications du réseau routier, étude sur la sécurité et le comportement des conducteurs, mise en place de plan de délestage, etc..). Bien que la plupart de ces données puissent être obtenues à l'aide de détecteurs divers (boucles d'induction, radars, etc...), leur installation pose de réels problèmes et leurs mesures ponctuelles ne permettent pas l'extraction de certaines données. C'est pourquoi on utilise, depuis longtemps, les images qui offrent une plus grande flexibilité et une plus grande richesse d'information. Cependant, l'exploitation manuelle de ces images est fastidieuse (un mois pour le suivi de 1500 véhicules) et l'intérêt d'un traitement automatique des images est vite apparu.

Le WADS a été testé sous de bonnes conditions climatiques et de jour. Les performances obtenues sont les suivantes : erreur de 4 % sur les détections et de 2 % sur les mesures de vitesse par temps ensoleillé, erreur de 10 % sur les détections par temps nuageux ou par éclairage artificiel.

Ce système, grâce à une réduction de données importante, devrait permettre une implantation temps réel en utilisant la technologie des microprocesseurs courants.

#### Projet de l'UMIST (G-B) [Wat84].

Le système proposé par l'Institut des Sciences et Techniques de l'Université de Manchester (UMIST) se limite actuellement au comptage des véhicules.

L'axe optique de la caméra, de résolution 100x100, est placé à 22,5 m du sol au dessus d'une double voie. La fréquence d'acquisition est réduite à 8 images/s.

Portant sur une séquence de 30 mn durant laquelle l'illumination a varié d'un facteur quatre, une erreur de 11 % a été mesurée sur l'ensemble des deux voies. Il semble cependant qu'une intervention humaine soit nécessaire pour ajuster un seuil lors des variations d'éclairage.

#### Projet du SRIT (Suède) [Abr84]

Le Laboratoire d'Instrumentation de l'Institut Royal Suédois de Technologie (SRIT) propose un système capable de détecter et de poursuivre les véhicules.

La caméra, de résolution 128x128 pixels, est placée à 7,5 m du sol de telle sorte que le balayage soit parallèle à la route dont on observe un tronçon long de 200 m. La fréquence d'acquisition est réduite à 8 images/s. L'article ne précise pas si le système nécessite un seuillage manuel.

La poursuite des véhicules peut échouer dans les situations suivantes : absence du véhicule dans la zone observée ; changement de voie ; changement apparent de la longueur dû à un manque de contraste ou à du bruit trop important ; occlusion partielle ou totale par un autre véhicule.

Les objectifs à atteindre pour ce système de vision sont :

- la détermination du nombre de véhicules de chaque classe (V.L., P.L., autocar...) et de leurs positions par rapport à la caméra et aux autres véhicules ;

- la poursuite (jusqu'à 1 km) de chaque véhicule afin de déterminer : ses éventuels changements de voies, son origine et sa destination aux éventuelles intersections ainsi que sa vitesse.

Pour rendre le système de vision satisfaisant pour l'ensemble de ces tâches, il est nécessaire que celles-ci s'effectuent en temps réel.

Afin de parvenir à ces objectifs, les systèmes existants réalisent généralement la prise de vue à partir d'une caméra placée en bordure de la chaussée, à une hauteur de quelques dizaines de mètres et visant l'axe de celle-ci à l'horizon. Par conséquent, les véhicules sont vus soit de face, soit de l'arrière, et se déplacent selon une direction privilégiée.

Depuis la tentative de Onoe *et coll.* en 1973 [Ono73], de nombreux systèmes ont été proposés. Le paragraphe suivant décrit les résultats obtenus avec les systèmes les plus récents.

#### I.3.3.2- Etat de l'art

Les résultats des 5 principaux systèmes développés de par le monde sont exposés ci-dessous. Parmi ceux-ci seul le dernier a fait l'objet d'une implantation sur site.

#### Projet du JPL (USA) : le WADS [Hil80a, Hil80b]

Le WADS, pour "Wide Area Detection System", prototype développé par le JPL (Jet Propulsion Laboratory, USA), semble actuellement le plus performant.

Ce système est capable de détecter les véhicules sur trois voies et d'en calculer la vitesse. L'objectif initial de poursuite des véhicules a été abandonné car il nécessite un matériel trop coûteux, des calculs trop importants et est de plus d'une robustesse insuffisante face aux variations d'éclairage ainsi qu'aux incertitudes sur la hauteur de chaque véhicule [Sch82].

L'approche utilisée ainsi que les algorithmes employés sont exposés dans [Hil78, Sch82]. La prise de vue est réalisée à l'aide d'une caméra placée à 7,5 m du sol.

#### Projet de l'UT (Japon) [Tak82, Tak83].

Le système proposé par l'Université de Tokyo (UT) peut mesurer le volume de trafic (nombre de véhicules par unité de temps en un point de la voie) ainsi que la vitesse et la densité (nombre total de véhicules dans une zone donnée à un instant donné).

La caméra est placée quasiment à la verticale d'un carrefour régulé par des feux tricolores à environ 25 m du sol sur un toit d'immeuble. La fréquence d'acquisition choisie est de 30 images/s.

Des tests effectués en laboratoire sur une séquence de 40 mn, enregistrée par temps clair sur bande vidéo, ont conduit aux résultats suivants : sur un total de 317 véhicules durant 11 cycles de feux, l'erreur du système de détection fut inférieure à 5%. Lors de changements notables de l'illumination, le système a donné de "bons résultats" selon l'auteur (aucune mesure indiquée) bien qu'il n'ait pas été conçu pour cela. En ce qui concerne les estimations de vitesses, une erreur maximale de 10% a été obtenue.

#### Système opérationnel d'OOYAMA-SHIGETA (Japon) [Tak84].

Le seul système implanté sur site réel dans le monde est celui de Ooyama et Shigeta. Ce système est capable de détecter les véhicules et d'en mesurer la vitesse. D'autres paramètres tels que le taux d'occupation des voies, la longueur de files et la densité peuvent également être mesurés.

Le système a été entièrement conçu pour cette application. Le capteur, spécifique, est une matrice d'éléments photoélectriques dont la géométrie compense l'effet de perspective produit pour une position donnée de la caméra équipée d'un objectif donné. Les éléments ont été choisis de sorte que leur réponse spectrale soit satisfaisante 24 h sur 24 h (crépuscule, aube et nuit compris). La portion de route couverte par la caméra s'étend de 17,5 m à 275 m.

Installé pendant au moins 2 ans (719 jours à la date du 19 février 1985) à Tokyo, le système a permis d'effectuer des tests sous différentes conditions climatiques, telles que typhons, neige, brouillard et séismes. Durant cette période, l'erreur maximale absolue sur les mesures de vitesse fut de 3 km/h et le pourcentage d'erreur maximal de 6 %. Les vitesses des véhicules restaient inférieures à 50 km/h.

#### Autres systèmes

D'autres systèmes expérimentaux sont décrits en [Ono73, Hou80, Kud81, Abr82, Aok84, Hou87]. Signalons notamment les travaux de Blosseville [Blo87] qui, employant une approche tout à fait originale basée sur la morphologie mathématique, obtient une fiabilité et une robustesse exceptionnelles aux variations d'éclairage ainsi qu'aux conditions climatiques. L'inconvénient majeur réside, actuellement, dans le temps de traitement très long. Cependant, d'après les auteurs, le temps réel est envisageable à court terme grâce à l'élaboration d'un processeur spécifique câblé.

En marge du contrôle de trafic, citons les travaux de Mott concernant la détection d'autobus dans le trafic londonien. L'algorithme de reconnaissance, testé dans de bonnes conditions d'éclairages, est assez fiable et autorise même une occlusion partielle de l'autobus. Cependant, l'algorithme ne fonctionne pas en temps réel et les orientations et positions possibles de l'autobus sont fortement limitées.

Enfin, Oscarsson [Osc82] utilise le traitement d'images pour la détection et le comptage de piétons en vue du contrôle des feux de circulation.

#### I.3.3.3- Aspect temps réel

Afin d'aboutir au temps réel, tous ces systèmes procèdent à une importante réduction de données pour compenser la puissance insuffisante des processeurs actuels : le WADS n'utilise que 1140 pixels/image, le système SRIT en utilise 1152, les autres systèmes décrits ci-dessus demeurent dans le même ordre de grandeur. Cette réduction est obtenue de manières diverses selon l'approche utilisée : par un échantillonnage spatial de l'image [Hil80a,b ; Wat84, Abr84, Tak82,83,84]; par l'extraction des seuls points en mouvement [Ono73] ; par réduction du champ de la caméra afin de réduire la quantité de véhicules apparaissant dans l'image [The80].

Malheureusement, la réduction de données effectuée conduit généralement à l'impossibilité de réaliser la poursuite des véhicules. C'est le cas du système de l'UT, de celui d'Ooyama-Shigeta ainsi que de celui décrit en [The80].

#### I.3.3.4- Conclusion

Si un seul système a été implanté jusqu'à présent, il semble que ce soit essentiellement à cause des effets de certaines conditions climatiques qui dégradent fortement leurs performances.

De plus, le coût d'un système de vision est encore actuellement supérieur à celui des autres systèmes de contrôle du trafic routier réalisés à l'aide de capteurs classiques.

#### **I.3.4-** Conclusion : spécificité du projet

Le problème de la détection des obstacles pouvant perturber le fonctionnement d'un système de transport guidé présente des particularités importantes le différenciant de chacun des trois domaines traités ci-dessus.

Par rapport à la poursuite de cibles, la particularité majeure réside dans l'aspect multiple des "cibles". Celles-ci sont de plus très variées, en tailles, aspects, positions et orientations. D'autre part, aucune connaissance *a priori* n'est disponible à leur propos (forme, dimension, luminance, texture, etc ...).

Par rapport aux véhicules autonomes, notons les différences majeures suivantes :

- les véhicules de transport guidé circulent en site propre sauf dans des zones relativement restreintes et en nombre limité ;

- les véhicules de transport guidé circulent toujours à l'extérieur.

Par rapport au contrôle de trafic, notons les restrictions majeures suivantes :

- on doit tenir compte des piétons et cycles ;

- les véhicules ont une orientation quelconque par rapport à la caméra ;

Les trois paragraphes précédents mettent également en évidence l'influence de la stratégie de prise de vue sur l'approche à envisager pour analyser les mouvements.

Avant d'entrer dans le problème du traitement d'images proprement dit, il est par conséquent indispensable de définir la stratégie de prise de vue la mieux adaptée à notre application.

## **I.4- CHOIX D'UNE STRATEGIE DE PRISE DE VUE**

Pour préciser la stratégie de prise de vue, il faut définir :

- le type de caméra à employer ;

- leur nombre et leurs positions.

Les travaux réalisés à ce sujet par Stelmaszyk [Ste85] ont permis de définir avec précision et certitude ces deux éléments.

#### I.4.1- Choix de la caméra

La caméra peut être choisie en fonction du domaine spectrale dans lequel elle travaille : infra-rouge ou visible.

Les caméras opérant dans l'infrarouge présentent, par rapport à celles opérant dans le visible, l'avantage d'une insensibilité aux variations d'éclairages et aux ombres.

Quoique encore onéreuses, le développement actuel de leurs domaines d'application risque de faire baisser le prix de ces capteurs. Leur emploi n'est donc pas à écarter définitivement.

Notons toutefois que l'utilisation de tels capteurs conduira à une interprétation difficile des différents points chauds détectés par ces caméras. En effet, une automobile, par exemple, sera détectée uniquement par des tâches représentant chacune de ses parties chaudes (pneus, freins, pot d'échappement, moteur, etc...). La "reconstitution" de l'automobile à partir de ces tâches disjointes ne sera pas nécessairement évidente dans la mesure où l'on n'a aucune connaissance sur le nombre d'objets présents dans la scène (véhicules ou piétons). Certaines détections "parasites" ne correspondant pas nécessairement à une partie d'objet en mouvement (gaz d'échappement, plaque d'égoût, etc ...) risque de surcroît d'aggraver le problème. Par conséquent, l'emploi de ce type de caméra ne peut être envisagé qu'en complément de celui des caméras opérant dans le visible.

Parmi les caméras opérant dans le visible, notre choix s'est porté vers les caméras à semi-conducteurs de type CCD. Celles-ci suppléent peu à peu les caméras à tube pour des raisons de confort d'utilisation (encombrement et poids réduits), de fiabilité (durée de vie décuplée, absence de distorsions géométriques et d'effets de rémanence), et d'insensibilité aux sources lumineuses intenses, telle que la réflexion du soleil sur les pare-brises.

#### I.4.2.- Prise de vue

Les études précédentes [Ste85] ont montré que la solution d'une prise de vue par caméra embarquée (simple substitution d'un oeil "électronique" à l'oeil du conducteur) ne peut satisfaire à la contrainte temps réel dans l'état actuel des techniques de traitement d'images. De plus, l'estimation de la troisième dimension (distance séparant l'obstacle du tramway) à partir d'une seule caméra est quasiment impossible dans le cas de scènes naturelles et il faut alors faire appel à la stéréovision qui nécessite non seulement deux caméras (surcoût) mais surtout des algorithmes d'analyse de l'image très complexes.

La deuxième solution consiste à fixer une ou plusieurs caméras au niveau de chaque carrefour traversé. La position idéale semble alors être à l'aplomb du carrefour. Cette position offre alors les avantages majeurs suivants par rapport à la solution d'une caméra embarquée :

- une vue plus globale du carrefour traversé et de ses abords, permettant ainsi une meilleure prévision des risques de conflit ;

- une diminution du risque d'occlusions entre mobiles ;

- une détermination plus aisée et plus précise du mouvement des mobiles et de leur surface au sol.

Par conséquent, il est évident que, sans même considérer le problème technique de l'analyse du mouvement par traitement d'images, cette deuxième solution est en fait idéale.

D'un point de vue économique, le nombre de sites non protégés (environ 70 stations et carrefours) étant supérieur au nombre de rames (30 motrices), la solution choisie entraîne un besoin supplémentaire de caméras. Cependant, si l'on considère la tendance actuelle des transports guidés et plus particulièrement ceux à pilotage automatique vers une réduction de la taille de leurs rames en faveur d'une augmentation de leur nombre, cet aspect économique risque de devenir tout à fait secondaire.

Les paramètres restant à déterminer sont la hauteur à laquelle doivent être placées les caméras et la distance focale des objectifs.

Pour cela, il faut tenir compte de :

- la sensibilité du support au vent ;

- la déformation des objets observés en fonction de leur position par rapport à l'axe optique ;

- l'occlusion des mobiles entre eux ou par le tramway ;

- la taille minimale des objets à repérer.

La stratégie de prise de vue définie, il est alors nécessaire de définir une méthode d'analyse du mouvement dans la scène afin de modéliser celle-ci pour générer la commande du tramway.

#### **I.4.3- Analyse du mouvement**

Les recherches en vision artificielle, et plus particulièrement sur le mouvement, se décomposent en deux axes :

- les recherches effectuées par les biologistes et psychophysiologistes pour comprendre le processus de perception du mouvement par les systèmes biologiques ;

- les recherches effectuées par les automaticiens pour développer des algorithmes de perception et d'analyse du mouvement sans toutefois rechercher à reproduire systématiquement les processus physiologiques.

L'analyse du mouvement par les techniques de traitement d'images s'est développée dans divers domaines que l'on peut diviser en deux classes selon la nature du mouvement considéré :

- l'analyse du mouvement "naturel" concerne principalement les trois domaines suivants : la médecine (étude du mouvement du coeur essentiellement), la biologie

#### Choix d'une stratégie de prise de vue

(étude du mouvement des micro-organismes), la météorologie (étude du mouvement des nuages). La caractéristique du mouvement dans ces domaines est essentiellement l'aspect déformable des mobiles étudiés (coeur, micro-organismes, nuages);

- l'analyse du mouvement "artificiel" concerne également trois domaines : la poursuite de cibles (fusées, avions, chars, etc...), la conduite automatique de véhicules (robots, voitures) et le contrôle du trafic routier (principalement autoroutier). Les mobiles sont ici indéformables.

La diversité des méthodes d'analyse du mouvement démontre l'absence d'une méthodologie universelle adaptée à tous les problèmes d'analyse de scènes dynamiques. Cette diversité se retrouve même à l'intérieur de chacun des domaines. C'est pourquoi une étude approfondie de l'état de la recherche est nécessaire.

I-20

# - CHAPITRE II -

## ANALYSE DU MOUVEMENT DANS LES SEQUENCES D'IMAGES

L'analyse de séquences d'images acquises à intervalles de temps réguliers a fait l'objet de très nombreux travaux qui peuvent être répertoriés en cinq principaux domaines de recherche selon l'objectif visé. La figure II.1 indique ces domaines de recherche et leurs relations. Pour deux d'entre eux, l'analyse du mouvement constitue la finalité essentielle recherchée (détection, estimation). Pour les trois autres, le mouvement constitue un moyen d'analyse pour parvenir à une description de la scène (segmentation, détermination de formes, reconnaissance d'objets).

Cependant, quoique très distincts, les objectifs de ces domaines de recherche passent tous par l'interprétation de résultats issus d'un nombre limité de techniques de traitement de séquences d'images.

#### **II.1- TECHNIQUES DE TRAITEMENT DE SEQUENCES D'IMAGES**

Dans les systèmes de vision biologiques, la perception du mouvement se fait selon trois phases successives :

- périphérique ;
- attentive ;
- cognitive.

La phase périphérique détecte les zones du champ de vision soumises à des variations. Durant la phase attentive, l'observateur concentre son attention sur chacune de ces zones afin d'y étudier en détail les variations. Enfin, la phase cognitive met en relation les résultats de la phase attentive avec la connaissance que possède l'observateur sur la scène étudiée afin d'interpréter les variations [Mar78].



Figure II.1 : Principaux domaines de l'analyse de séquences d'images

De façon similaire, en vision artificielle, l'analyse de séquences d'images débute toujours par une phase de prétraitement parfois appelée phase périphérique par analogie aux systèmes biologiques. Ce prétraitement peut s'effectuer selon deux approches fondamentalement différentes qui sont : l'analyse structurelle d'une part, l'étude des variations locales de l'intensité d'autre part.

L'analyse structurelle ("feature-based method") utilise des techniques de reconnaissance de formes statiques. Par conséquent, c'est une approche qui analyse chaque image de la séquence indépendamment les unes des autres et qui nécessite, en général, un minimum de connaissance *a priori* sur les objets mobiles. C'est pourquoi, cette approche est parfois dite orientée objet ("object-based").

L'approche basée sur l'étude des variations locales de l'intensité a été développée exclusivement pour l'étude du mouvement. C'est une approche qui utilise le mouvement lui même comme moyen d'analyse et qui ne nécessite aucune connaissance *a priori* sur les objets étudiés. C'est pourquoi, cette approche est parfois dite orientée mouvement ("motion-based").

II-2

#### **II.1.1- Analyse structurelle**

D'une manière générale, cette approche vise à établir des correspondances entre certains éléments ("features") contenus dans deux images. Dans le cas d'analyse de séquences d'images, ces éléments caractérisent les objets dont on désire étudier le mouvement.

Cette approche se conduit en deux étapes : dans la première, après avoir défini des éléments structurels caractéristiques des objets étudiés, ceux-ci sont recherchés dans les images de la séquence, c'est l'extraction des éléments structurels. Puis dans la deuxième étape, l'on recherche la correspondance physique entre les éléments extraits d'une image et ceux extraits de la suivante, c'est la mise en correspondance.

Ces deux étapes s'effectuent pour toutes les paires d'images successives de la séquence.

La correspondance peut être établie entre de simples éléments tels que points, segments de droite, surfaces, et éléments de contour. Elle peut également être établie à partir d'éléments plus complexes tels que le contour extérieur ou l'image complète de l'objet à retrouver lorsque celui-ci est parfaitement connu et relativement simple ("object-based methods"). Dans ce cas, une description plus ou moins fine de celui-ci est utilisée (sous forme de graphe par exemple).

L'utilisation d'éléments complexes à le mérite de réduire la quantité de données à traiter car un élément complexe a usuellement un nombre limité d'homologues dans l'image suivante, alors qu'un élément simple peut avoir de nombreux candidats lors de la mise en correspondance avec les éléments de l'image suivante. Cette dernière situation complique notablement le processus de mise en correspondance et peut même conduire à des erreurs.

#### **II.1.1.1-** Extraction d'éléments structurels

Un élément structurel est défini comme un ensemble de pixels (région, contour, coin, frontière, etc...) ayant entre-eux des relations particulières les différenciant de leurs voisinages. Plus ces relations sont caractéristiques de l'ensemble de points à repérer, meilleur est l'élément structurel.

Exploitant cette propriété, Moravec a développé un "opérateur d'intérêt" pour extraire des caractéristiques importantes (les coins essentiellement) [Mor77, Mor79, Mor83, Dav79]. On peut citer également parmi les opérateurs d'extraction de points anguleux statiques ceux de Kitchen-Rosenfeld [Ros75] et Dreschler-Nagel [Dre83].

Notons également le gabarit adaptatif de Potter [Pot75, Pot77], les maxima locaux de variabilité [Bar79, Bar80,Mor77, Mor79, Mor83], les contours locaux [Jac80, Mar77, Kor80], les sommets [Agg75, Asa80], les statistiques locales [Jai77,Sny81] et enfin les régions [Pri75, Pri77, Wil80].

Le choix de ces éléments structurels dépend de l'application et des hypothèses admises sur les objets recherchés. En effet, un choix judicieux peut atténuer, le cas échéant, les effets des variations d'illumination ou ceux des changements géométriques des objets ou de leur image (rotation, changement d'échelle).

Selon l'élément structurel recherché, on travaille soit directement sur l'image originale, soit sur une transformation globale de cette image telle que la différentiation spatiale (image de contours [Mai84, Law83]), la transformation de Hough (image de formes géométriques [Fen79]) ou la binarisation (image seuillée). Généralement, excepté dans le cas de l'utilisation de la transformée de Hough, un filtrage visant à supprimer le bruit de mesure est effectué car l'analyse structurelle y est très sensible.

Une technique originale assimilable à l'analyse structurelle est la morphologie mathématique [Bri81]. Outil récent du traitement d'images, ses développements sont encore assez limités car elle n'est vraiment séduisante que pour les images binaires. En effet, son utilisation sur des images à niveaux de gris nécessite d'énormes moyens de calcul.

#### **II.1.1.2-** Mise en correspondance

Le problème de la mise en correspondance, deuxième et dernière étape de l'approche, se résoud par le choix d'un critère qui permet d'établir la correspondance entre les éléments structurels de deux images successives. Les travaux d'Ullman sur le processus de mise en correspondance par le système visuel humain [Ull78, 79, 81] l'ont conduit à proposer un modèle algorithmique décrit brièvement ci-dessous :

- si les éléments structurels choisis sont des **points isolés**, la correspondance est établie uniquement selon la distance entre points. A autres paramètres égaux (niveau de gris, taille, etc...), chaque point de l'image est mis en correspondance avec le point le plus proche dans l'image suivante.

- si les éléments structurels choisis sont des segments, des règles supplémentaires interviennent. Au critère de proximité, s'ajoute la longueur et l'orientation des segments. A autres paramètres égaux, un segment sera mis en correspondance avec le segment de longueur et d'orientation les plus semblables. Le critère de proximité n'est pas appliqué dans ce cas.

La difficulté réside dans la construction d'une fonction métrique d'affinité qui combine les différents paramètres de l'élément structurel (ici longueur, orientation et position). Un exemple bien connu où cette fonction d'affectation du système visuel humain est mise en défaut est la roue à rayons au cinéma, où la rotation apparente est inverse de la rotation réelle (effet stroboscopique). Ce phénomène est lié à la préférence qu'a le système visuel humain à choisir entre deux mises en correspondance, celle qui entraîne les variations d'angles et de déplacements minimales, et ceci malgré la connaissance *a priori* du mouvement réel.

- enfin, si les éléments structurels choisis deviennent **plus complexes** (régions, groupes d'éléments, structures, etc...), la correspondance ne fait plus intervenir de simples fonctions métriques d'affinité, mais des processus de mise en correspondance de plus haut niveau qui prennent en compte un grand nombre de paramètres [Lev83, Roa79, Jac80].

De très nombreux critères optimaux ont été utilisés pour trouver la meilleure mise en correspondance selon l'élément structurel choisi. Ces critères vont de la simple fonction d'intercorrélation [Gen79,Yak78] à des procédures complexes à base de graphes [Jac80]. Des méthodes de relaxation ont également été utilisées avec succès et ce dans des scènes très variées [Bar80, Dre83, Pra79]. Celle de Rosenfeld, Hummel et Zucker [Ros76] est parfois utilisée pour une mise en correspondance par approximations successives. Pour une revue plus détaillée des différentes méthodes de mise en correspondance, on pourra se reporter à [Agg81].

Des méthodes qui recherchent les correspondances en introduisant une résolution variable de l'image [Gen77] ou l'introduction d'hypothèses telles que des limites aux variations possibles entre deux images successives (*e.g.* déplacements et taille [Bou88]), permettent d'accélérer le processus de mise en correspondance. Toutefois, la mise en correspondance reste très couteuse en calculs.

Lorsque les changements de luminance ou de géométrie des objets entre deux images sont trop importants, la plupart de ces méthodes de bas niveau échouent. Dans ce cas, il peut être avantageux d'utiliser une représentation symbolique des images [Pri75, Pri77].

#### II.1.2- Etude des variations locales de l'intensité

#### **II.1.2.1.-** Différence d'images

Les variations locales d'intensité dues au mouvement peuvent être mises en évidence de nombreuses manières [Nag86] allant de la simple soustraction [Jai81a, Jai81b] au calcul d'un rapport de vraisemblance basé sur des statistiques de deuxième ordre d'une petite région [Jai77, Jai79].

Cependant, la technique la plus employée, car la plus directe et la plus rapide, reste la soustraction pixel à pixel [Jai79, Jai79b, Lil72].

En éliminant tous les points stationnaires, ceci permet de concentrer ensuite les ressources du calculateur sur les points en mouvement. Cette méthode sera par conséquent d'autant plus intéressante que les zones occupées par des objets en mouvement dans l'image seront faibles.

De plus, étant une opération locale, la soustraction d'images nécessite très peu de temps de calcul.

Notons que la sensibilité au bruit de la méthode peut être réduite par la détection des régions connexes en mouvement suivie d'un filtrage dimensionnel [Jai81a, Ste85]. En effet, le bruit ne générant que des régions de petite taille, celles-ci sont

éliminées et ne subsistent alors dans l'image que les régions correspondant réellement à des objets en mouvement dans la scène.

La soustraction pixel à pixel de la fonction intensité f à l'instant mT peut se réaliser à un ordre quelconque. Ainsi, Jain et Nagel ont utilisé simultanément l'ordre 1 (D(m) = f(m)-f((m-1)) et 2 (D(m) = f(m)-2f(m-1)+f(m-2)) pour l'analyse de scènes de trafic urbain [Jai79a], alors que Snyder et Rajala [Sny83] utilisent l'ordre 3 (D(m) = f(m)-3.f(m-1)+3.f(m-2)-f(m-3)) pour un bon compromis filtrage/simplicité, en vue de la détection de cibles représentées par quelques pixels dans l'image.

Cette méthode connait hélas trois limitations importantes vis à vis des applications, à savoir :

- l'impossibilité de mesurer le mouvement 3-D.

- la nécessité d'une prise de vue à partir d'un poste fixe par rapport au fond.

- la nécessité d'une illumination constante entre l'acquisition des images intervenant dans le calcul.

De plus, la soustraction d'images n'assure pas la détection systématique de tous les points en mouvement. C'est pourquoi, si l'on désire garantir la détection de tous les points des objets en mouvement, ces techniques requièrent généralement des déplacements entre les images supérieurs aux dimensions des objets dans le sens des déplacements.

La simplicité de mise en oeuvre de la différence d'images basée sur une soustraction pixel à pixel a conduit cependant à de très nombreuses investigations visant à s'affranchir de cette contrainte. Nous développerons celles-ci dans le chapitre III.

#### II.1.2.2. Corrélation

Cette méthode compte parmi les plus anciennes techniques de mesure de mouvement entre des paires d'images. Elle a été appliquée à la mesure des mouvements des nuages [Lee70, Smi72, End71, Lee71], au contrôle de trafic [Jai79a, Jai79b] et au suivi d'un véhicule dans une scène réelle [Ber80, Riv86].

La corrélation est proposée en [Lee70, Smi72] pour une estimation du mouvement 2-D, en [Sny80] pour la segmentation en vue de la poursuite d'objets et en [Han74] pour la stéréovision.

Dans cette méthode, une fenêtre de la première image est intercorrélée avec la seconde image. Un pic de la fonction d'intercorrélation signifie une correspondance de l'image du même objet dans les deux images. Les coordonnées du pic donnent la valeur du déplacement. La fonction est calculée pour un certain nombre de fenêtres. L'ensemble des déplacements ainsi obtenus constitue le champ des déplacements.

Bien que très coûteuses en temps de calcul, la corrélation est assez utilisée car elle ne nécessite pas de connaissances à priori sur les objets en mouvement. C'est pourquoi elle est surtout adaptée aux objets déformables (nuages).

Seule la taille de l'objet à suivre est une information quasi indispensable car c'est elle qui détermine la taille de la fenêtre de calcul de la fonction d'intercorrélation. En effet, une fenêtre trop petite ne comporte pas assez d'information caractéristique de l'objet recherché alors qu'avec une fenêtre trop grande, les distorsions empêchent un calcul fiable.

Outre une sensibilité accrue aux changements d'échelle et d'orientation, la fonction d'intercorrélation définie précédemment est sensible aux changements de réflectance qui accompagnent souvent le mouvement. C'est pourquoi d'autres mesures de corrélation, telles que la corrélation centrée et la corrélation filtrée, ont été utilisées au prix d'un accroissement de la quantité de calcul.

Le temps de calcul peut être considérablement réduit en limitant :

- le domaine de recherche aux zones en mouvement ou à la zone où il est probable de trouver l'objet : après détermination de la position initiale, une prédiction des positions suivantes peut être effectuée au fur et à mesure à l'aide de l'estimation de la vitesse de déplacement ;

- les déplacements interimages par une fréquence d'acquisition élevée.

#### **II.1.2.3-** Gradient spatio-temporel

Cette troisième méthode, appelée également méthode différentielle, utilise les changements d'intensité en chaque point de l'image dans le temps et l'espace afin d'en déduire le déplacement des régions constituant l'objet.

<u>II-8</u>

Techniques de traitement de séquences d'images

Sous l'hypothèse que la luminance des points de l'objet ne varie, ni avec le temps, ni avec l'espace (c'est à dire le mouvement) on peut établir, dans le cas bidimensionnel et en chaque point l'égalité suivante :

$$\frac{\partial f}{\partial x}u + \frac{\partial f}{\partial y}v + \frac{\partial f}{\partial t} = 0 \qquad (2)$$

où f = f(x,y,t) est la fonction intensité, t le temps,  $\partial$  l'opérateur de dérivée partiel, u et v les composantes de la vitesse de chaque point selon les axes x et y respectivement.

L'équation (2) ne permet pas à elle seule de déterminer les deux inconnues u et v. C'est pourquoi, des contraintes supplémentaires variées ont été proposées parmi lesquelles : (a) le vecteur vitesse est continu ; (b) le vecteur vitesse est constant le long d'un segment ; (c) le vecteur vitesse est dû à un mouvement restreint (e.g. translation pure).

Plus récemment, une méthode prometteuse qui établit un système d'équations similaires à l'équation (2) en proposant pour *f*, outre la fonction intensité, diverses fonctions (fonctions multispectrales R-V-B, opérateurs spatiaux,...), a été proposée [Woh83, Mit84, Mit87].

Les méthodes différentielles supposent qu'il existe une fonction continue (pas nécessairement la fonction intensité f(x,y)), à dérivées définies et non-nulles, accessible à partir de l'image. Ces méthodes échouent bien sûr dans le centre des régions homogènes où le gradient est nul ou d'amplitude voisine à celle du bruit.

Outre ces problèmes, les méthodes basées sur le gradient spatio-temporel sont sensibles aux variations du gradient le long du contour, à celles des vecteurs déplacements et à celles de la luminance de l'objet. Divers auteurs proposent des techniques pour réduire l'influence de ces variations : Kearnes [Kea82] démontre qu'un filtrage [Yac83] ou une défocalisation [Fen79], de préférence adaptatifs, suivi d'un seuillage du gradient peut réduire cette sensibilité. Toutefois, ces opérations s'effectuent au détriment de la précision de la mesure.

Très récemment, Ma [Ma87] a proposé une équation basée sur la réflectance des objets permettant de traiter le cas des variations soit spatiales, soit temporelles de l'illumination, généralisant ainsi l'équation (2).

Conduisant directement au flux optique, le calcul du gradient spatio-temporel (cf. II.1.3) est très fréquemment utilisé pour l'estimation du mouvement 2-D ou 3-D.

#### II.1.2.4- Méthode de Fourier

Cette approche, oeuvrant dans le domaine fréquentiel, s'appuie généralement sur l'analyse de la phase de la transformée de Fourier pour évaluer les déplacements [Hua81, Has74].

Arkin l'utilise pour l'estimation du mouvement des nuages, quoique les méthodes de corrélation semblent mieux adaptées à ce genre de problème [Mar78].

Rajala, quant à lui, utilise une transformée de Fourier monodimensionnelle pour poursuivre des cibles de très faibles dimensions ("sub-pixel targets") dans un environnement bruité par détection des pics de la transformée de Fourier [Sny83, Raj83]. Chaque pic correspond à une cible mobile de vitesse proportionnelle à l'abscisse de celui-ci.

Malgré des temps de calculs réduits grâce à l'utilisation de circuits câblés de transformée de Fourier rapide ("FFT"), la méthode reste très peu employée. Les limitations principales résident dans la nécessité d'un fond uniforme et dans un temps de calcul prohibitif lorsque l'on considère les rotations et les changements d'échelles [Woh83].

#### II.1.3- Comparaison des deux approches

Globalement, chacune des deux approches présentées possède des avantages ("+") et des inconvénients ("-") spécifiques résumés dans le tableau II.1 qui montre les intérêts respectifs de l'analyse structurelle et de l'analyse des variations locales de l'intensité. Ce tableau est à nuancer selon la technique employée pour chacune des approches.

Il faut toutefois noter que, quelque soit l'approche utilisée, les résultats peuvent être obtenus sous deux formes très différentes qui sont :

- le champ des déplacements d'une part ;

- l'image différence d'autre part.
| Approche                          | analyse locale de l'intensité   | analyse structurelle   |
|-----------------------------------|---|--|
| Rapidité                          | ++ plus rapide  | plus lente   |
| Précision                         | - à cause de l'ambiguïté<br>due aux mesures locales                   | + surtout si la structure<br>est bien repérable  |
| Robustesse                        | - Nombreuses sources d'erreurs<br>sensibilité au bruit                | + plus l'élément structurel<br>est complexe plus rapide est la mise en<br>correspondance               |
| Volume de<br>données              | - élevé   | + réduit, concentré sur les<br>structures  |
| Adaptabilité                      | ++ peu ou pas de connaissances<br>à priori nécessaíres                | connaissances à priori sur les objets<br>mobiles nécessaires   |
| Déplacements<br>autorisés         | - doivent être faibles<br>sauf pour la différence d'images            | + peuvent être importants  |
| Changements de<br>formes autorisé | + importants<br>s   | t variable selon l'élément<br>structurel choisi  |
| Utilisations                      | - détection de mouvement, segmentation<br>estimation du mouvement 2-D | + détection de mouvement, segmentation<br>mouvement 2 & 3-D, structure 3-D,<br>reconnaissance d'objets |
| Limites                           | - bords, surfaces<br>d'intensité uniforme                             | - élément structurel désiré<br>non extractible   |

# Tableau II.1 : Comparaison des deux approches

La figure II.2 indique à quelle forme on aboutit en fonction de la technique utilisée pour chacune des approches.

# II.1.4- Flux optique et image différence

#### II.1.4.1- Le flux optique

Le champ des déplacements est l'image obtenue par projection du vecteur déplacement de chaque point de la scène sur le plan de l'image.

Le flux optique est la projection des vecteurs vitesse des points de la scène sur le plan image [Law83]. Par conséquent, on l'obtient en divisant chaque vecteur du champ des déplacements par le temps séparant les images qui ont permis sa détermination.



Figure II.2 : Principales techniques de traitement d'une séquence d'images

Les méthodes conduisant au flux optique nous placent en présence du paradoxe suivant : pour calculer les vitesses instantanées, il faut que les déplacements entre les deux images soient faibles. Or dans ce cas, le bruit contenu dans l'image est du même ordre de grandeur que les déplacements, ce qui fausse le calcul du flux optique.

Notons qu'une opération de lissage (ou de défocalisation) autorise une légère augmentation des déplacements mais entraîne une perte de précision [Fen79].

Cette contrainte de petits déplacements constitue la principale limitation des méthodes du gradient spatio-temporel.

Le calcul du flux optique se détermine généralement en deux étapes :

1) calcul du vecteur vitesse en certains points de l'image qui peuvent être des points particuliers tels que les coins [Yac83, Dav79], des structures [Pot75], des contours [Spo83, Bur82, Fen79, Kor80] ou des régions [Tho80, Fen79].

II-12

Selon la scène étudiée, cette étape est effectuée soit par analyse structurelle, soit par des techniques basées sur l'analyse des variations locales d'intensité telles que la corrélation [Sny80], la transformée de Fourier [Has72] ou le gradient spatio-temporel [Tho80, Hor81, Tho81].

2) extrapolation de ce vecteur aux autres points des objets mobiles [Dav79, Tho80, Yac83, Bou86a-b-c].

Lorsque l'on emploie la méthode du gradient spatio-temporel, le calcul peut s'effectuer en chaque point de l'image [Sny80, Spo83], ce qui conduit directement au flux optique [Clo80, Praz79].

Cependant, le calcul du vecteur vitesse en chaque point de l'image est inévitablement entaché d'erreurs. Par conséquent, il est nécessaire de combiner les mesures locales afin d'obtenir des valeurs représentatives du vecteur vitesse en chaque point. Pour cela, les méthodes suivantes sont fréquemment employées :

- les méthodes d'optimisation locale [Sny80, Lim75, Luc81] qui minimisent un terme d'erreur pour résoudre un ensemble d'équations linéaires telles que la méthode des moindres carrés [Gla81];

- les méthodes d'optimisation globale : Horn et Schunk furent les premiers à utiliser ces méthodes en analyse de mouvement [Hor81]. Sous l'hypothèse que les changements d'intensité sont continus (ce qui n'est pas vérifié sur les contours), ils combinent le gradient spatio-temporel avec une procédure d'optimisation itérative pour estimer le vecteur de déplacement en chaque point de l'image.

- la méthode d'agrégation qui vise à regrouper les points voisins de même vitesse [Caf76,Fen79,Tho80].

Le calcul du flux optique effectué par des techniques de transformée de Fourier [Has72] ou de corrélation [Sny80] est très coûteux. C'est pourquoi, on préfère généralement les méthodes différentielles ou méthodes du gradient spatio-temporel [Hor80, Tho80].

# II.1.4.2- L'image de la différence

L'image différence permet d'extraire les points en mouvement de la scène sans aucune évaluation de la direction, du sens, et de l'amplitude de leurs déplacements.

C'est pourquoi, son interprétation sera assez limitée puisqu'elle ne permettra pas de déduire directement les différentes composantes du mouvement des objets mobiles.

Cependant, en détectant tous les points en mouvement y compris ceux des régions uniformes, l'image différence permet une extraction immédiate de l'image de chaque objet mobile ainsi que l'estimation de leur position dans le plan de l'image.

Si besoin est, l'estimation du mouvement 2-D (tout au moins la translation du centre de gravité dans le plan de l'image) peut être déduit par une mise en correspondance des zones en mouvement dans leurs positions successives.

Par conséquent, l'image différence est plutôt utilisée pour la détection de mouvement, la segmentation et l'estimation du mouvement 2-D.

Fondamentalement différents, le flux optique et l'image différence seront par conséquent interprétés et exploités différemment. Comme la figure II.3 l'indique, le flux optique autorise une exploitation beaucoup plus étendue que l'image de la différence. La figure indique également les relations existants entre ces diverses exploitations.

# **II.2- ANALYSE DE SEQUENCES D'IMAGES**

Les résultats obtenus par les techniques de traitement des séquences d'images exposées dans la première partie de ce chapitre ne constituent pas une finalité. En fait, les séquences d'images doivent être analysées afin de répondre à un besoin précis dans le cadre de chaque application. Selon l'objectif recherché, on peut se contenter de détecter des mouvements, mais on peut également segmenter les images dynamiques pour en extraire les objets mobiles. Un objectif réaliste mais plus difficile à atteindre concerne l'estimation des paramètres qui définissent les mouvements de chacun des objets mobiles dans la scène. L'objectif le plus ambitieux demeure cependant la reconnaissance des objets mobiles et des objets non-rigides ou articulés.

Nous allons maintenant examiner comment ces techniques de traitement permettent de répondre à ces différents objectifs.





II-15

#### II.2.1- Détection du mouvement

La détection du mouvement revêt une importance particulière lorsque la présence d'objets en mouvement constitue un danger.

Au niveau des applications de la détection de mouvement, citons la conduite des véhicules autonomes et la surveillance. Pour la première application, la détection du mouvement signifie la présence d'un obstacle mobile à proximité d'un véhicule autonome. Pour les tâches de surveillance, la détection d'un mouvement permet de déclencher une alarme. Par exemple, dans le cas d'un système à base de plusieurs caméras, la détection automatique d'un mouvement sur l'une d'entre elles permettra au surveillant d'y porter son attention.

Notons que ces applications sont tout de même assez réduites. Ceci s'explique par la pauvreté de l'information déduite de l'analyse de la séquence d'images, cette tâche étant avantageusement effectuée par des capteurs classiques moins onéreux (ultrasons, cellules, télémètres, etc...).

La méthode la plus employée pour la détection du mouvement reste la soustraction pixel à pixel car elle est la plus directe et la plus rapide [Jai79a, Jai79b, Jai81a, Lil72]. Cette méthode est très utilisée comme "signal d'alarme" [Jai79, Jai81b, ].

La transformée de Fourier a été utilisée dans le cas très particulier d'objets représentés par un ou quelques pixels noyés dans le bruit [Sny83, Raj83].

Une troisième approche s'effectue en deux temps : dans le premier temps on détermine un modèle paramétré des variations des niveaux de gris dans un voisinage de certains points. Dans le second temps, des tests statistiques établissent si les descriptions, calculées pour deux images consécutives aux mêmes points, sont compatibles ou non [ Jai77, Jai79a, Nag78]. Les zones d'incompatibilité seront alors identifiées comme des zones perturbées par un mouvement.

Enfin, une approche peu utilisée en détection de mouvement est le gradient spatio-temporel car il nécessite une cadence d'acquisition des images élevée, de nombreuses hypothèses sur le mouvement et des calculs assez importants [Fen79, Fri72, Hor81, Lim75].

#### **II.2.2-** Segmentation

La détection des zones affectées par le mouvement permet une segmentation "binaire" de la scène. Cependant, elle ne permet pas toujours d'établir une correspondance univoque entre zones en mouvement et objets mobiles. En effet, plusieurs zones en mouvement détectées peuvent correspondre au même objet et à l'opposé, une zone en mouvement peut correspondre à deux objets connexes ou paraissant connexes (phénomène d'occlusion).

Si l'on dispose d'une certaine connaissance *a priori* de la scène (nombre d'objets, dimensions, distances entre objets, niveaux de gris et textures des objets, etc...), une segmentation plus fine peut être obtenue en appliquant aux zones en mouvement un algorithme de segmentation de scènes statiques tels que ceux décrits en [Yac80, Bar86, Man87].

Ainsi, Jain utilise des mesures de niveaux de gris pour la segmentation des zones en mouvement détectées par **différence d'images** [Jai79a, Jai79c]. La différence d'images est également utilisée pour séparer le fond des objets mobiles en [Yal80] et [Lim75]. En [Yal80], le processus de segmentation est basée sur une technique d'accroissement des régions ("region growing") détectées.

Korn et Kories [Kor80] segmentent une scène naturelle (véhicule circulant sur fond d'immeubles) grâce à une mise en correspondance de portions de contours.

D'autres auteurs distinguent les différentes régions ayant subi un mouvement par des techniques de statistiques locales [Jai77, Jai79a, Nag78].

Cafforio utilise quant à lui un algorithme de programmation dynamique (algorithme de Viterbi) pour détecter les zones en mouvement [Caf76].

Si en revanche, on ne dispose pas d'information *a priori* suffisante, alors la simple détection des zones en mouvement ne permet pas une segmentation correcte de chaque objet mobile.

On sait que dans les systèmes visuels biologiques, les variations temporelles d'une scène apportent des indices sur ce qui appartient à un même objet ou non. Des expériences ont montré que le système visuel humain utilise la vitesse de parties d'objets mobiles ne présentant ni contraste, ni changement de texture avec le fond, pour segmenter ceux-ci. L'illustration est donnée par le dessin animé d'un dalmatien se déplaçant sur un fond ayant le même motif que son pelage. Lorsque le chien est à l'arrêt, si l'observateur n'est pas averti de sa présence, il ne détecte rien dans l'image. Par contre dès que le dalmatien se déplace, l'observateur est tout de suite capable d'identifier un chien.

Inspirés par ces expériences, de nombreux chercheurs ont développé des techniques de segmentation sous l'hypothèse que les éléments de la scène qui ont la même vitesse appartiennent au même objet (sous-entendu rigide) [Lim75, Pot75a, Pot75a, Pot75a, Pot75a, Pot77, Nag78, Ull79, Jai79c, Tho80, Sny80]. Sous certaines conditions telles que mouvement de pure translation, mouvement parallèle au plan de l'image ou effet de perspective négligeable, l'hypothèse précédente est généralement vérifiée.

Par conséquent, l'analyse directe du flux optique [Tho80, Fac87] ou par l'intermédiaire de sa transformée de Hough [Fen79, Neu80, Oro81] permet de segmenter les différents objets mobiles.

Ainsi, Thompson suggère une procédure combinant l'intensité et la vitesse [Tho80]. Adiv utilise également le flux optique pour segmenter une scène contenant plusieurs objets selon une procédure d'hypothèses et de tests. La méthode a été appliquée avec succès sur des images synthétisées et réelles [Kap83].

La célèbre transformée de Hough associée à des techniques de "clustering" [Fen79, Oro81] permet de détecter les modes d'histogrammes multidimensionnels correspondant aux mouvements majeurs de l'image. Le principal avantage de cette méthode est la possibilité de manipuler plusieurs mobiles simultanément, chaque mode de l'histogramme correspondant à un mobile. Cependant, ceci nécessite une relative uniformité du flux optique. Cette condition est vérifiée lorsque la surface des objets mobiles est importante.

De nombreux auteurs utilisent cette Transformée de Hough car la segmentation est assez précise et peu sensible au bruit [Caf76, Tho81, Neu80, Fen79, Oro81, Tho80, Kor80, Sny83].

En autorisant les erreurs au niveau des correspondances (une grande majorité de correspondances exactes suffisent), la méthode ne nécessite pas l'emploi de techniques de correspondance sophistiquées telles qu'en [Ras79, Ull79] pour conserver une bonne insensibilité au bruit. De surcroît, la méthode permet de traiter le problème très délicat des occlusions partielles.

Une approche récente et intéressante consiste à analyser directement le champ des déplacements par une méthode de coalescence non supervisée : la méthode des nuées dynamiques. Cette nouvelle approche semble prometteuse, mais bute sur le problème des mouvements de rotation et nécessite des calculs importants [Fac87].

#### **II.2.3- Estimation du mouvement**

De nombreuses applications nécessitent une estimation du mouvement. Ce domaine a suscité de nombreux travaux qui peuvent être divisés en trois axes de recherche : l'estimation du mouvement 2-D, l'estimation du mouvement 3-D et la navigation passive (conduite automatique).

#### II.2.3.1- Estimation du mouvement 2-D

L'estimation du mouvement 2-D consiste en la détermination des paramètres définissant le mouvement des objets de la scène dans le plan de l'image. Ces paramètres sont des paramètres de translation (direction, sens, amplitude) et des paramètres de rotation (centre, angle, rayon). La vitesse et l'accélération des objets peuvent éventuellement être déduites de ces déplacements si l'on connait le temps (généralement constant) séparant les images successives.

De simples soustractions d'images peuvent être utilisées pour estimer les paramètres de mouvement 2-D [Jai79a, Jai79b, Jai79c]. Jain et Nagel utilisent cette approche pour analyser des scènes de trafic [Jai79a]. Quoique le mouvement des véhicules et des piétons puisse être composé d'une rotation et d'une translation, seule cette dernière est estimée.

La technique du gradient spatio-temporel est également employée sous diverses formes pour estimer les paramètres du mouvement 2-D par l'intermédiaire du calcul du flux optique [Fen79, Fri72, Hor81, Lim75, Lon80, Pra81].

En réalisant la poursuite d'éléments structurels d'une image à l'autre, on peut calculer la translation d'une surface [Jer84, Lee70, Mor77].

Chien et Jones proposent un système de poursuite temps réel d'automobiles s'appuyant sur des répartitions caractéristiques de niveaux de gris [Chi75].

Analyse de séquences d'images

Davis et *coll.* estiment le mouvement de translation par des mesures d'intensité aux coins [Dav79]. Pour obtenir une solution plus précise, les estimations du mouvement des coins sont ensuite extrapolées le long du contour principal de l'objet sous l'hypothèse d'un mouvement local  $2^{1}/_{2}$ -D.

Jain, Militzer et Nagel mettent en correspondance des surfaces de différentes images de la séquence par comparaison de leurs statistiques de premier et second ordre [Jai77, Jai79a, Nag78].

Citons enfin, Fennema et Thompson [Fen79], Snyder et Rajala [Sny83], Adiv [Adi83] qui, en utilisant la **transformée de Hough** généralisée, calculent directement les paramètres de mouvement 2-D dans des scènes contenant plusieurs objets mobiles.

#### II.2.3.2- Estimation du mouvement 3-D

La poursuite d'objets se déplaçant dans un espace qui ne peut être considéré comme bidimensionnel nécessite la détermination de vecteurs de translation et de rotation tridimensionnels. Par conséquent, les paramètres de mouvement à déterminer sont au nombre de cinq : trois pour la rotation et deux pour la translation (un des paramètres de translation pouvant être choisi arbitrairement). Par conséquent, il faut établir au moins cinq équations pour parvenir à une solution qui n'est pas nécessairement unique [Nag79, Nag81a, Nag81b, Pra81b].

S'il est aisé d'établir les équations à partir des relations liant le mouvement 3-D de l'objet dans l'espace à celui de son image, il est en revanche très difficile de les résoudre à cause de leur caractère essentiellement non linéaire. De surcroît, les solutions sont généralement très sensibles au bruit de mesure.

D'autre part, les informations de profondeur (distances selon l'axe optique) étant perdues lors de la projection sur le plan de l'image, le vecteur translation ne peut être déterminé qu'à un facteur d'échelle près à moins de connaître la profondeur d'au moins un point de l'objet. De plus, la solution n'est pas unique puisqu'elle dépend de l'origine du référentiel choisi. En revanche, les paramètres de rotation peuvent être théoriquement déterminés de façon unique puisqu'ils sont indépendants de ce référentiel.

Les inconnues du système étant au nombre de cinq, la plupart des chercheurs estiment qu'il suffit de cinq points dans deux images ou quatre points dans trois images [Gen79b, Law80, Mei80, Roa80, Ull79] pour le résoudre . Seul Shariat propose une méthode ne nécessitant que trois points dans trois images [Sha86].

Meiri a établi les relations entre le nombre d'inconnues, le nombre de points de l'objet, et le nombre d'images pour aboutir à une solution unique [Mei80] tandis que Fan et Huang ont développé un théorème sur l'unicité de la solution [Fan84].

Les techniques de résolution du problème sont largement dépendantes de la sa formulation. Dans le cas général, les équations du mouvement établies ci-dessus étant non linéaires, leur résolution est par conséquent une tâche ardue et passe presque toujours par l'emploi de méthodes itératives. Toutefois, la résolution d'équations non linéaires par une telle méthode soulève le problème de la convergence qui dépend de l'algorithme employé et de celui de l'unicité de la solution qui dépend de la formulation initiale.

Certains auteurs résolvent le système d'équations analytiquement en effectuant diverses hypothèses qui permettent généralement d'aboutir à un système d'équations linéaires. Ces hypothèses sont souvent des hypothèses restrictives sur le mouvement : pure translation [Clo80, Lee80] ; simple rotation autour d'un axe vertical parallèle au plan de l'image suivi d'une translation [Ull79] ; combinaison de translations et de rotations autour d'un axe fixe sur plusieurs images [Web81]. L'hypothèse de projection

parallèle (absence d'effet de perspective lorsque l'objet est suffisamment loun également employée [Ull79, Web81].

En supposant que la correspondance de cinq points ou plus ait pu être établie entre deux images, Nagel développe des relations pour le calcul du vecteur de translation et la matrice de rotation [Nag81a, Nag81b].

De manière voisine, Prazdny utilise des considérations d'ordre géométrique pour déterminer la direction instantanée du mouvement à partir du flux optique instantané, observé en un minimum de cinq points de l'objet [Pra80].

Dans le cas fréquent d'objets rigides, le fait que la distance entre deux quelconques de leurs points soit constante permet d'établir des équations supplémentaires [Law80]. Utilisant cette hypothèse de rigidité, Tsai et Huang [Hua81b, Tsa81] aboutissent à un système linéaire de huit équations. L'approche proposée est malheureusement extrêmement sensible aux mesures extraites des images : une erreur de 2 % sur ces mesures entraîne une erreur de 30 % sur les paramètres de rotation et de 95 % sur ceux de translation. L'hypothèse de rigidité est également appliquée par Roach et Aggarwal à des images de cubes mobiles [Roa80] et par Asada et *Coll*. qui analysent en plus le mouvement complexe d'objets connexes [Asa80].

En supposant de **petites rotations** (inférieure à 5°), Adiv établit un système d'équations reliant le mouvement 3-D de l'objet et le flux optique 2-D instantané correspondant [Adi85]. Les paramètres de translation sont calculés par approximations successives alors que ceux de rotation sont obtenus par la résolution d'un système d'équations linéaires. La méthode proposée est sensible au bruit du fait de l'analyse de faibles déplacements. C'est pourquoi Medioni et Yasumoto se sont attachés à améliorer la robustesse de cette méthode [Med85].

Lorsque l'information profondeur est connue, une formulation particulièrement aisée est alors possible [Jac80]. Aggarwal et Magee obtiennent rapidement cette information en quelques points caractéristiques de l'objet (extraits par analyse structurelle) par l'emploi d'un télémètre. Après avoir établi un modèle de l'objet sous forme de graphe grâce à différentes vues, ils estiment alors les paramètres de mouvement par comparaison du modèle courant au modèle établi [Nag79, Agg86]. O'rourke [Oro80] ainsi que Wallace [Wal80] proposent une approche également basée sur un modèle préétabli de l'objet grâce à une connaissance à priori sur celui-ci.

Citons aussi Ballard et Kimball qui estiment les paramètres de rotation et de translation en utilisant trois images et l'information profondeur [Bal83].

O'Rourke suggère une approche pour résoudre simultanément les problèmes de la mise en correspondance, de la segmentation, de la détermination de la structure et de l'estimation du mouvement. Cette approche, basée sur la **transformée de Hough**, permet l'estimation simultanée du mouvement de plusieurs objets et est peu sensible au bruit [Oro81].

Toutefois, cette méthode nécessite une place mémoire importante liée au grand nombre de dimensions de l'espace des paramètres, ce qui réduit ses possibilités d'application pour résoudre le problème du mouvement 3-D en général [Oro81].

Signalons également la tentative de Ballard et Kimball de décomposer le problème en plusieurs sous-problèmes résolus séquentiellement [Bal83].

# **II.2.3.3-** Navigation passive

La navigation passive consiste à estimer le mouvement de l'observateur par rapport aux éléments de l'environnement, mobiles ou non. L'application directe est évidemment la conduite (navigation) automatique de véhicules autonomes (observateurs).

Les changements apparents de la position des éléments composant la scène résultant du mouvement de l'observateur ("parallax motion") sont une source importante d'information. Cette information permet de déterminer le mouvement de l'observateur et de construire un modèle de l'environnement afin d'éviter les obstacles.

Eskenazi et Cunningham utilisent le système de vision du robot du "Jet Propulsion Laboratory " pour la poursuite temps réel d'objets dans des séquences d'images stéréo et déterminent ainsi la trajectoire du robot afin d'éviter les obstacles [Esk78]. Dans une approche similaire, afin de piloter son robot mobile, le "Stanford Cart ", autour d'obstacles, Moravec obtient des images stéréos à partir d'une seule caméra qui peut être décalée perpendiculairement à son axe optique selon des déplacements bien quantifiés [Mor77, Mor79, Mor80].

Prazdny fait valoir qu'en utilisant le **flux optique** produit par le mouvement de l'observateur, on peut calculer cinq des six paramètres définissant ce mouvement [Lon80] : ce sont les trois paramètres de vitesse angulaire et deux des trois paramètres du vecteur vitesse (sens et direction). L'amplitude de ce dernier n'est pas estimable sans la connaissance de la distance séparant un point quelconque de l'environnement de l'observateur.

La formulation et la solution au problème de navigation passive sont similaires, dans une certaine mesure, à ceux de l'analyse du mouvement 3-D. En effet, la plupart des équations nécessaires pour l'analyse du mouvement 3-D sont déduites du mouvement relatif entre l'objet et la caméra.

#### II.2.4- Reconstitution de formes 3-D

Le flux optique produit par un objet en mouvement contient suffisamment d'informations pour permettre la détermination de sa forme tridimensionnelle à un facteur d'échelle près. De la même manière, le flux optique induit par le mouvement de l'observateur peut être utilisé pour construire un modèle 3-D de l'environnement.

Les techniques permettant de reconstituer la structure des objets s'appuyant sur le mouvement ont fait l'objet de nombreuses investigations : des algorithmes ont été développés et les conditions nécessaires ont été étudiées afin de déterminer la profondeur, c'est à dire les dimensions des objets selon l'axe optique pour en déduire leur structure visible [Lon80, Wil80].

#### II.2.4.1- Profondeur de l'objet

Connaissant le mouvement de la caméra entre deux acquisitions, la profondeur des éléments visibles de la scène dans les deux images peut être calculée et un modèle de l'environnement peut être établi de manière très similaire à celle utilisée en stéréovision. En effet, les changements produits par le mouvement de la caméra entre deux images sont similaires à ceux obtenus par stéréovision.

Par exemple, Moravec a développé un programme de navigation passive pour le "Stanford Cart" qui peut construire une carte de l'environnement, indiquant la position, la profondeur, et la hauteur approximative de tous les obstacles visibles [Mor77, Mor79, Mor80].

#### II.2.4.2- Structure de l'objet

Le problème de retrouver la structure d'un objet à partir du mouvement ("Structure From Motion") et celui de l'analyse du mouvement 3-D proprement dit sont tellement imbriqués que la plupart de leurs solutions sont obtenues simultanément par la résolution d'un système d'équations [Bal83, Dre81, Nag81a, Nag81b, Neu80, Oro81, Pra80, Mit86].

Rappelons cependant que l'on ne peut en aucun cas déterminer la taille de l'objet, quelque soit le nombre de vues disponibles, sans connaître la distance le séparant de la caméra.

# II.2.5- Reconnaissance d'objets non-rigides

Ce dernier domaine de l'analyse de séquences d'images est encore assez peu développé. Il concerne la reconnaissance d'objets non-rigides ou articulés à partir de l'étude de leurs mouvements. Les recherches concernent essentiellement l'analyse des **mouvements humains**.

On sait que l'homme est capable de reconnaître certains objets uniquement d'après les caractéristiques de leurs mouvements. Dans les expériences de Johansson, un observateur réussi à deviner les attitudes d'une personne plongée dans le noir uniquement par les mouvements de 13 points lumineux fixés à ses articulations [Joh73]. Des études ultérieures montrent que même une détermination du sexe est possible selon ce même protocole [Kos77].

Fort de ces expériences, Rashid tente le premier une interprétation par ordinateur de telles scènes. Pour cela, il mesure la distance séparant dans le plan de l'image chaque paire de points puis teste la cohérence de cette distance sur un grand nombre d'images [Ras80]. Webb et Aggarwal [Web81b] utilisent une méthode similaire mais interprètent l'image de manière tridimensionnelle.

O'rourke et Badler emploient un modèle précis de silhouette humaine pour interpréter les mouvements [Oro80]. L'utilisation d'un tel modèle permet d'interpréter la structure complète de la silhouette même lorsque certaines de ses parties sont cachées (ex : main derrière le dos).

Clocksin a développé une méthode heuristique pour retrouver la structure d'objets articulés [Clo80]. Huffman et Flinchbaugh utilisent l'hypothèse que le mouvement se fait selon un plan pour retrouver à la fois la structure des articulations et la structure tridimensionnelle des objets [Huf80].

# **II.3- CONCLUSION ET CHOIX D'UNE METHODE**

Nous avons montré comment les différents objectifs de l'analyse de séquences d'images sont en liaison avec les différentes techniques de traitement, par l'intermédiaire du flux optique et de l'image de la différence. Etant donné l'aspect temps réel que nécessite notre application, parmi toutes les méthodes décrites précédemment seules celles basées sur la différence d'images semblent satisfaisantes (cf. II.1.2.1). Les rares systèmes fonctionnant actuellement en temps réel basés sur d'autres méthodes concernent tous des scènes simples : cible unique, cibles connues, cibles multiples mais semblables, cibles de textures différentes de celle du fond, orientations des objets peu variables, objets lents par rapport à la cadence d'acquisition, etc....

Dans notre application, les méthodes basées sur la recherche d'éléments caractéristiques ne sont pas exploitables car la variété des mobiles ainsi que de leur orientation par rapport à la caméra (quelle soit embarquée ou fixe) ne permet pas de définir, pour chacun d'eux, un ensemble d'éléments structurels pouvant les différencier des éléments statiques de la scènes.

De plus, notre choix de stratégie de prise de vue nous place dans les conditions quasi-idéales d'application des méthodes basées sur la différence d'images : caméra fixe, extraction du mouvement 2-D. Seules les contraintes de l'illumination constante et de déplacements minimaux des objets entre l'acquisition des images successives pourrait éventuellement remettre en question ce choix.

Afin de tenter de lever les limitations des méthodes éprouvées de différence d'images face à ces contraintes, il est nécessaire d'étudier celles-ci dans le détail et d'en bien comprendre le fonctionnement. Ce travail nous permettra de développer une procédure de détection des contours des objets en mouvement sur fond non uniforme et en éclairage non contrôlé. C'est à quoi s'attache le chapitre suivant.

# - CHAPITRE III -

# DETECTION DES CONTOURS D'OBJETS EN MOUVEMENT SUR FOND NON UNIFORME ET EN ECLAIRAGE NON CONTROLE

La segmentation de scènes dynamiques basée sur la différence d'images peut être abordée de deux manières différentes selon que l'on cherche à extraire les régions en mouvement ou les contours en mouvement.

La première approche est exposée en détail dans la première partie de ce chapitre.

Après en avoir exposé ses limites, on aborde dans la deuxième partie du chapitre l'approche par détection des contours en mouvement. Une nouvelle procédure, permettant de mettre en évidence le mouvement sur fond non uniforme et en éclairage non contrôlé est alors proposée.

#### **III.1- EXTRACTION DES REGIONS EN MOUVEMENT**

Sous les hypothèses suivantes : caméra fixe et illumination constante, la différence point à point entre deux images acquises à des instants différents permet d'isoler les points affectés par le mouvement entre ces deux acquisitions. Ces points sont généralement groupés en régions correspondant, soit aux objets en mouvement en entier, soit à différentes parties de ces objets.

Selon le choix des images opérandes constituant la différence, l'interprétation à donner aux zones isolées sera différente. On distinguera donc deux variantes de la méthode : celle utilisant une image de référence et celle utilisant plusieurs images successives. L'objectif reste dans les deux cas d'isoler les objets en mouvement de l'image courante, notée C.

# III.1.1- Avec image de référence.

Dans la méthode basée sur la différence d'images, l'idéal est de pouvoir disposer d'une image de référence, notée R. Cette image doit posséder les caractéristiques suivantes :

- être exempte de tout objet mobile ;

- avoir été acquise sous les mêmes conditions d'éclairement que les images avec lesquelles elle sera comparée.

L'image D, résultat de la différence entre cette image R et l'image courante C, est constituée de deux types de régions (voir planche III.1). Le premier type (type 1) est formé de points résultants de la différence entre des points des objets en mouvement dans leurs positions courantes et des points du fond correspondants de l'image R. Le deuxième type (type 2), caractérisé par un niveau de gris très faible sinon nul, correspond aux points stationnaires et dont les niveaux de gris sont très voisins dans les deux images C et R.

Les niveaux de gris des régions de type 2 sont d'autant plus faibles que la variation d'éclairement l'a été entre les deux images. Par conséquent, dans le cas de faibles variations, un simple seuillage permet d'extraire les objets mobiles. En effet, l'histogramme de l'image différence comporte un mode important dans les faibles valeurs de niveaux de gris qui correspondent aux éléments stationnaires de la scène et la valeur d'un seuil permettant de les isoler peut être aisément déterminée.

Cependant, la stabilité de l'éclairement ne peut être garanti dans les scènes extérieures car la lumière naturelle présente des variations aussi bien temporelles (lever et coucher du soleil, ciel nuageux avec éclaircies, etc...) que spatiales (ombres portées d'objets ou de nuages, reflets, etc...).

Il est également parfois impossible d'avoir accès à une image ne contenant aucun élément mobile. C'est pourquoi on a souvent recours à des procédures permettant de se rapprocher du cas idéal exposé ci-dessus. Les paragraphes ci-dessous décrivent les procédures utilisées lorsqu'on se trouve confronté soit à des variations d'éclairage, soit à des problèmes lors de l'acquisition d'une image de référence sans objet mobile. Ces méthodes sont répertoriées sous le nom des chercheurs les ayant introduites et selon l'ordre de leur apparition dans la littérature.



Planche III.1 : Différence d'images avec image de référence.

# III.1.1.1- Eclairement non constant.

# Aoki [Aok84]

Face à des variations continuelles d'éclairage de faible amplitude, Aoki propose une actualisation de l'image de référence. L'étape de binarisation de l'image D est effectuée selon deux seuils distincts : le premier, élevé, isole les points correspondant aux mouvements ; le second, très faible isole les points affectés par un changement de luminance du fond et non par le mouvement. Ceux-ci sont utilisés pour actualiser les points correspondant de l'image R, qui constitue ainsi la nouvelle image de référence.

Cette méthode est appliquée au comptage de véhicules sur une voie rapide. L'auteur utilise une caméra ligne dont le champ de vision est une ligne perpendiculaire à la voie. Le nombre limité de pixels sur cette ligne permet ainsi un traitement temps réel.

# III.1.1.2- Impossibilité de disposer d'une image de référence exempte d'éléments mobiles

# Dreschler et Nagel [Dre78].

Dreschler et Nagel proposent de considérer initialement deux images dans lesquelles le déplacement de l'objet est tel que ses positions respectives dans chacune des images ne se chevauchent pas. La première acquisition est alors considérée comme image de référence pour les images suivantes.

On somme alors les valeurs des points de contour des régions de type 1 qui apparaissent dans l'image D, en se référant à l'image C. Si la valeur obtenue est importante, c'est que la région correspondante fait partie d'un objet en mouvement dans l'image C.

Cette approche suppose que les niveaux de gris des points de contour du fond restent faibles par rapport à ceux des objets en mouvement.

#### Tang et Snyder [Tan82]

La méthode proposée par Tang et Snyder afin d'extraire les objets en mouvement d'une scène réelle utilise des sous-images appelées fenêtres de référence, obtenues de la manière suivante. La différence entre les deux premières images successives est seuillée. Les régions de type 1 ainsi détectées et supposées générées par le mouvement d'un même objet (selon un critère de proximité généralement), sont encadrées par un rectangle qui constitue ainsi une fenêtre de référence. Tant que l'objet n'est pas totalement sorti de cette fenêtre, il apparaîtra, dans la différence entre les images successives suivantes, des régions de type 1 incluses dans cette fenêtre. Dans le cas contraire, c'est que l'objet sera totalement hors de la fenêtre et ne pourra alors disposer du fond initialement recouvert par cet objet dans la fenêtre et reconstruire ainsi une image de référence ne contenant que des objets stationnaires. La différence entre cette image et la dernière image courante permettra alors l'extraction de la totalité de l'objet mobile.

La méthode proposée peut poser des problèmes lorsque la scène contient plusieurs objets en mouvement dans les situations suivantes :

- les objets ne sont pas suffisamment éloignés les uns des autres pour qu'un simple critère de proximité permette de déterminer la fenêtre de référence relative à chacun des objets (ambiguïté sur les zones à encadrer par un même rectangle); - un objet apparaît dans la fenêtre de référence associée à un autre objet ("running occlusion") empêchant l'extraction de ce dernier tant que le premier n'a pas complètement quitté la fenêtre.

# Jain [Jai84].

Basée également sur la notion de fenêtre de référence, la méthode proposée par Jain introduit les notions d'image des différences cumulées positives (DCP) et négatives (DCN). La première image est considérée comme référence. Les points positifs de l'image D sont ajoutés sous forme binaire à la DCP, tandis que les points négatifs le sont à la DCN. On itère le processus jusqu'à ce que l'une des images des différences cumulées (positive ou négative, selon que le niveau de gris de l'objet est supérieur ou inférieur à celui du fond) cesse de s'étendre en surface alors que l'autre continue de croître. A ce moment, on peut affirmer que l'objet à totalement quitté sa position initiale. Les points de l'image de référence correspondant à ceux de l'image des différences cumulées qui a cessé de croître sont alors remplacés par les points correspondants de l'image courante, ces points étant des points d'éléments stationnaires. Les points de l'image des différences cumulées qui continue de croître, incrémentés lors de la dernière itération, sont des points de l'objet en mouvement qui peuvent alors être extraits de l'image courante.

La méthode proposée rencontre des difficultés lorsqu'un objet en mouvement possède à la fois des pixels plus clairs et des pixels plus sombres que le fond. Dans ce cas plusieurs régions apparaissent dans chacune des images des différence cumulées.

La méthode est également mise en défaut face au phénomène précité de "running occlusion". Dans ce cas, deux régions de la DCP ou de la DCN fusionnent et créent des problèmes d'interprétation.

Toutes ces méthodes visant à se rapprocher du cas idéal d'une image de référence sans objet mobiles et d'éclairement identique à celui de l'image courante, compliquent parfois la procédure d'extraction des objets en mouvement à un point tel qu'elles remettent en question l'opportunité d'une approche basée sur la différence d'images.

Selon la qualité des informations que l'on désire extraire sur les objets en mouvement, il est parfois préférable de ne pas avoir recours à une image de référence, comme dans les méthodes exposées au paragraphe suivant.

# III.1.2- Sans image de référence.

Dans certaines applications on peut avoir à faire face à des variations d'éclairement importantes entre l'acquisition de l'image de référence et celle de l'image courante. On peut aussi ne pas avoir accès à une image de référence exempt d'éléments mobiles (scène constamment occupée par des objets en mouvement) ou encore se trouver en présence des deux situations simultanément. Dans de tels cas, les techniques décrites précédemment,utilisant une image de référence, échouent. C'est pourquoi certains chercheurs ont développé des procédures permettant l'emploi de la technique de la différence d'images sans nécessiter d'image de référence. Ces techniques ne font alors appel qu'à des images successives.

#### **III.1.2.1-** Deux images successives.

Les deux images utilisées pour effectuer la différence sont ici deux images successives de la séquence. L'image de laquelle on désire extraire les paramètres de mouvement est appelée l'image courante, notée C, celle qui la précède dans la séquence, l'image précédente, notée P.

La différence entre deux images qui comportent un objet en mouvement de niveau de gris uniforme, fait généralement apparaître quatre zones comme l'illustre les images 3 et 6 de la planche III.2 :

- zone 1 : située à l'avant de l'objet dans le sens du déplacement, cette zone est constituée des pixels qui correspondaient au fond dans l'image précédente et qui sont recouverts par l'objet dans l'image courante. L'amplitude des niveaux de gris dans cette zone dépend du contraste entre l'objet et le fond ;

- zone 2 : située à l'arrière de l'objet par rapport au sens du déplacement, cette zone est constituée des pixels qui correspondaient à l'objet dans l'image précédente et qui ont été découverts par l'objet dans l'image courante ;

- zone 3 : cette zone est constituée des points correspondant aux éléments stationnaires (fond ou objet) des images précédente et courante ;

· ····



1 : Image précédente

2: Image courante



3 : Image différence Déplacement non parallèle



4 : Image précédente



5: Image courante





Extraction des régions en mouvement

- zone 4 : cette zone, constituée de points qui correspondent à la partie commune de l'objet dans les images P et C, est celle qui crée le plus de problème. Ce "trou" n'apparait que si il existe un chevauchement de la position de l'objet entre les images précédente et suivante.

Les zones 1 et 2 ont des niveaux de gris assez élevés (identiques en valeur absolue si l'objet se déplace sur fond uniforme) qui dépendent du niveau de contraste entre le fond et l'objet.

Les zones 3 et 4 ont des niveaux de gris dépendant de la variation d'éclairement entre l'acquisition des deux images. Le niveau de gris est nul dans ces 2 zones si l'éclairement est resté constant. Sinon, ils dépendent de la variation d'éclairement et des réflectances respectives du fond et de l'objet.

Dans le cas de l'image 3 de la planche III.2 où les zones 1 et 2 sont connexes, il est aisé de reconstituer le masque de l'objet par regroupement des régions 1 et 4 [Yal80, Yal82].

En revanche si, comme l'illustrent les images 4 et 5 de la planche III.2, le mouvement de l'objet se fait parallèlement à certains de ses côtés, les zones 1 et 2 ne séparent plus la zone 3 de la zone 4. Ces deux dernières zones ne forment alors plus qu'une seule et même zone. Dans ce cas, seules des considérations d'ordre géométrique sur les bords des régions 1 et 2 permettent la reconstitution de l'image de l'objet.

Ce procédé s'applique bien aux objets polyhédriques, homogènes et exempts de cavité. La généralisation devient difficile pour des objets de formes quelconques. Remarquons cependant que pour des objets à bords non rectilignes, le problème ne se pose pas.

Une technique de classification des régions de l'image D a été appliquée par Jain [Jai81a] dans le cas d'une scène contenant plusieurs objets mobiles de niveaux de gris et de mouvements variés. Cette technique permet de différencier neuf régions selon leurs origines parmi les situations suivantes :

- recouvrement du fond par suite du mouvement d'un objet mobile ;
- découverte du fond par suite du mouvement d'un objet mobile ;

- occlusion ;

- translation d'un objet ;

- rapprochement d'un objet ;

- éloignement d'un objet ;

- position courante d'un objet ayant totalement quitté sa position précédente ;

- position précédente d'un objet dès qu'il a totalement quitté celle-ci ;

- indéterminé.

La classification n'est possible que si les différentes régions de l'image D correspondent à l'une de ces situations.

D'autre part, la classification se faisant à partir des images C, P et D binarisées, le résultat dépend du choix de trois seuils de binarisation dont la détermination pose fréquemment de réels problèmes. De ce fait, dans le cas de scènes complexes (objets et/ou fond texturés, niveaux de gris relatifs du fond et des objets inconnus) la classification devient impossible.

Remarquons cependant que, dans le cas où les niveaux de gris relatifs du fond et des objets sont connus, le signe de la différence d'images permet de déterminer aisément si la région correspond à l'objet dans sa position courante ou dans sa position précédente [Yac78]. Dans le cas contraire l'utilisation de trois images successives permet de résoudre le problème.

### III.1.2.2- Plus de deux images successives.

Proposée par Lenz [Len83] dans un but de réduction de données en vue de la transmission d'images, cette méthode considère les deux différences d'images suivantes : entre les images P et C d'une part, entre les images C et S, où S est l'image suivant l'image C, d'autre part. Les deux images différences sont d'abord seuillées, puis l'application d'un simple "ET " logique entre les images binaires ainsi obtenues permet alors d'isoler les régions en mouvement de l'image C (régions correspondant aux objets en mouvement dans leurs positions courantes).

Bien que résolvant efficacement et simplement le problème de la détermination de la position courante de l'objet, on notera cependant le retard qu'entraîne inéluctablement cette méthode, le résultat étant connu avec une acquisition de retard. Les effets de ce retard peuvent être minimisés si le temps s'écoulant entre deux acquisitions successives est suffisamment court.

# III.1.3- Limites de l'approche par analyse de régions.

L'extraction des régions en mouvement à partir de différences d'images est idéale lorsque l'on peut disposer d'une image de référence qui est, rappelons-le, une image exempt d'éléments mobiles acquise sous les mêmes conditions d'éclairage que l'image courante.

Lorsque une telle image de référence n'est pas accessible, on se trouve confronté à des problèmes divers qui ne connaissent pas de solutions universelles. En effet, l'image D issue de la différence entre deux images successives est alors composée de 4 types de régions distinctes (*cf.* III.1.2.1). Un processus de segmentation en vue d'isoler ces régions n'est envisageable que dans la mesure où l'on est capable de déterminer le nombre de classes présentes dans l'image. Le nombre de classes étant fonction à la fois des niveaux de gris des objets en mouvement et de leurs déplacements relatifs, sa détermination devient rapidement très difficile.

#### **III.2- EXTRACTION DES CONTOURS EN MOUVEMENT**

#### **III.2.1- Opérateurs existants.**

Lorsque la connaissance sur les niveaux de gris des objets et du fond est insuffisante pour permettre un processus fiable d'agrégation des régions appartenant à chaque objet, l'analyse des régions en mouvement devient difficile, voire impossible.

D'autre part, dans de nombreuses applications, seule la connaissance de la position des objets est recherchée et par conséquent, la seule connaissance des contours extérieurs des objets est suffisante.

C'est pourquoi, l'extraction des contours en mouvement est souvent préférée à l'extraction des régions elles-mêmes. Cette approche, assez récente, connait un engouement certain depuis la technique proposée en 1979 par Marr et Ullman [Mar79].

Pour procéder à une analyse des contours en mouvement, il est nécessaire de définir un opérateur qui, appliqué à l'image courante, permet d'extraire d'une scène dynamique l'ensemble des contours des objets en mouvement dans leurs positions courantes. Ce type d'opérateur fait appel au concept classique de détection de contour.

La détection de contours fait partie de la panoplie des outils de la vision artificielle, de la reconnaissance de forme et du traitement d'images depuis plusieurs décennies. On trouvera dans [Dav75, Dud75, Ros76b, Bas79, Mar79b] de nombreuses techniques de détection de contours accompagnés d'études comparatives. Sans s'étendre sur ce sujet qui fera l'objet d'un développement ultérieur en vue de notre application (cf. IV.2.1), rappelons en tout de même le principe : la détection de contours est généralement obtenue à l'aide d'un opérateur différentiel local agissant sur un faible voisinage du point considéré. La réponse de cet opérateur constitue une mesure du contraste ou de la "pente" du niveau de gris qui est en général importante au niveau des contours des objets.

Nous ne considérons, dans ce paragraphe, que des opérateurs permettant d'extraire des contours en mouvement à partir de différences d'images. Le choix des images opérandes constituant ces différences est basé sur les mêmes considérations que pour les techniques d'extraction de régions en mouvement. Par conséquent, nous ne traiterons pas le cas de méthodes où de tels opérateurs sont appliqués à une image de référence. En effet, dans ce cas les contours en mouvement s'obtiennent directement par application d'un opérateur classique d'extraction de contour à l'image différence. Seul le choix de l'opérateur différentiel conditionne alors l'efficacité de la méthode.

Les opérateurs d'extraction des contours en mouvement existants sont répertoriés ci-dessous, sous le nom de leurs inventeurs et sont présentés dans l'ordre de leur apparition dans la littérature. Nous noterons CM(I) l'image des contours en mouvement de l'image I, obtenue par l'application de ces opérateurs.

#### MARR et ULLMAN [Mar79]

Inspirés par les découvertes sur le système visuel des mammifères, Marr et Hildreth ont développé une technique de détection des contours très intéressante [Mar79b]. Marr et Ullman ont étendu cette technique afin de donner la direction du mouvement des éléments de contours détectés. Dans un premier temps, pour obtenir la direction approximative du mouvement ils utilisent la différence d'image. Dans un second temps, les points de passage à zéro du détecteur de contour sont collectés pour former des éléments caractérisés par leur orientation, leur longueur et leur amplitude. Dans le troisième temps, les dérivés temporelles de chaque point de passage à zéro sont calculées afin de préciser le sens du déplacement. Le problème de la combinaison des résultats obtenus avec les différents masques est approximativement résolu en ne conservant que les points détectés avec au moins deux des masques utilisés.

La technique de Marr et Ullman a pour principale limite de ne pouvoir trouver la direction du mouvement qu'à  $\pm 90^{\circ}$ . Ce problème, appelé le problème d'ouverture par les auteurs, est dû au fait qu'un élément de contour local ne peut donner qu'une information sur la direction du mouvement et non sur son sens.

# HAYNES et JAIN [Hay83]

L'opérateur d'extraction des contours en mouvement introduit par Haynes et Jain en 1983 est basé sur la combinaison de contours statiques et de contours dynamiques. Afin d'isoler les contours en mouvement de l'image courante, les auteurs proposent d'effectuer le produit entre l'image différence D(P,C) = |C-P| (contour temporel) et le l'image gradient G(C) (contour spatial) [Hay83] :

$$CM(C) = D(P, C) \times G(C)$$
 (1)

On peut considérer la différence D(P, C) comme un détecteur de mouvement et l'opérateur G(C) comme un détecteur de contour. La planche III.3 illustre les différentes étapes du calcul de l'image CM(C).

D'après les auteurs, le choix des détecteurs de mouvement D et de contour G n'est pas critique pour l'efficacité de la méthode. La seule contrainte concernant le choix de ces détecteurs est que les valeurs obtenues, qui seront multipliées entre elles, doivent correspondre aux mêmes points ou à leur voisinage immédiat. La méthode est peu sensible au bruit et les résultats sont peu affectés par le choix du détecteur de contour.

Dans l'image produit CM(C), la réponse de l'opérateur sera importante lorsque au moins un des deux facteurs (contour spatial ou contour temporel) sera important. Par conséquent, outre les objets présentant un fort contraste avec le fond et se déplaçant rapidement, l'opérateur met en évidence les objets présentant :

- un fort contraste avec le fond et se déplaçant lentement ;

- un faible contraste avec le fond et se déplaçant rapidement.





#### Extraction des contours en mouvement

Un seuillage approprié de l'image CM(C) permet de ne conserver que les points de contour des objets en mouvement. L'ajustement du seuil, qui pour des images de laboratoire peut apparaître peu critique (cf. planche III.4), reste toutefois délicat pour des images de scènes réelles peu contrastées [Ste85].

La technique rencontre des difficultés lorsque le déplacement d'un objet se fait parallèlement à l'un de ses bords rectilignes. Dans ce cas, malgré l'existence réel d'un bord dans l'image C, l'image D(P, C) prend sur celui-ci une valeur voisine de zéro (*cf.* III.1.2.1) et, par conséquent, les valeurs obtenues après multiplication par l'image G(C) y restent généralement faibles et voisines de celles dûes au bruit de mesure.

En faisant intervenir deux images temporellement proches, cet opérateur a le mérite de s'affranchir des conditions d'éclairage sous réserve que la vitesse de déplacement des objets est suffisante comparativement à la vitesse de variation de l'éclairement. Cependant il ne résoud pas totalement le problème du choix du seuil qui reste délicat.

# STELMASZYK et Coll. [Ste85].

Inspiré par l'opérateur précédent, Stelmaszyk propose d'utiliser comme détecteur de mouvement non plus l'image D(P,C) mais l'image G(D(P,C)), c'est à dire le gradient de l'image différence. Cette image fait par conséquent apparaître tous les points de contour en mouvement. Le produit par l'image G(C) est alors effectué. L'image CM(C) obtenue fait donc apparaître, comme pour l'opérateur précédent, les points de contour des objets en mouvement dans leur position courante.

L'opérateur s'exprime donc de la manière suivante :

$$CM(C) = G(D(P,C)) \times G(C)$$
 (2)

L'opérateur de gradient peut être quelconque à condition toutefois qu'il reste le même pour les deux images D et C. L'auteur a testé les opérateurs de Roberts, Prewitt et Sobel. La planche III.5 illustre les différentes étapes du calcul de l'image CM(C).







Planche III.5 : Détection des contours en mouvement selon l'algorithme de Stelmaszyk et Coll..

La planche III.6 illustre l'influence du seuil. L'auteur affirme que le seul calcul de la valeur maximale de niveau de gris Max{CM(C)}, mesuré sur l'ensemble des points de l'image CM(C), permet de déterminer un seuil correct en l'ajustant à la valeur Max{CM(C)}/2. Notons toutefois que ce seuil proposé par l'auteur est relativement élevé et tend de ce fait à supprimer un certain nombre de points de contour pertinents, ne laissant subsister, pour les objets en mouvement présentant un faible contraste avec le fond, qu'un nombre insuffisant de points de contour.

D'autre part, le produit s'effectuant entre deux images différentielles (images de contours), les valeurs combinées correspondent forcément aux mêmes points de l'image courante. Par conséquent, tous les points de contour en mouvement répondront fortement alors que pour l'opérateur de Haynes et Jain environ la moitié des points de contour en mouvement sont perdus. Ce dernier point constitue certainement le plus gros apport de l'opérateur de Stelmaszyk par rapport à celui de Haynes et Jain.

# III.2.2- Critique des opérateurs existants

Si le dernier opérateur décrit ci-dessus apporte de réelles améliorations par rapport au précédent, il pose un délicat problème lorsque le fond sur lequel se déplacent les objets n'est pas uniforme. Considérons pour expliquer ce problème, l'opérateur de Stelmaszyk appliqué aux images 1 et 2 de la planche III.7.

L'image G(C) contient une partie de points de contours du fond situés sous position l'objet dans l'image précédente. L'image G(D) contient également ces points de contours puisque l'image D contient les points du fond à cette même position, découverts par le mobile durant le mouvement. Par conséquent, ces points de contours du fond, communs aux deux images, seront conservés lors de la multiplication des images G(D) et G(C).

En plus des points de contours de l'objet dans sa position courante, apparaîtront donc dans l'image CM(C), des points de contours de l'objet dans sa position précédente qui cachaient les points de contour du fond. L'image 6 (image en niveaux de gris) de la planche III.7 met en évidence le phénomène.

Un raisonnement analogue peut être suivi pour l'opérateur de Jain.



Planche III.6 : Détection des contours en mouvement selon l'algorithme de Stelmaszyk et Coll. : influence du seuil.



Planche III.7 : Détection des contours en mouvement sur fond non uniforme selon l'algorithme de Stelmaszyk et Coll..

Ces fausses détections constituent la grande faiblesse des opérateurs de détection des contours en mouvement présentés ci-dessus. La littérature ne propose aucune solution à ce problème que l'on rencontre par exemple lorsque des véhicules se déplacent sur un fond non uniforme, tel que les voies d'un tramway ou un passage pour piétons.

L'opérateur proposé ci-dessous apporte une solution à ce problème, tout en conservant la robustesse de l'opérateur de Stelmaszyk.

#### III.2.3- Opérateur proposé.

# III.2.3.1- Description de l'algorithme.

Une solution efficace au problème des fausses détections dues à la non uniformité du fond passe par l'utilisation d'une troisième image et en utilisant l'opérateur:

$$CM(C) = G(D(P,C)) \times G(D(C,S))$$
(3)

La suppression des fausses détections est alors assurée à tout coup. En effet, il ne peut y avoir de zone de l'objet commune aux images P et S, qui soit extérieure au contour de l'objet dans sa position courante (excepté le cas peu probable d'une inversion du sens de la vitesse de l'objet après l'acquisition de l'image C).

Pour une écriture plus concise nous noterons : D l'image différence obtenue à partir des images courante C et précédente P (D=|C-P|), et D' l'image différence obtenue à partir des images courante C et suivante S (D'=|C-S|). L'image des contours en mouvements de l'image courante CM(C) sera notée CM.

La planche III.8 permet de mieux comprendre le fonctionnement de l'algorithme à partir d'une scène de laboratoire : l'image 1 constitue l'image suivante S, les images 2 et 3 de la planche III.7 constituant respectivement les images précédente et courante. L'image 2 représente G(D) et l'image 3 représente G(D'). L'image  $CM = G(D) \times G(D')$ (image en niveaux de gris) est représentée par l'image 4. On peut déjà remarquer qu'il ne subsiste aucun point visible du fond à l'extérieur du contour de l'objet mobile dans sa






Planche III.9 : Comparaison de l'algorithme proposé avec l'algorithme de STELMASZYK et Coll. sur images de laboratoire.

position courante dans l'image C. Par contre, la procédure fait apparaître des points à l'intérieur de ce contour mais ceux-ci ne nuisent pas à la détermination du mouvement.

# III.2.3.2- Comparaisons et résultats.

La planche III.9 permet de confirmer sur les images de laboratoire précédentes l'intérêt de l'opérateur proposé par rapport à celui de Stelmaszyk. Avec l'algorithme de Stelmaszyk, il est impossible d'isoler par un seuillage les véritables points de contours en mouvement (images 1, 2 et 3). Avec l'algorithme proposé, il parait possible de supprimer la quasi totalité des "fausses" détections sans pour autant éliminer une quantité importante de véritables points de contours en mouvement (images 4 et 5).

Les planches III.10.a et III.10.b permettent quand à elles de constater l'efficacité de l'opérateur proposé sur des images de scènes réelles à un carrefour urbain : les images 4 et 5 de la planche III.10.b montrent, qu'avec l'algorithme de Stelmaszyk et *coll.*, il est impossible de supprimer totalement les points de contours dûs à la présence des rails sans supprimer de nombreux points de contours du minibus. Les images 6 et 7 montrent en revanche qu'avec l'algorithme proposé, il est possible de déterminer un seuil de manière à éliminer les détections dûes aux rails sans pour autant éliminer des points de contours du minibus.

Par conséquent, on peut affirmer qu'un fond non uniforme résultant de la présence de rails, de pavés ou de signalisations au sol ne constitue plus un obstacle à l'extraction des objets en mouvement dans leurs positions courantes.

### **III.2.3.3-** Limites de l'opérateur

Si l'opérateur ainsi défini permet d'éluder le problème de la non-uniformité du fond, il a un inconvénient majeur au niveau de la contrainte temps réel. En effet, il est nécessaire d'avoir acquis l'image suivante pour analyser l'image courante. Cet effet peut être négligeable si le déplacement de l'objet entre l'image C et S est faible c'est à dire si la fréquence d'acquisition est élevée.



Planche III.10.a : Comparaison de l'algorithme proposé avec l'algorithme de STELMASZYK et Coll. sur images de trafic urbain.



Planche III.10.b : Comparaison de l'algorithme proposé avec l'algorithme de STELMASZYK et Coll. sur images de trafic urbain.

# **III.3- CONCLUSION.**

Tous les opérateurs de contours en mouvement décrits ci-dessus font intervenir un détecteur de mouvement basé sur la différence d'images et, de ce fait, connaissent la même limitation que les opérateurs de détection de régions en mouvement. En effet, les points de contours en mouvement de vitesse parallèle au contour ne seront pas détectés car la différence y est nulle et c'est l'unique raison pour laquelle cette approche n'est pas souhaitée dans certaines applications.

Pourtant, comme l'attestent les résultats obtenus, seuls les contours d'objets mobiles très rectilignes présentent de grosses discontinuités. En ce qui nous concerne, ce n'est le cas ni des piétons, ni des automobiles.

L'avantage principal de ces opérateurs est la possibilité de s'affranchir des variations d'éclairement de la scène permettant ainsi le travail en extérieur.

L'opérateur proposé résoud quand à lui efficacement le problème de la non uniformité du fond.

Nous verrons dans le chapitre suivant, dans quelles mesures l'opérateur proposé est apte à résoudre le problème de la détection d'objets en mouvements (piétons et véhicules les plus divers) au niveau d'un carrefour urbain complexe.

# - CHAPITRE IV -

# **MISE EN OEUVRE ORIENTEE TEMPS REEL**

Les procédures de traitement des séquences d'images exposées dans le chapitre précédent ont été développées dans l'optique d'une mise en oeuvre orientée temps réel afin d'être utilisables dans le contexte général de la surveillance du trafic routier.

Dans le cadre d'une automatisation, le temps réel est en fait le temps maximum que l'on peut octroyer au système de commande pour qu'il génère l'action nécessaire en fonction de l'environnement.

En conduite automatique, ce temps doit être comparable au temps de réaction d'un conducteur, c'est à dire de l'ordre de la seconde. Cela revient à dire que la procédure globale de conduite automatique doit permettre de traiter et d'interpréter chaque image de la séquence, puis de générer la commande adaptée en environ une seconde.

Nous tenterons donc de minimiser le temps de traitement afin de laisser le maximum de temps disponible pour les processus d'interprétation des images qui doivent permettre de délivrer les ordres de conduite du véhicule guidé en fonction de l'état de la scène.

# **IV.1- DESCRIPTION MATERIELLE**

Le choix des algorithmes est fortement conditionné par les éléments de la chaîne de traitement d'images utilisée. La figure IV.1 montre les différents éléments qui composent le système de traitement d'images utilisé et qui sont décrits dans les paragraphes suivants.



Figure IV.1 : Schéma bloc de la chaîne de traitement d'images.

# **IV.1.1- Acquisition des images**

Les images sont acquises par une caméra matricielle de type CCD qui délivre un signal vidéo standard à la cadence de 20 ms par trame, soit 40 ms par image, une image vidéo étant composée d'une trame paire et d'une trame impaire. Les images sont ensuite numérisées à cette même cadence vidéo sur 6 bits et selon une résolution de 256 x 256 pixels correspondant à une trame vidéo, soit une demi-image. Seules les images numérisées correspondant aux trames paires sont analysées. Le délai séparant deux images consécutives sera par conséquent de 40 ms.

Tous les temps annoncés par la suite s'appliquent à cette résolution de  $256 ext{ x}$  256 pixels.

# IV.1.2- Description du système de traitement d'images.

Le système utilisé est un "open system" Intel 380 bâtit autour de la carte iSBC 286/12 supportant les cartes spécialisées de la série IP-512 d'Imaging Technology. Le schéma bloc de cette partie matérielle est représenté à la figure IV.2.

# IV.1.2.1- Carte iSBC 286/12.

Cette carte et les logiciels fournis permettent la programmation dans les langages Pascal, Assembleur et PL/M. La vitesse de calcul en Pascal est de l'ordre de 300 000 additions d'entiers par seconde. Ce matériel n'est en rien spécifique au traitement d'images. C'est pourquoi compte tenu de ses performances, on ne peut y implanter des algorithmes complexes dans une optique temps réel. A titre d'exemple, une simple addition de deux images 256 x 256 déjà stockées sur la carte nécessiterait plus de  $2/10^{e}$  s.

# IV.1.2.2- Cartes IP-512.

Ces cartes composent la partie du système spécifique au traitement d'images. Elles sont aux nombre de six, réparties de la manière suivante selon leur fonction :

- 1 carte unité arithmétique et logique ALU-512 ;
- 4 cartes mémoires FB-512;
- 1 carte de conversion analogique/numérique AP-512.

L'architecture de type pipeline de ces cartes permet le transfert à travers l'unité arithmétique et logique de quatre images de résolution 256 x 256 codées sur 8 bits en une trame vidéo, soit 20 ms. Cette rapidité de transfert est le premier point fort de la machine.

La carte unité arithmétique et logique ALU-512 permet d'effectuer à cette même cadence une opération élémentaire sur ces images. Nous entendons par opérations



Figure IV.2 : Schéma bloc de la partie matériel.

#### Description matérielle

élémentaires des opérations sur une image telles que transfert, binarisation ou produit par une constante, et sur deux images telles que sommé, différence ou produit. Ces opérations se font évidemment pixel à pixel. Le schéma bloc de la carte ALU-512 est représenté figure IV.3.

Les quatre cartes FB-512 permettent le stockage de 16 images 256 x 256 codées sur 8 bits. Ces cartes sont configurées en deux groupes de deux cartes. Un registre pour chacun des groupes permet de définir par logiciel le point d'origine des images. Cette possibilité qui constitue le deuxième point fort de la machine, offre la possibilité de décaler les images les unes par rapport aux autres avant de réaliser des opérations entre elles.

La carte AP-512 permet, comme son nom l'indique, la conversion analogique/numérique du signal vidéo. Avant son stockage dans la mémoire d'image d'une carte FB-512, l'image passe à travers une "LUT " configurable par logiciel. Ceci



#### Figure IV.3 : Schéma bloc de la carte ALU-512.

\_IV-4

permet à l'utilisateur de transformer l'image numérisée issue du convertisseur en faisant correspondre à chaque niveau de gris d'entrée de la LUT un niveau de gris de sortie. La correspondance entre le niveau de gris d'entrée et celui de sortie peut suivre une loi mathématique ou être quelconque, au choix de l'utilisateur.

#### IV.1.2.3- Transfert des données entre calculateur et mémoires d'images.

Le transfert des données entre mémoires d'images et calculateur se fait via le registre 16 bits des cartes FB-512 et le bus CPU à la cadence de 0,4 s par image 256 x 256 (soit 164 000 mots/s) et cela quelque soit le langage de programmation utilisé. A cause de cette lenteur, nous limiterons au maximum ces transferts entre mémoires d'images et calculateur.

Toutes ces cartes sont commandées au moyen de la carte processeur central iSBC 286/12 via le bus CPU.

#### **IV.2- OPERATEUR DE DETECTION DE CONTOURS.**

L'efficacité de l'opérateur de détection des contours en mouvement présenté au chapitre précédent est évidemment liée à la qualité des contours mis en évidence dans les images G(D) et G(D') où, rappelons le, D représente la valeur absolue de la différence entre l'image courante C et l'image précédente P (D=|C-P|) et D' représente la valeur absolue de la différence entre l'image courante C et l'image courante C et l'image courante C et l'image courante C et l'image suivante S (D'=|C-S|). Par conséquent, le choix de l'opérateur de détection de contour qui sera appliqué aux images D et D' est fondamental.

# **IV.2.1-** Détection des contours

La détection des contours constitue l'une des procédures de base de la vision artificielle. Une littérature aussi abondante [Dav75, Dud75, Ros76b, Bas79, Mar79b] que récente [Neu86] l'atteste.

Un contour peut être défini comme une zone de transition entre deux régions relativement homogènes d'une image. Dans le cas où l'on cherche à différencier des objets du fond sur lequel ils apparaissent, un contour peut être considéré comme une zone de l'image à fortes variations locales de niveaux du gris.

D'une manière générale, on peut décomposer la détection de contours en deux étapes :

- d'un point de vue local, il s'agit de détecter des éléments de contours, c'est à dire les points autour desquels la fonction intensité f(x,y) présente de fortes variations et qui sont donc susceptibles d'appartenir à des contours. Cette détection locale de points de contour fait appel à des opérateurs différentiels qui ont tendance à répondre fortement au niveau des contours réels des objets mais également dans d'autres zones de l'image car ils sont sensibles au bruit. Notons que les contours ainsi détectés ne correspondent pas nécessairement aux bords physiques d'objets, mais peuvent représenter des variations de niveaux de gris à l'intérieur de ceux-ci ;

- d'un point de vue plus global, il faut ensuite relier entre eux ces points pour obtenir des contours continus et supprimer les détections locales isolées non cohérentes, assimilées à du bruit. Cette étape de reconstitution des contours peut nécessiter l'utilisation d'informations connues *a priori*. Mais il existe également des procédures de suivi de contour qui permettent de reconstituer ces éléments fondamentaux de l'image sans cette information.

Nous n'aborderons ici que l'étape de détection locale, la reconstitution des contours des objets stationnaires ou en mouvement dans leurs positions précédentes et suivantes ne présentant aucun intérêt dans le cadre de cette étude. Par contre, la nécessaire reconstitution des contours des objets en mouvements dans leurs positions courante sera examinée après application de l'opérateur de détection des contours des objets en mouvement.

#### **IV.2.2-** Opérateurs différentiels de type gradient

Le contour est ici détecté grâce à l'information apportée par le gradient (ou plutôt une de ses approximations par différences finies) de la fonction représentant le niveau de gris. On définit ainsi en chaque point de l'image, un vecteur contour perpendiculaire au gradient en ce point.

De nombreux masques de gradient ont été proposés dans la littérature depuis une quinzaine d'années. On en trouvera les définitions en [Pra78, Rob77]. De manière théorique, le vecteur gradient G est défini par :

- son amplitude ou hauteur :  $|\mathbf{G}| = \sqrt{X^2 + Y^2}$  (1)
- son orientation :  $\theta = \arctan Y/X$  (2)

où X et Y sont les deux composantes orthogonales du vecteur G.

En pratique, on n'utilise que rarement l'orientation et on ne considère qu'une approximation du module telle que :

$$|G| = |X| + |Y|$$
(3)

ou bien

$$|\mathbf{G}| = Max\{|\mathbf{X}|, |\mathbf{Y}|\}$$
 (4)

Les différents opérateurs sont représentés sous forme de masque que l'on convolue avec chaque point de l'image.

Par exemple, le masque  $\begin{bmatrix} 0 & -1 \\ 0 & 1 \end{bmatrix}$  où la valeur en caractère gras correspond à la position du point transformé, représente l'opérateur :

$$G(i,j) = |I(i,j) - I(i-1,j)|$$
(5)

où *i* est l'indice de ligne et *j* l'indice de colonne.

Les opérateurs ci-dessous, représentés par leur masque dans une seule direction, sont certainement parmi les plus utilisés :

- Roberts [Rob65] : 
$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

- Kirsch [Kir71] :

- Sobel (C= 2) [Abd78] et Prewitt (C= 1) [Pre70] :

$$\left[\begin{array}{ccc}
1 & 0 & -1 \\
C & 0 & -C \\
1 & 0 & -1
\end{array}\right]$$

1 1 1 1 -2 1 -1 -1 -1

- Robinson [Rob77] :

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \quad \text{ou} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad \text{ou encore}$$

Plus récemment, Neuvo a proposé la masque suivant [Neu86] :

-2	1	1]	
-2	1	1	
-2	1	1	
		ر ا	

Les masques correspondant aux différentes directions possibles du contour se déduisent par rotation des masques présentés ci-dessus. Ainsi, le masque correspondant à la deuxième direction possible du masque de Roberts est :

ſ	n	_1	T
	1	-1	
L		v	J

Les deux masques correspondant aux deux composantes orthogonales X et Y définissent ainsi le module et l'orientation du vecteur gradient G selon les expressions (1) et (2) respectivement.

Il est possible de déduire de chacun des masques présentés ci-dessus des masques de dimensions supérieures. Par exemple, l'opérateur de Neuvo correspondant à une fenêtre 5 x 5 conduit au masque [Neu86] :

-3 -3 2 2 2 -3 -3 2 2 2
-3-3222
-3-3222
-3-3222

Deux types d'utilisation de ces masques de gradient sont à distinguer : la première consiste à évaluer le module par une combinaison de deux masques orthogonaux obtenus par une rotation de  $\pi/2$ . L'expression (3) correspond à une telle utilisation. L'orientation peut alors être déduite par l'expression (2).

La deuxième utilisation consiste à évaluer le module par recherche de la valeur maximale obtenue par applications successives des masques correspondant à différentes orientations possibles. L'expression (4) correspond à cette autre utilisation.L'orientation du contour est, dans ce cas, donnée par l'orientation du masque ayant fourni la valeur attribuée au module, c'est à dire la valeur maximale.

Dans les deux méthodes, on décide de la présence d'un point de contour si le module évalué dépasse un certain seuil.

Il faut toutefois noter que les opérateurs différentiels répondent à toute discontinuité d'intensité, quelle qu'en soit l'origine. Ainsi, un changement local de réflectance ou d'illumination d'une surface sera détecté de la même manière qu'un contour d'objet. Il existe des procédures plus fines qui prennent en compte, de manière séquentielle, les valeurs de |G| et  $\theta$  afin de sélectionner les véritables contours de l'image. Nous n'aborderons pas ici ces méthodes de suivi de contour.

D'autre part, les réponses des opérateurs différentiels varient selon la dimension du masque utilisé. Plus la dimension n du masque est faible, plus les hautes fréquences sont favorisées. Les plus hautes fréquences spatiales seront par conséquent détectées par un masque de dimension minimale (n=2).

Si l'on ne désire pas favoriser certaines fréquences, une solution consiste à utiliser plusieurs masques de dimensions différentes [Mar79, Ros71] dont on recombine ensuite les différentes réponses.

### IV.2.3- Opérateurs différentiels de type laplacien

Inspirés par des découvertes de neurophysiologie sur les systèmes visuels chez les mammifères, Marr et Hildreth [Mar79] ont développé une technique de détection de contours très intéressante. Cette technique utilise les passages à zéro du Laplacien de la fonction "niveau de gris" de l'image. Le masque le plus standard est le suivant :

1	1	1	
1	-8	1	
1	1	1	

Mais on peut également faire appel à des filtres gaussiens de différentes tailles afin de détecter les contours de différentes fréquences spatiales.

Robinson quant à lui utilise le masque [Rob77] :

$$\left(\begin{array}{c}
1 -2 & 1 \\
-2 & 4 -2 \\
1 -2 & 1
\end{array}\right)$$

comme approximation discrète d'un opérateur bilaplacien du type  $d^4/dx^2dy^2$ .

Dans le même ordre d'idée, Modestino et Fries [Mod77] ont proposé un filtre pour une image modélisée par la somme d'un champ aléatoire i(x) et d'un bruit blanc b(x). Son action approche au mieux, au sens des moindres carrés, l'action de l'opérateur laplacien sur i(x).

# IV.2.4- Choix du détecteur de contours

Le choix d'une méthode de détection de contours adaptée à un problème particulier dépend non seulement du type d'images étudiées et des contours recherchés, mais aussi du matériel informatique dont on dispose (architecture pipeline, microprocesseurs parallèles ou en tranches, ...) et du mode de traitement requis (temps réel, implantation en réalisation câblée, etc...) [Bas79].

Etant donné l'architecture pipeline du matériel utilisé et l'aspect temps réel requis, il apparait que les opérateurs de type gradient sont les plus adaptés à notre problème.

Parmi les nombreux masques disponibles, un choix reste à faire. Celui-ci s'effectue en fonction des points suivants :

- les caractéristiques des contours recherchés (orientation, épaisseur, longueur, hauteur);

- le temps de calcul disponible, fonction du délai autorisé par le cahier des charges entre l'acquisition de l'image courante et la modélisation de la scène et, par conséquent, du temps de calcul incompressible nécessaire aux autres traitements (filtrage, reconstitution des contours, modélisation, génération de la commande);

- la robustesse par rapport au bruit qui est importante lorsque l'on considère des images de scènes extérieures.

Le masque proposé par Neuvo [Neu86] présente la propriété remarquable de filtrer le bruit sans diminuer la netteté des contours. Son implantation sur le matériel utilisé ne conduit malheureusement pas à un temps de calcul compatible avec les contraintes temps réel.

Pour nous rapprocher du but recherché, nous avons été conduits à définir le gradient défini par les deux masques de la figure IV.4 ci-dessous.

$ \left[\begin{array}{ccccc} 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\$	$ \left[\begin{array}{ccccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] $

Figure IV.4 : Masques du gradient adopté.

Le module choisi est, selon la notation adoptée pour (5) :

$$G(i,j) = |I(i-2,j) - I(i+2,j)| + |I(i,j-2) - I(i,j+2)|$$

Avec cet opérateur, qui répond dans une gamme de fréquences plus basses que les traditionnels masques 3 x 3, les contours obtenus manquent de détails. Cependant, cela ne constitue en aucun cas un inconvénient car le contour extérieur des objets mobiles ne nous intéresse que pour leur localisation, et non pas pour leur reconnaissance. Bien au contraire, une réponse importante sur une zone étendue au niveau des contours tend à renforcer la continuité des contours détectés.

D'autre part, cet opérateur différentiel peut être aisément implanté sur la machine dont nous disposons, sans encombrement mémoire ni temps de calcul supplémentaires malgré l'étendue de la fenêtre  $(5 \times 5)$ . Le temps de calcul est identique à celui auquel on aboutirait avec un masque  $2 \times 2$ . La planche IV.1 permet une comparaison des réponses des deux opérateurs classiques de Roberts et Sobel et celle de l'opérateur proposé appliqués à une image de scène réelle. Pour une meilleure visualisation des résultats, les images de gradient obtenues ont été binarisées selon deux seuils : les images 2, 3 et 4 correspondent à un seuil S=3, les images 5,6 et 7 correspondent à un seuil S=10 pour une plage de variation possible de 0 à 64. On notera l'obtention d'une meilleure continuité du contour extérieur des véhicules avec l'opérateur proposé.

Le seul inconvénient de cet opérateur est le débordement extérieur des contours détectés par rapport à la position réelle du bord des objets. Par conséquent, deux objets séparés d'une distance inférieure à 5 points images verront leurs contours extérieurs fusionner. La procédure d'analyse et d'interprétation de la scène observée présentée par ailleurs [Bou88], prend heureusement en compte ce phénomène éventuel de fusion.

Ayant défini l'opérateur différentiel à utiliser pour détecter les contours dans les images différences D et D' nous abordons maintenant l'implantation de l'opérateur de détection des contours en mouvement.





5: Roberts



6 : Gradient proposé Seuil = 10



- 7 : Prewitt
- Planche IV.1 : Comparaison de l'opérateur de gradient proposé avec l'opérateur de Roberts et celui de Prewitt.



# **IV.3- OPERATEUR DE DETECTION DES CONTOURS EN MOUVEMENT**

#### **IV.3.1-** Extraction des contours en mouvement

L'étape de modélisation de la scène courante aura pour objectif de définir les paramètres de dimensions au sol et de position de chacun des objets en mouvement de la scène. Ces paramètres peuvent être aisément déterminés par la connaissance de l'ensemble des seuls points de contours extérieurs de l'objet.

Revenons à l'image  $CM = G(D) \times G(D')$ . Dans cette image, les points de contours en mouvement sont caractérisés par de fortes valeurs de niveaux de gris. Il s'agit maintenant d'exploiter cette image CM afin d'en extraire les contours en mouvement.

Dans le cas idéal où le contour extérieur d'un objet est une courbe fermée et continue, celle-ci peut être extraite de l'image CM par une procédure de chaînage portant sur les points de forts niveaux de gris.

L'opération comporte cependant certains risques car, comme nous l'avons vu aux paragraphes III.2 et III.3, l'opérateur de détection de mouvement utilisé ne répond pas uniquement aux points de contours en mouvement. De plus, il arrive qu'il ne réponde que faiblement au niveau de certains contours en mouvement. C'est la raison pour laquelle un simple seuillage ne permet pas d'extraire correctement les contours recherchés. Cette opération conduit généralement à des résultats assez peu significatifs et difficiles à interpréter et ce, quelque soit le choix du seuil. En effet, avec un seuil trop bas on trouve dans l'image binarisée de nombreux points ne correspondant pas à un objet en mouvement de l'image courante. Ces points indésirables, qui ont pour origine le bruit de mesure, sont plus ou moins isolés et répartis sur l'ensemble de l'image. Les images 1 et 2 de la planche IV.2 illustrent ce phénomène.

En revanche, avec un seuil trop haut, on fait disparaître de l'image non seulement tous ces points indésirables mais également des points significatifs, ce qui a pour effet de créer des discontinuités du contour extérieur de certains objets, particulièrement pour ceux présentant un faible contraste avec le fond. Les images 2 et 3 de la planche IV.2 illustrent également ce phénomène. Ces discontinuités ont pour conséquences fâcheuses de faire correspondre plusieurs amas de points disjoints à chaque objet en mouvement au lieu d'un seul, ce qui complique la tâche ultérieure de modélisation de la scène. Dans



1: Seuil = 5



**2**: Seuil = 15



3 : Seuil = 25 Avant Moyennage



Planche IV.2 : Comparaison du résultat avant et après lissage de l'image CM par le masque proposé.

de telles conditions, il devient difficile, à partir de ces ensembles disjoints de retrouver les paramètres recherchés, à savoir : la surface au sol et la position du centre géométrique de chacun des objets en mouvement.

Il est donc impossible par un simple seuillage de ne sélectionner que les points correspondant réellement au contour extérieur des objets en mouvement sans en éliminer une quantité non négligeable. Il est nécessaire d'envisager un processus de sélection plus élaboré afin de rendre plus fiable le résultat de la modélisation de la scène. Il existe des procédures de suivi de contour séquentielles qui, par optimisation de divers critères, permettent d'extraire d'une image, des contours présentant des discontinuités. L'aspect temps réel de ce projet en interdit malheureusement l'emploi.

Les méthodes de relaxation permettent également d'améliorer l'extraction des contours. Elles sont hélas inutilisables pour la même raison [Tou87].

Dans ces conditions, nous nous sommes portés vers des méthodes plus simples qui, bien que moins élégantes, permettent d'extraire les contours recherchés en temps réel.

# IV.3.2- Filtrage et mise en forme

Afin d'améliorer la connexité des contours, on procède à un lissage de l'image.

On peut penser à utiliser les masques classiques (a) et (b) de la figure IV.5. Ces masques peuvent être implantés avantageusement sur la machine utilisée car les pondérations proposées sont des puissances de 2. Le masque (a) nécessite 9 trames et le masque (c) 12 trames.

Cependant, si l'on cherche à tirer au maximum profit des possibilités de la machine, on s'aperçoit que, du point de vue temps de calcul, l'utilisation de fenêtres de dimensions égales à une puissance de 2 est très intéressante. Ainsi le masque (c) de taille  $4 \times 4$ , ne nécessite que 12 trames, c'est à dire autant que le masque (b) de taille  $3 \times 3$ . Son effet de lissage sera donc supérieur pour un temps de calcul identique.



Figure IV.5 : Filtres de lissage avec et sans pondération.

Notons que l'introduction d'une pondération améliorerait l'efficacité de ce masque de lissage. Cependant, étant donné le matériel utilisé, elle n'est pas souhaitable car elle pénaliserait fortement le temps de calcul.

Les résultats obtenus avec le masque (c) de la figure IV.5 sont assez satisfaisants. Les images 4 et 5 de la planche IV.2. montrent l'intérêt du lissage par ce masque : la continuité des contours extérieurs des objets en mouvement est considérablement améliorée et le bruit est inexistant.

La planche IV.3 illustre la robustesse du résultat obtenu par rapport au seuil : les images 3 et 5 montrent qu'une augmentation de près de 80% par rapport à un seuil minimal satisfaisant (S = 9) n'affecte pas considérablement le résultat.

## IV.3.3- Conclusion.

Le bruit présent dans l'image des contours en mouvement CM trouve son origine dans les images qui ont servi à la calculer, à savoir les images courante C, précédente P et suivante S. Il n'est cependant pas opportun de filtrer le bruit dès son origine car cette opération, non seulement serait à réaliser sur ces 3 images au lieu d'une seule, mais surtout elle dégraderait la netteté des contours. Pour les mêmes raisons, un filtrage au niveau du calcul des gradients des images différences D et D' a été rejeté et c'est pourquoi nous avons opté pour un filtrage de l'image CM.

La taille relativement importante du masque choisi augmente sensiblement la continuité des contours, ce qui, comme nous le verrons au paragraphe suivant, permet une amélioration notable du processus de modélisation, tant sur le plan fiabilité que sur le plan temps de calcul.



1: Seuil = 3



**2** : Seuil = 6



3: Seuil = 9



**4 :** Seuil = 12







<u>IV-17</u>

Dans la suite de ce mémoire, nous appellerons désormais CM l'image résultant du filtrage et de la binarisation de l'image appelée jusqu'à présent CM. Cette image est par conséquent une image binaire dans laquelle les points de valeur '1' (points blancs) représentent des points de contours d'objets en mouvement.

# **IV.4- MODELISATION DES OBJETS EN MOUVEMENT.**

Le problème que nous abordons maintenant n'est pas un problème classique de reconnaissance des formes en vision artificielle. Il ne s'agit pas de reconnaître des objets en les décrivant avec rigueur pour les identifier. La variété des éléments mobiles dans une scène de trafic interdit une telle approche en temps réel.

A ce stade de l'étude, il s'agit maintenant de décrire la scène dynamique observée afin de fournir un ensemble de données qui permettent de comprendre le mouvement des objets mobiles.

Dès le début de cette étude, nous nous sommes placés dans le cadre d'une vision 2-D. Nous allons donc passer en revue les techniques de description des formes planes en vue de les utiliser pour modéliser les objets en mouvement dans la scène observée.

## **IV.4.1- Descriptions géométriques**

Il existe un grand nombre de propriétés géométriques simples qui permettent de caractériser une forme plane. Les attributs géométriques les plus utilisés car les plus aisés à déterminer sont : la surface A, le périmètre P, le degré de compacité exprimé par la formule  $P^2/A$ .

On peut également modéliser la forme des objets par des polygones à partir desquels on détermine différents paramètres caractéristiques tels que :

- la mesure des angles entre les cotés ;

- la longueur des cotés ;

- la régularité définie comme le rapport de la variance de la longueur des cotés sur la variance de tous les angles ;

- la variabilité angulaire définie comme la moyenne des valeurs absolues des différences entre les angles adjacents du contour extérieur, les angles convexes prenant le signe positif, les angles concaves prenant le signes négatif [Att57].

De nombreux algorithmes ont été développés pour quantifier ces paramètres [Pra78, Dud73, Ros766b, Kul83, Ros80, Baj84].

On trouvera en [Zus70] une excellente revue de ces différentes descriptions basées sur les propriétés géométriques des formes.

Le polygone le plus simple est le rectangle exinscrit. L'enveloppe convexe constitue un modèle polygonal beaucoup plus élaboré sur lequel nous reviendrons ultérieurement.

Citons également : les moments invariants de Hu [Hu62], les descripteurs de Fourier [Wal80b, Zah72], les profils centroïdes de Freeman [Fre78], les matrices de forme de Goshtasby [Gos85] ou encore la signature polaire [Gos85]. Nous ne nous étendrons pas sur ces descriptions car leur propriété de conservation de l'information ne présente pas un intérêt déterminant dans le cadre de cette étude où la forme des objets mobiles évolue en fonction de leur position par rapport à la caméra. De plus,elles ne conduisent généralement pas à une réduction de données suffisante pour envisager une application temps réel.

La plupart des techniques de description évoquées ci-dessus n'utilisent que le contour extérieur et ne tiennent pas compte des caractéristiques internes de la forme. Si cette propriété peut être pénalisante pour certaines applications, elle est par contre tout à fait intéressante pour le problème envisagé. En effet, la modélisation a pour objectif d'extraire de chaque objet en mouvement dans la scène analysée un ensemble de paramètres suffisants pour générer la commande d'un tramway lors de sa traversée d'un carrefour, en évitant les collisions et en minimisant la durée du trajet. Cet ensemble, établi par Boutry [Bou88], se compose des éléments suivants :

- la position du centre de gravité de chaque objet mobile ;

- ses dimensions au sol.

qui peuvent être obtenus à partir des seuls points de contour extérieurs des objets mobiles.

Des informations supplémentaires permettant de reconnaître le type de mobile (piéton, cycle, automobile, etc...) apporteraient une plus grande souplesse du processus de commande par l'utilisation de connaissances *a priori* sur le comportement dynamique de chacun de ces types de mobiles. Ces informations supplémentaires ne seront pas recherchées pour l'instant.

A partir de l'image CM des points de contours en mouvement obtenue dans l'étape précédente, la modélisation choisie doit donc permettre d'établir ces éléments avec un maximum de précision.

Dans la mesure où les objets mobiles sont vus du dessus, il apparait que seuls les points de leur contour extérieur suffisent pour déterminer leurs dimensions au sol et pour en déduire la position de leur centre géométrique. La mise en correspondance des objets mobiles entre les différentes images d'une séquence est obtenue à partir de ces paramètres [Bou88]. Ceci permet de calculer ensuite leur vitesse à partir de la variation de position du centre géométrique entre les images successives et ainsi de générer la commande du tramway.

Afin d'obtenir ces paramètres tout en respectant les contraintes de la commande en temps réel, il est nécessaire de définir un modèle simple établi à partir des seuls points du contour extérieur de chaque objet mobile. Différentes approches basées sur l'encadrement du contour extérieur par un polygone convexe sont abordées dans le paragraphe suivant.

# IV.4.2- Modélisation du contour extérieur par un polygone

Le polygone de périmètre minimal exinscrit à une forme, appelé également enveloppe convexe, a fait l'objet de nombreuses études [Baj84, Skl72, Orl85, Gos83, Tou82]. D'autres polygones moins élaborés tels des rectangles ou des octogones peuvent également être utilisés mais ces formes géométriques simples sont surtout employées pour localiser une forme ou délimiter une zone d'intérêt dans une image [Stu80].

# IV.4.2.1- Modèle polygonal convexe de périmètre minimal

L'approximation polygonale convexe du contour extérieur des formes ("convex hull") a été introduite au début des années soixante-dix [Skl70, Skl72b, Mun68]. En [Skl72b], Sklansky propose un premier algorithme pour déterminer le polygone de périmètre minimal exinscrit ("CCP: circumscribed convex polygon") à un polygone de convexité quelconque. Cet algorithme échouant dans certaines situations, d'autres algorithmes plus complexes ont été développés depuis [Sha75]. La simplicité de l'algorithme de Sklanski a incité Toussaint et Avis [Tou82], puis Orlowski [Orl83] à reprendre ces travaux. Toussaint et Avis [Tou82] définissent ainsi une condition suffisante d'application et propose une modification de l'algorithme pour les formes satisfaisant à cette condition ("weakly externally visible polygons"). Orlowski, quant à elle, définit une condition nécessaire et suffisante d'application puis propose une modification de l'algorithme, le rendant ainsi applicable à tous les polygones [Orl85].

La quantité de calcul nécessaire à la détermination de ce modèle croit avec le nombre de sommets du polygone : il a été montré que quelque soit l'algorithme employé, le coût en calcul est fonction de  $n \log n$ , où n est le nombre de sommets du polygone [Sha78]. Lorsque les objets à modéliser s'éloignent de cette forme polygonale, la quantité de calcul devient rapidement imposante, rendant cette modélisation inexploitable en temps réel.

D'autre part, notre application ne semble pas devoir nécessiter d'une description aussi élaborée des contours extérieurs. En effet, la description par un polygone convexe ne permet pas une discrimination des objets mobiles : deux mobiles différents mais de même type (deux roues, automobile, piéton, etc...) peuvent avoir approximativement le même polygone exinscrit. De plus, les variations de l'image d'un objet mobile au cours du déroulement de la séquence analysée peuvent être très importantes de telle sorte qu'un modèle polygonal de périmètre minimal manquerait de stabilité dans l'optique d'un algorithme de poursuite multicible.

C'est pourquoi, aucune implantation d'une telle modélisation n'a été tentée.

#### **IV.4.2.2-** Modèle rectangulaire

Le rectangle est, après le triangle, la forme géométrique la plus élémentaire permettant de modéliser d'un objet considéré comme bidimensionnel.

Dans une optique robotique, Stuck [Stu80] a développé une procédure permettant d'encadrer tout amas de points connexes par un rectangle de cotés parallèles et perpendiculaires au balayage vidéo. Cette modélisation, très peu couteuse en calculs a été utilisée par Stelmaszyk [Pos86]. Elle est malheureusement très sensible aux variations d'orientation de l'objet. En effet, les dimensions du modèle s'éloignent d'autant plus de celles de l'objet que celui-ci se déplace en diagonale par rapport au balayage vidéo (cf. images 1 et 2 de la planche IV.4).

Le rectangle exinscrit dont les dimensions correspondent le mieux à celles de l'objet à modéliser est le rectangle exinscrit de périmètre minimal qui tient compte de l'orientation de l'objet. Quoique idéal pour notre application, la détermination de ce rectangle nécessite de nombreux calculs : calcul du tenseur d'inertie, de ses axes principaux d'inertie, etc... . Il fut utilisé par Ferrie [Fer82] pour la modélisation de cellules en complément à d'autres paramètres.

# IV.4.2.3- Modèle octogonal.

Un octogone exinscrit à un ensemble de points connexes peut être défini de diverses manières. L'octogone de périmètre minimal est évidemment l'octogone idéal, cependant comme pour le rectangle du même nom, sa détermination nécessite de trop nombreux calculs. C'est pourquoi nous proposons avec Boutry [Bou88] le modèle suivant, plus approximatif mais cependant beaucoup plus rapide à déterminer. Il correspond à un octogone composé de deux cotés parallèles, de deux cotés perpendiculaires au balayage vidéo et de quatre cotés de pente +1 ou -1. Par conséquent, les huit cotés forment deux à deux des angles de  $45^{\circ}$  et sont définis par les huit valeurs suivantes :

-Xmin, correspondant au point de contour le plus à gauche, définit le coté Ouest;

- Xmax, correspondant au point de contour le plus à droite, définit le coté Est ;

- Ymin, correspondant au point de contour le plus haut, définit le coté Nord ;

- Ymax, correspondant au point de contour le plus à bas, définit le coté Sud ;

- (X-Y)min, correspondant au point de contour présentant la plus petite différence entre ses deux coordonnées, définit le coté Sud-Ouest ;

- (X-Y)max, correspondant au point de contour présentant la plus grande différence entre ses deux coordonnées, définit le coté Nord-Est ;

 - (X+Y)min, correspondant au point de contour présentant la plus petite somme de ses deux coordonnées, définit le coté Nord-Ouest;

- (X+Y)max, correspondant au point de contour présentant la plus grande somme de ses deux coordonnées, définit le coté Sud-Est.



Figure IV.6 : Représentation du modèle octogonal.

La figure IV.6 ci-dessus représente un tel octogone et met en évidence les huit valeurs décrites ci-dessus.

Le modèle octogonal ainsi défini, bien qu'indépendant de l'orientation de l'objet à modéliser, est cependant beaucoup moins sensible aux rotations des objets que le modèle rectangulaire proposé par Stuck (*cf.* IV.4.2.2). Les images 1 et 2 de la planche IV.4 illustrent l'approximation réalisée par le modèle rectangulaire déjà décrit. Les images 3 et 4 de la même planche illustrent l'approximation réalisée par le modèle octogonal proposé. La comparaison établit la supériorité du modèle octogonal qui est moins sensible à l'orientation de la forme que le modèle rectangulaire et plus proche de la forme des objets.

La réduction des données, deux fois plus faible que celle obtenue avec le modèle rectangulaire de Stuck, reste malgré tout compatible avec une application temps réel.



Planche IV.4 : Comparaison des modélisations par rectangle et octogone exinscrit selon leurs définitions respectives données en IV.5.2.2 et IV.5.2.3.

# **IV.4.3- Implantation**

L'implantation de cette modélisation, qui nécessite la détermination des huit valeurs maximales et minimales ci-dessus, a été effectuée selon les deux algorithmes suivants :

- le premier suit une démarche analogue à celle de l'algorithme de Stuck [Stu80] au cours de laquelle l'image est explorée ligne par ligne ;

- le deuxième s'appuie sur une technique de suivi du contour extérieur des ensembles de points connexes de l'image CM.

La rapidité d'exécution dépend, pour chaque algorithme, du contenu de l'image CM (nombre et répartition des points). On peut cependant affirmer que le premier algorithme est en général le plus rapide. Le second algorithme n'est en effet plus rapide que lorsque l'image contient de gros amas de points. Cette situation correspond à la traversée de véhicules de dimensions importantes (autobus, camions, tramway) et constitue un événement assez rare dans les scènes étudiées.

Grâce à une exploitation poussée du matériel, le temps nécessaire pour modéliser les objets mobiles est d'environ 0,6 s (langages Pascal et PL/M) pour une scène contenant quelques véhicules et objets divers en mouvement (piétons, cycles, ...). Lorsque la scène ne contient aucun objet en mouvement on arrive à un temps voisin de 0,2 s qui correspond au temps de scrutation de l'image. Dans des situations particulièrement encombrées (plus de dix mobiles), il est possible de dépasser la seconde. Heureusement, ces situations demeurent exceptionnelles.

# **IV.5- PERFORMANCES TEMPORELLES**

Nous avons vu que l'analyse de l'image s'effectue en trois grandes étapes successives :

- l'extraction des contours en mouvement ;

- le filtrage et la mise en forme ;

- la modélisation.



Figure IV.7 : Diagramme temporel.

Chacune de ces procédures a été développée afin de minimiser le temps d'exécution. Dans ces conditions, le délai actuellement obtenu entre l'acquisition d'une image et l'obtention des modèles octogonaux des objets mobiles est d'environ 1,3 s qui se décomposent comme suit :

- 0,2 s pour l'acquisition de l'image suivante ;

- 0,24 s pour la détection des points de contours en mouvement ;

- 0,26 s pour le filtrage et la mise en forme ;

- 0,6 s pour la modélisation.

Ces différents temps sont résumés sur le chronogramme de la planche IV.5.

Le diagramme temporel de la figure IV.7 indique les délais séparant les différentes opérations ainsi que la répartition des tâches entre les deux processeurs du système. Les images précédente  $P_i$ , courante  $C_i$  et suivante  $S_i$  sont acquises par groupes



Planche IV.5 : Chronogramme des différentes phases de l'analyse de l'image C<sub>i</sub>.



IV-27

de trois et sont séparées entre elles d'un délai de 0,2 s. Le délai séparant l'image courante  $C_i$  de l'obtention des modèles octogonaux correspondant à cette image est de 1,3 s. Le délai séparant l'obtention de la modélisation de l'image  $C_i$  de l'obtention de celle de l'image  $C_{i+1}$  est de 1,5 s.

Ces temps de calcul restent supérieurs aux objectifs initiaux. Nous étudierons les possibilités d'accélération de la procédure dans le prochain chapitre où sont exposés les résultats obtenus sur différentes séquences de scènes réelles.

# - CHAPITRE V -

# APPLICATION A LA SURVEILLANCE DU TRAFIC URBAIN EN VUE DE LA CONDUITE AUTOMATIQUE D'UN TRAMWAY

La méthodologie de détection des obstacles développée dans les chapitres précédents a été appliquée à un certain nombre de séquences d'images représentant des scènes réelles de trafic urbain. Ces séquences, enregistrées sur magnétoscope, correspondent à des angles de vue ainsi qu'à des conditions d'illumination variées.

Les images présentées sont les images courantes telles que définies dans le chapitre précédent. Elles sont espacées de 1,5 s c'est à dire le temps maximum nécessaire pour extraire, de trois images successives  $P_i$ ,  $C_i$  et  $S_i$  espacées de 0,2 s, les modèles des objets en mouvement de l'image courante  $C_i$  (cf. figure IV.6).

#### V.1- ANALYSE DE DIVERSES SCENES REELLES.

# V.1.1- Séquence 1

La première séquence présentée a été filmée sur la ligne de tramway reliant Lille à Roubaix, à la hauteur d'un carrefour traversé par les voies.

La caméra était placée sur un immeuble bordant la scène observée, à une hauteur d'environ 15 m au dessus de la chaussée. Cette installation ne permettait pas d'obtenir une prise de vue à la verticale de la scène, mais, comme nous le verrons, l'inclinaison de la caméra, lorsqu'elle demeure faible, ne porte pas préjudice à la méthode.

Le champ de la caméra a été ici restreint à la zone voisine des voies du tramway.

La planche V.1 composée des images 35 à 46 correspondant à cette séquence, illustre une évolution normale de véhicules à travers le carrefour.








Dans cette scène, on discerne essentiellement le passage de trois véhicules, à savoir une voiture noire, une voiture blanche et enfin, une camionnette.

Remarquons d'abord que, quelque soit la couleur du mobile passant dans le champ de la caméra, celui-ci est détecté sans ambiguïté. Cette propriété vient de l'utilisation des opérateurs différentiels pour localiser les contours en mouvement.

Les modèles octogonaux de chacun des mobiles restent relativement stables tout au long de la traversée des véhicules, même lorsque ceux-ci changent de direction, comme c'est le cas pour la camionnette.

Enfin, il importe de noter que le fond de l'image, constitué d'une chaussée revêtue en partie de macadam, en partie de pavés, et traversée par deux voies de tramway est loin d'être uniforme. En dépit de ces importantes variations de niveau de gris et de texture du fond, les objets mobiles ont été correctement détectés.

Quelques véhicules font des apparitions fugitives en bas et à gauche de la scène. Eux aussi sont correctement détectés, à l'exception d'un camion sur les images 35 et 36. La raison de cette mauvaise détection à déjà été évoquée au cours du chapitre III. En effet, en effectuant la différence des images successives, les contours parallèles aux déplacements des objets ne peuvent plus être mis en évidence par des opérateurs différentiels.

#### V.1.2- Séquence 2

Pour la séquence de la planche V.2, composée de 9 images numérotées de 20 à 28, les conditions de prise de vue sont les mêmes que pour la séquence précédente.

Les images 23 à 28 permettent d'illustrer les résultats obtenus lors de la traversée d'un tramway. Ceux-ci sont satisfaisants malgré la grande longueur du véhicule. En effet, on pourrait craindre que des éléments du contour extérieur restent parallèles au déplacement, comme pour le camion des images 35 et 36 de la séquence précédente, et qu'ils ne soient pas détectés. En fait, l'effet de perspective et l'irrégularité du contour ont permis la détection de la totalité du contour extérieur du tramway. On notera cependant le biais apporté au modèle à cause du pantographe sur les images 26 à 28 et de l'ombre du tramway sur les images 23 à 26.



On remarquera également dans la partie supérieure gauche des images 20, 21 et 28, quelques octogones de très faibles surfaces : ceux-ci sont dûs à l'ombre portée du feuillage mobile d'un arbre situé à proximité de la voie.

#### V.1.3- Séquence 3

La planche V.3, composée de 5 images, illustre les résultats obtenus lors de la traversée d'un piéton. On voit aisément que ceux-ci peuvent être utilisés pour effectuer une poursuite automatique fiable de ce piéton tout au long de son évolution au travers du carrefour et ce malgré la faible vitesse de déplacement et le biais apporté au modèle par la présence d'une ombre portée importante.

Notons que la détection des piétons constitue une contrainte impérative pour les utilisateurs potentiels d'un tel système de surveillance.

### V.2- EFFETS DES CONDITIONS D'ECLAIRAGE

Les séquences présentées ci-dessus ont été filmées en lumière naturelle, par temps d'éclaircies accompagnées de passages nuageux. Il est apparu que de telles conditions ne sont pas les plus favorables car les variations d'éclairement de la scène sont très importantes lorsque le soleil est occulté par un nuage. De plus, comme les ombres portées nuisent à la qualité de la modélisation, les résultats sont généralement meilleurs par temps couvert.

Ce phénomène, qui peut paraître paradoxal, tient au fait que l'on segmente l'image par le mouvement lui-même plutôt que par des techniques usuelles de recherches de régions ou d'extraction de contours.

La planche V.4 montre que même sous de très mauvaises conditions de prise de vue, il est encore possible d'identifier et de modéliser correctement les objets mobiles.

L'image 1 représentant l'image courante à analyser est très pâle et on y distingue difficilement les deux automobiles et le cycliste qui traversent le carrefour. Cette mauvaise qualité de l'image est confirmée par la forme de l'histogramme de l'image 2 qui montre que toute l'information visuelle est concentrée dans une étroite plage de niveau de gris.







Malgré ce manque évident de contraste, l'opérateur de détection des contours en mouvement permet de mettre les objets mobiles en évidence sans ambiguïté, comme l'illustre l'image 3.

L'exploitation de cette image fournit les octogones représentés sur l'image 4 dont la dynamique a été portée à la pleine échelle pour faciliter la lecture. Ces résultats reflètent correctement le contenu mobile de la scène.

Notons toutefois que le seuil de binarisation de l'image CM (selon la notation adoptée au paragraphe IV.3.3), maintenu constant et égal à la valeur 10 pour les trois séquences présentées précédemment, a dû être abaissé ici à la valeur minimale 1.

### V.3- ACCELERATION DU PROCESSUS DE TRAITEMENT DES SEQUENCES.

Tous les résultats présentés ont été obtenu avec un espacement constant entre les images courantes et égal à 1,5 s. Cet intervalle est le temps maximum nécessaire pour extraire d'une image les modèles octogonaux des objets mobiles qu'elle contient. Ce temps, qui se rapproche des objectifs à atteindre dans le cadre d'une exploitation de la procédure pour la conduite automatique d'un véhicule guidé, doit encore être réduit. En effet, il faudra lui ajouter le temps d'exécution de la procédure de compréhension de la scène et de génération des ordres de conduite [Bou88].

Afin d'accélérer le processus de traitement des séquences d'images, une première solution consisterait à tenter de réduire le temps total qui sépare l'acquisition d'une image courante C; de la mise à disposition des modèles des objets mobiles.

Ce temps, qui est actuellement de l'ordre de 1,5 s, pourrait effectivement être réduit en utilisant, par exemple, des processeurs câblés pour effectuer les phases de détection des points de contour, de filtrage et de mise en forme. Ce travail est actuellement en cours de développement au Centre d'Automatique.

Il est une autre approche au problème de l'accélération du processus qui s'appuie sur une distinction entre le " temps de réaction " et le " temps de réactualisation ".

On définit le " temps de réaction " comme l'intervalle de temps qui s'écoule entre l'acquisition d'une l'image et l'obtention des modèles des mobiles de cette image. Ce temps est actuellement de 1,3 s (*cf.* Figure IV.6).



Planche V.4 : Résultat sous mauvaises conditions de prise de vue.

On définit le "temps de réactualisation" comme l'intervalle de temps qui s'écoule entre l'obtention des modèles des mobiles correspondant à une image  $C_i$  et ceux correspondant à l'image  $C_{i+1}$  lui succédant. Ce temps est actuellement de 1,5 s (cf. Figure IV.6).

On peut envisager de réduire ce temps en confiant à deux processeurs travaillant en parallèle, la phase de détection des points de contours en mouvement et de filtrage, appelée phase de traitement d'une part, et la phase de modélisation d'autre part.

Le système de vision utilisé permettrait d'envisager une telle implantation. En effet, la phase de traitement peut être confiée au processeur d'images (cartes IP-512) pendant que l'unité centrale du calculateur hôte (carte iSBC 286/12) effectue les calculs nécessaires à la modélisation.

On aboutit ainsi au diagramme temporel de la figure V.1 qui peut être comparé à celui de la figure IV.6. On constate que cette implantation permet d'abaisser le temps qui s'écoule entre deux actualisations des modèles à 0,6 s.



Figure V.1 : Diagramme temporel proposé.

Cette minimisation du temps de réactualisation est très importante car elle permet de réduire les variations des modèles entre les différentes images analysées. Plus ce temps sera réduit et plus facile et fiable sera la procédure de suivi des mobiles et de prédiction de leurs déplacements [BOU88].

Cette modification de la mise en oeuvre de la procédure globale nécessiterait un système de génération d'interruptions au niveau des cartes images qui n'existe malheureusement pas sur le modèle dont nous disposons. Il serait pourtant facile pour le constructeur d'utiliser le registre d'état des cartes IP-512 pour générer ce signal d'interruption.

#### **V.4- CONCLUSION**

Les séquences d'images présentées dans ce chapitre ont permis de mettre en évidence les qualités et les défauts de la procédure de traitement des séquences d'images dynamiques sur des scènes de trafic réelles.

Les images présentées ne constituent qu'une faible fraction des séquences enregistrées sur sites réels. Elles représentent des flots de circulation moyens. L'analyse de tous les enregistrements a mis en évidence les limites de la méthode lorsque les carrefours sont embouteillés. En effet, dans de telles conditions, il devient difficile sinon impossible, d'individualiser les modèles des véhicules qui circulent serrés les uns contre les autres. Il est cependant important de noter que, en plaçant la caméra exactement à l'aplomb de la scène observée, il serait plus facile de différencier les véhicules rapprochés.

# CONCLUSION

Dans cette étude, nous avons présenté une méthode d'analyse de séquences d'images orientée vers l'interprétation de scènes dynamiques où évoluent des objets quelconques, sur fond non uniforme, et sous éclairage non contrôlé.

Les procédures proposées ne nécessitent aucune connaissance *a priori* sur les objets mobiles, leur nombre, leurs trajectoires, ni sur le fond de l'image. En particulier, aucune image de référence représentant ce fond n'est nécessaire.

Une attention très particulière a été apportée à la robustesse des procédures mises au point qui sont peu sensibles aux variations d'éclairage des scènes observées ainsi qu'aux variations d'aspect des mobiles dues aux effets de perspectives au cours de leurs mouvements

L'approche utilisée s'appuie sur une procédure de détection des contours en mouvement. En considérant les différences entre trois images successives prises deux à deux, on met en évidence les objets en mouvement ainsi que les régions du fond découvertes ou recouvertes par ces objets pendant leurs mouvements. Les gradients de ces différences d'images successives sont ensuite combinés afin d'isoler uniquement les contours des objets mobiles, même lorsque le fond de l'image n'est pas uniforme.

Ces contours sont ensuite modélisés par des octogones exinscrits qui permettent de limiter le nombre de paramètres nécessaires pour décrire l'évolution dynamique de la scène.

Au niveau de l'implantation de ces procédures, il est apparu judicieux de distinguer le temps nécessaire pour obtenir les modèles octogonaux relatifs à une image, du temps de réactualisation de ces modèles.

En effet, on pourrait tirer parti de l'architecture multiprocesseur des systèmes de traitement d'images pour effectuer en parallèle les procédures de bas niveau qui permettent d'extraire les contours en mouvement, et les procédures de plus haut niveau qui permettent de modéliser ces contours. Une telle implantation, bien qu'elle ne

modifierait pas le temps s'écoulant entre l'acquisition d'une image et sa modélisation, permettrait de réactualiser les modèles à une plus haute cadence. C'est pourquoi, nous avons proposé une modification très simple des matériels qui permettrait cette nouvelle implantation.

Les développements théoriques et leur implantation sur machine de vision orientée temps réel ont été validés dans le cadre d'une étude sur la surveillance du trafic routier au niveau de carrefours urbains.

De nombreuses scènes, enregistrées sur bande magnétique, ont ainsi été analysées en temps réel. Les résultats obtenus ont montré que les procédures proposées permettent de décrire le "contenu mobile" de scènes relativement complexes où évoluent des véhicules de différentes tailles et couleurs, des tramways, des piétons et des cyclistes.

Dans l'optique d'une conduite automatique d'un système de transport guidé de type tramway, les procédures développées ne permettent pas encore d'envisager un temps de réaction sufissamment bref puisqu'il s'écoule environ 1,5 s entre l'acquisition de l'image et l'obtention des modèles des mobiles. C'est la raison pour laquelle, les algorithmes de bas niveau proposés dans cette étude sont actuellement en cours d'implantation sur des processeurs frontaux spécialisés. Toute la phase d'extraction des contours pourra ainsi être réalisée effectivement en temps réel.

Par ailleurs, les récents progrès technologiques au niveau des microprocesseurs (passage du microprocesseur Intel 286 au 386) permettront à très court terme d'abaisser notablement le temps de calcul des modèles octogonaux.

Sur un plan plus fondamental, quelques points restent encore à explorer parmi lesquels nous mentionnerons le problème des mouvements de la caméra lorsque celle-ci est fixée à une structure légère telle qu'un poteau électrique. Ces mouvements mettraient la méthode complètement en défaut. Une procédure de recalage de l'image serait dans ce cas absolument nécessaire et son implantation est envisagée à très court terme.

Notons également que, bien que l'ajustement du seuil qui permet d'extraire les points de contours des objets en mouvement ne soit pas critique, il serait intéressant d'implanter une procédure de sélection automatique de ce seuil. En effet, les séquences analysées, bien qu'enregistrées sous éclairage naturel et variable, ne reflètent pas l'ensemble des conditions atmosphériques envisageables. Rien ne garantit donc que le seuil choisi arbitrairement et maintenu constant pour toutes les séquences présentées, soit encore valable sous des conditions de prise de vue plus défavorables. Une approche originale pour la sélection de ce seuil est actuellement en cours de validation au Centre d'Automatique.

Le dernier point important soulevé par cette approche réside dans les possibilités de fusion des modèles octogonaux lorsque des mobiles évoluent de manière très rapprochée. Ce phénomène, qui peut être limité en plaçant la caméra exactement à la verticale du carrefour, ne peut être totalement éliminé à cause des ombres portées qui accompagnent les mobiles.

Ces ambiguïtés dans l'analyse du mouvement ne seront levées qu'en implantant des procédures de suivi des mobiles qui intégreront des connaissances *a priori* sur le comportement dynamique des véhicules et sur les règles de conduite dictées par le code de la route. Cette phase de "compréhension" du mouvement constitue l'objet d'un important travail qui sera bientôt présenté à l'Université [BOU88].

## BIBLIOGRAPHIE

- Abd78 ABDOU I.E., "Quantitative methods of edge detection ", USCIPI, Report 830, University of Southern California, L.A., USA, 1978.
- Abr82 ABRAMCZUK T., " A TV-detection for monitoring traffic flow ", Final report on feasibility studies, Royal Institute of Technology, Stockolm, TRITA-ILA, Janvier 1982.
- Abr84 ABRAMCZUK T., "A Microcomputer based TV-Detector for Road Traffic", Symposium on Road Research Program, OECD, Tokyo, Japan, Octobre 1984.
- Adi83 ADIV G., "Recovering Motion Parameters in Scenes Containing Multiple Moving Objects", *IEEE*, Computer and Information Sciences Dept, Univ. Massachusetts, 1983.
- Adi85 ADIV G., "Determinig 3-D motion and structure from optical flow generated by several moving objects", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 7, N°4, pp. 384-401, Juin 1985.
- Agg75 AGGARWAL J.K., "Computer analysis of moving polygonal images ", *IEEE Computer*, Vol C-24, N°10, pp. 966-976, Octobre 1975.
- Agg81 AGGARWAL J.K., "Correspondence processes in dynamic scene analysis", *IEEE PAMI*, vol 69, n°5, Mai 1981.
- Agg86 AGGARWAL J.K. "Determining Motion Parameters Using Intensity Guided Range Sensing", *Pattern Recognition*, Vol. 19, pp. 169-179, 1986.
- Aok84 AOKI M., "Detection of Moving Objects using Line Image Sequence", 7<sup>th</sup> International Conference on Pattern Recognition, 1984.
- Asa80 ASADA M., YACHIDA M. et TSUJI S., "Three dimensional motion interprétation of the sequence of line drawings", *Proceedings of the International Conference on Pattern Recognition*, pp. 1266-1273, 1980.
- Att57 ATTENEAVE F., "Physical determinations of the judged complexity of shapes ", J. Exp. Psychol, Vol. 53, N°4, pp. 221-227, Avril 1957.
- Baj84 BAJON J., "A concavity characterization method for digital objects ", Sig. Proc. Vol. 9, 1985.
- Bal83 BALLARD D.H. et KIMBALL O.A., "Rigid body motion from depth and optical flow ", Computer Vision Graphics and Image Processing, Vol. 22, N°1, pp. 95-115, Avril 1983.
- Bar86 BARBA D., "Automatic Region Construction by Edge Detection and Contour Following in Image segmentation", 8<sup>th</sup> Int. Conf. on Pat. Rec., 1986.

- Bar79 BARNARD S.T. et THOMPSON W.B., "Disparity estimation using feature point matching", Workshop on Computer Analysis of Time-Varying Imagery, Ed. Aggarwal J.K. et Badler N.I., Philadelphia, PA, pp. 2, Avril 1979.
- Bar80 BARNARD S.T. et THOMPSON W.B., "Disparity analysis of images ", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 2, n°4, pp. 333-340, Juillet 1980.
- Bas79 BASSEVILLE M., "Détection de contours : méthodes et études comparatives ", Annales des Télécommunications, Vol. 34, N°11-12, 1979.
- Ber80 BERS K.H., "Object detection in image sequences ", IEEE 5th Int. Conf. on Pat. Rec., 1980.
- Blo87 BLOSSEVILLE J.M. et coll. "TITAN : Système de mesure de trafic par analyse d'image ", Rapport INRETS, Arcueil, 1987.
- Bou86a BOUTHEMY P., "A Method of Integrating Motion Information Along Contours Including Segmentation", 8th Conf. Pat. Rec., 1986.
- Bou86b BOUTHEMY P., "Un nouveau schéma d'estimation du champ des vitesses sur les contours dans une séquence d'images ", Sem. Int. Im. Elec. Avril 1986.
- Bou86c BOUTHEMY P., "Un nouveau schéma d'estimation du champ des vitesses sur les contours dans une séquence d'images ", 2e Coll. Image CESTA, Nice, pp.237-245, Avril 1986.
- Bou87 BOUTRY F., KUHN F., DEPARIS J.P. et POSTAIRE J.G., "New tendancies in L.R.T. system : Automatic driving by means of computer vision ", *International Conference and Exhibition on Local Public Transport*, Liverpool, Novembre 1987.
- Bou88 BOUTRY F., "Analyse de scènes dynamiques par représentatoin espace-temps des éléments mobiles. Application à la conduite des véhicules guidés ", *Thèse de l'Université*, Université des Sciences et Techniques de Lille Flandres Artois, A paraître : Mai 1988.
- Bri81 BRILLAULT B. et MERCIER H., "Eléments de la théorie morphologie mathématique sur des images binaires ", Rapport de stage, Centre Edf, Clamart, Juillet 1981.
- Bur82 BURT P.J., "Local correlation measures for motion analysis A comparative study", *IEEE Conf. Publi.*, Vol 14, Las Vegas, 1982.
- Caf76 CAFFORIO C., " Methods for measuring small displacement of television images, IEEE Transf. Info. Theory, Vol IT 12, n°2, Septembre 1976.
- Chi75 CHIEN R.T. et JONES V.C., "Acquisition of moving objects and hand-eye coordination", Proceedings of the 4th International Joint Conference on Artificial Intelligence, Tbilissi, URSS, pp. 737-741, Septembre 1975.
- Clo80 CLOCKSIN W.F., "Perception of surface slant and edge label from optical flow : a computational approach ", *Perception*, Vol. 9, pp. 253-269, Novembre 1980.

- Cou84 COURTNEY J.W., "Robot Guidance Using Computer Vision", Pattern Recognition, Vol.17, pp. 585-592, 1984.
- Dav75 DAVIS L.S., "A survey of edge detection techniques ", Computer Graphics and and Image Processing, Vol. 4, pp. 248-270, 1975.
- Dav79 DAVIS L.S., WU Z. et SUN H., " Contour-based motion estimation ", Computer Graphics Image Processing, Vol. 9, 1979.
- Dep84 DEPARIS J.P., KUHN F., "Perspectives de Développement des Automatismes dans les Transports Guidés de Surface. 1ère Partie : Résultats d'une étude sur les conditions de circulation du Mongy ", Rapport Interne INRETS/CRESTA, Juin 1984.
- Dic84 DICKINSON K.W., "Image Processing Applied to Traffic: 1. A General Review", Engineering & Control, Janvier 1984.
- Dra85 DRAKE K.E., "Sensing Error for a Mobile Robot Using Line Navigation ", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-7, pp. 485-490, 1985.
- Dre78 DRESCHLER L., "Using "Affinity" for extracting Images of Moving Objects from TV-Frame Sequences ", Institut für Informatik, Hamburg, Février 1978.
- Dre83 DRESCHLER L. et NAGEL H.H., "Volumetric model and 3-D trajectory of a moving car derived from monocular TV-frame sequence of a scene street", Int. Joint Conf. on Artificial Intelligence, Vancouver, Canada, Août 1981.
- Dre83 DRESCHLER L. et NAGEL H.H., "On the selection of critical points and local curvature. Extrema of region boundary for interframe matching", *Image sequence processing*, Ed. T.S. Huang, Berlin, pp. 457-470, 1983.
- Dud73 DUDA R. et HART P., "Pattern classification and scene analysis ", J. Wiley and Sons, N.Y., Chap. 9, pp. 327-377, 1973.
- End71 ENDLICH R.M., "Use of a pattern recognition technic for determining cloud motions from sequences of satellite photographs ", J. Appl. Meteorol., Vol. 10, pp. 105-117, Février 1971.
- Esk78 ESKENAZI R. et CUNNINGHAM R.T. "Real-time tracking of moving objects in TV-images ", *IEEE Worshop on Pattern Recognition and Artificial Intelligence*, Princeton, N.J., pp.4-6, Avril 1978.
- Fac87 FACON J., "EDOSI: Un système d'étude du déplacement d'objets à partir de séquences d'images ", *Thèse de doctorat*, Université de Compiègne, Novembre 1987.
- Fan84 FANG J.Q. et HUANG T.S., "Solving 3-D small-rotation motion equations : uniqueness algorithms and numerical results", Computer Graphics and Image Processing, Vol. 26, pp. 183-206, 1984.

- Fen79 FENNEMA C.L. et THOMPSON W.B. "Velocity determination in scenes containing several moving objects", Computer Graphics and Image Processing, Vol. 9, pp. 301-315, Avril 1979.
- Fer82 FERRIE F.P., "Cell Tracking: A Modeling and Minimization Approach ", *IEEE Trans. on Pattern Analysia and Machine Intelligence*, Vol. PAMI, Mai 1982.
- Fre78 FREEMAN H., "Shape determination via the use of critical points", Pattern Recognition, Vol. 10, pp. 159-166, 1978.
- Fri72 FRISBY J.P., "The effect of stimulus orientation on the phi phenomenom ", Vision Research, Vol. 12, pp.1145-1166,1972.
- Gam84a GAMBOTTO J-P., "Correspondence analysis for target tracking in infrared images", 7th Conf. on Pattern Recognition, 1984.
- Gam84b GAMBOTTO J-P. "Motion analysis of targets in infrared sequences", 7th Conf. on Pattern Recognition, 1984.
- Gen77 GENNERY D.B., " A stereo system for an autonomous vehicule ", Proceedings of the 5<sup>th</sup> International Conference on Artificial Intelligence, pp.576-582, Août 1977.
- Gen79 GENNERY D.B., "Object detection and measurement using stereo vision", Proceedings 6<sup>th</sup> Joint International Conference on Artificial Intelligence, Août 1979.
- Gen79b GENNERY D.B., "Stereo-camera calibration", Proceedings of the Image Understanding Workshop, pp. 101-107, Novembre 1979.
- Gil80 GILBERT A.L., GILES M.K., FLACHS G.M., "A real-time video tracking system ", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-2, N<sup>°</sup>1, Janvier 1980.
- Gil85 GILBERT A.L. et al, "A real-time video tracking system using image processing techniques", Proceedings of SPIE, Photo and Electro-optics in Instrumentation, Vol. 134, Mars 1985.
- Gir79 GIRALT G., SOBEK R. et CHATILA R., "A multilevel planning and navigation system for a mobile robot : a first approach to HILARE ", Proceedings 6<sup>th</sup> Joint International Conference on Artificial Intelligence, Août 1979.
- Gos83 GOSH S.K., "A linear time algorithm for obtaining the convex hull of a simple polygon ", *Pattern Recognition*, N°6, 1983.
- Gos85 GOSHTASBY A.S., "Point pattern matching using convex hull edges", IEEE Trans. Systems Man and Cybernetics, Septembre 1985.
- Han74 HANNAH M.J., "Computer matching of areas in stereo images", Report AIM-239, Stanford University, Juin 1974.
- Has72 HASKELL B.G. et LIMB J.O., "Predictive video encoding using measured subject velocity", U.S. Patent 3.632.865., Janvier 1972.

- Has74 HASKELL B.G., "Frame to frame coding of television pictures using twodimensional Fourier transform", *IEEE Transactions on Information Theory*, IT-20, pp. 119-120, 1974.
- Hay83 HAYNES S.M., "Detection of Moving Edges", Computer Graphics and Image Processing, Vol. 21, 1983.
- Hil78 HILBERT E.E., et al, "Wide area detection system-conceptual design study ", JPL Interim Report for FHWA, Report FHWA-RD-77-86, Février 1978.
- Hil80a HILBERT E.E., et al, "A Sensor for Control of Arterial and Networks.", IEEE Transaction on Vehicular Technology, Vol. VT-29, Mai 1980.
- Hil80b HILBERT E.E., et al, "Sensor for Control of Arterial and Networks ", Final Rep. FHWA/RD-80/024, Octobre 1980.
- Hor81 HORN B.K.P, "Determining optical flow ", Artificial Intelligence, Vol 17, 1981.
- Hou87 HOUGHTON A., HOBSON G.S., SEED L. and TOZER R.C., "Automatic monitoring of vehicles at road junctions", *Traffic Engineering and Control*, Octobre 1987.
- Hou80 HOUKES A., "Measurement of speed and time headway of motor vehicles with video camera and computer", Proc. of 6<sup>th</sup> Int. Conf. on Digital Computer Applications to Process Control, Düsseldorf, Octobre 1980.
- Hu62 HU M.K., "Visual pattern recognition by moment invariants" *IEEE Trans. Inform. Theory*, pp. 179-187, 1962.
- Hua81b HUANG T.S. et TSAI R.Y, "Image sequence analysis: Motion estimation", Springer Verlag, New-York, Chap. 1, 1981.
- Hua81 HUANG T.S., "Image Sequence Analysis: what can whe learn from applications ?", Image Sequence Analysis, pp. 104-116. T.S HUANG, Springer Verlag, Berlin, 1981.
- Huf80 HUFFMAN D.D. et FLINGHBAUGH B.E., "The interpretation of biogical motion", *Technical Report on Artificial Intelligence*, Memo 608, Massachusetts Institute of Technology, Décembre 1980.
- Ini82 INIGO R.M., McVEY E.S. et MOORE J.W., "Application of image processing to vehicle guidance and safety", *Final Report*, Contract DTRS5680-C-00033, DOT/RSPA/DMA-50/82/7, Septembre 1982.
- Ini84 INIGO R.M., McVEY E.S., BERGER B.J. et WIRTZ M.J., "Machine Vision Applied to Vehicle Guidance ", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, N°6, Novembre 1984.
- Ini85 INIGO R.M., "Traffic Monitoring and Control using Machine Vision : A survey ", IEEE Transactions on Industrial Electronics, Vol. IE-32, N°3, Août 1985.

- Jac80 JACOBUS C.J., CHIEN R.T. et SELANDER J.M., "Motion detection and analysis of matching graphs of intermediate-level primitive", *IEEE Transactions* on Pattern Analysis and Machine Intelligence, Vol. 2, N°6, pp. 495-510, Novembre 1980.
- Jai77 JAIN R., MILITZER D. et NAGEL H.H., "Separating non-stationary from stationary scene components in a sequence of real world TV-images", *Proceedings of the 5<sup>th</sup> Int. Conf. on Artificial Intelligence*, Cambridge, Massachusetts, pp. 612-618, Août 1977.
- Jai79a JAIN R., "On the Analysis of Difference Pictures from Image Sequences of Real World Scenes ", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Avril 1979.
- Jai79b JAIN R., "Extraction of moving objects images trough change detection", Proceedings of the 6th International Conference on Artificial Intelligence, pp. 425-428, 1979.
- Jai79c JAIN R., "Segmentation through the Detection of Change Due to Motion", Computer Graphes and Image Processing, Vol. 11, 1979.
- Jai81a JAIN R., "Dynamic Scene Analysis Using Pixel-Based Processes ", Computer, Août 1981.
- Jai81b JAIN R., "Extraction of Motion Information from Peripheral Processes ", IEEE Trans. PAMI, Septembre 1981.
- Jai84 JAIN R., "Difference and Accumulative Difference Pictures in Dynamic Scene Analysis", Image & Vision Computer., Mai 1984.
- Jer84 JERIAN C. and JAIN R., "Determining motion parameters for scheme with translation and rotation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, N°4, pp. 523-529, Juin 1984.
- Joh73 JOHANSSON G., "Visual perception of biological motion", Perception and psychology, Vol. 14, N°2, pp. 202-211, 1973.
- Kap83 KAPPAGANTULA S., "Motion Compensated Interframe Image Prediction", IEEE Trans. on Communications, Septembre 1985.
- Kea82 KEARNEY K.K., "Gradient-based estimation disparity", *IEEE Conf. Publi.*,vol 14, Juin 1982.
- Kir71 KIRSCH R. "Computer determination on the constituent structure of biological images", Computer and Biomedical Research, Vol. 4, N°3, pp. 315-328, USA, 1971.
- Kos77 KOSLOWSKI L.T., "Recognizing the sex of a walker from dynamic point-light display", *Perception and Psychology*, Vol. 21, N°6, pp. 575-580, 1977.
- Kor80 KORN A., "Motion Analysis in Natural Scenes Picked up by a Moving Optical Sensor ", 5th Int. Conf. on Pattern Recognition, Miami, 1980.

- Kud81 KUDO Y., "Traffic Flow Measurement System using Image Processing", National Convention of IIP-Japan, 1981.
- Kul83 KULPA Z., "More about areas and perimeters of quantized objects", Computer Graphics and Image Processing, Vol. 22, N°2, pp. 268-276, Mai 1983.
- Law83 LAWTON D.T., "Constraint-based inference from image motion", Proceedings of the 1<sup>st</sup> annual Conference on Artificial Intelligence, Stanford University, Août 1980.
- Law83 LAWTON D.T., "Processing translational motion sequences", Computer Vision Graphics and Image Processing, pp. 116-144, Avril 1983.
- Lee80 LEE D.N., "The optic flow field: the foundation of vision", *Philosophical* Transactions of Royal Society of London, Vol. B, N° 290, pp. 169-179, 1980.
- Lee70 LEESE J.A., "The detection of cloud pattern motions from geosynchronous satellite image data ", *Pattern Recognition*, Vol 2, 1970.
- Lee71 LEESE J.A., "Analysis automatic technique for obtaining cloud motion from geosynchronous satellite data using cross-correlation", J. Appl. Meteorol., Vol. 10, pp. 118-132, Février 1971.
- Len83 LENZ R. et GERHARD A.," Image sequence coding using scene analysis and spatio-temporal interpolation ", *Image Sequence Processing*, Ed. T.S. Huang, pp. 663-709, 1983.
- Lev83 LEVINE M.D., "A rule-based system for characterizing blood cell motion", Image sequence processing, Ed T.S. HUANG, pp. 663-709, 1983.
- Lil72 LILLESTRAND R.L., "Technics for change detection", *IEEE Transactions on Computer*, Vol. C-21, pp. 654-659, 1972.
- Lon80 LONGUET-HIGGINS H.C. et PRAZDNY K., "The interpretation of a moving retinal image", *Proceedings of the Royal Society of London*, Vol. B, N°208, pp. 385-397, 1980.
- Ma87 MA J., "A motion constraint equation under space-varying or time-varying illumination ", *Pattern Recognition Letters*, Vol. 5, pp. 203-205, 1987.
- Mai84 MAITRE H. et LOPEZ K.J., "Transformation de Hough : théorie et application à la détection de symétries circulaires ", 4<sup>e</sup> Congrès de Reconnaissances des Formes, Paris, Janvier 1984.
- Man87 MANTAS J., "Methodologies in pattern recognition and image analysis a brief survey ", *Pattern Recognition*, Vol. 20, n°1, pp. 1-6, 1987.
- Mar77 MARR D. et POGGIO T., "A theory of human stereovision", Artificial Intelligence Memo 451, Massachusetts Institute of Technology, Novembre 1977.
- Mar79b MARR D. et HILDRETH E., "Theory of edge detection ", MIT AI Memo 518, Avril 1979.

- Mar79 MARR D. et ULLMAN S., "Directional selectivity and its use in early visual processing", MIT, AI, Memo 524, Juin 1979.
- Mar78 MARTIN W.N. et J.K. AGGARWAL, "Dynamic scene analysis: Survey", Computer Graphics Image Processing, Vol. 7, pp. 356-374, 1978
- McV86 MC VEY E.S., "Range Measurement by a Mobile Robot Using a Navigation Line ", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-8, pp. 105-109, 1986.
- Med85 MEDIONI G et YASUMOTO Y., "Robust estimation of 3-D motion parameters of sequence of image frames using regularization", Proceedings of Images Understanding Workshop, DARPA, Miami Beach, pp. 117-128, Décembre 1985.
- Mei80 MEIRI A.Z., "On monocular perception of 3-D moving objects", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-2, pp. 582-583, 1980.
- Mit84 MITICHE A. "Computation of optical flow and rigid motion", Proceedings Workshop on Computer Vision: Representation and Control, Annapolis, MD, pp. 63-71, 1984.
- Mit86 MITICHE A., "On kineopsis and computation of structure and motion ", Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, n°1, Jan. 86.
- Mit87 MITICHE A. "Experiments in computing optical flow with the gradient-based, multiconstraint method ", *Pattern Recognition*, Vol. 20, n°2, pp. 173-179, 1987.
- Mod77 MODESTINO J.W. et FRIES R.W. "Edge detection in noisy images using recursive digital filtering", *Computer Graphics and Image Processing*, Vol. 6, pp. 409-433, USA, 1977.
- Mor77 MORAVEC H.P., "Towards Automatic Visual Obstacle Avoidance", Proceedings of the 5<sup>th</sup> Int. Conf. on Artificial Intelligence, pp. 584, Août 1977.
- Mor79 MORAVEC H.P., "Visual mapping by a robot Rover", Proceedings of the Int. Joint Conf. on Artificial Intelligence ", Tokyo, Japan, pp. 598-600, 1979.
- Mor80 MORAVEC H.P. "Obstacle avoidance and navigation in the real world by seeing robot Rover", PhD Dissertation, Stanford University, Stanford, 1980.
- Mor83 MORAVEC H.P., "The stanford Cart and the CMU Rover", Proceedings of the IEEE, Vol. 71, N°7, Juillet 1983.
- Mot8? MOTT D.H. " An Initial Study into Bus Detection in London Traffic ", source non disponible.
- Mun68 MUNSON J.H., "Experiments in the recognition of hand printed text : part 1character recognition", *AFIPS Conf. Proc.*, Washington D.C., Thompson, Vol. 33, pp. 1125-1138, 1968.

- Nag78 NAGEL H.H., "Formation of an object concept by analysis of systematic time variations in the optically perceptible environment", *Computer Graphics and Image Processing*, Vol.7, pp. 149-194, Avril 1978.
- Nag79 NAGEL H.-H., "Evaluation of several dissimilarity grading approaches for the analysis of image sequences", *IEEE Workshop on Computer Analysis of Time-Varying Imagery*, Philadelphia, PA., Avril 1979.
- Nag81a NAGEL H.-H., "On the derivation of 3-D rigid point configuration from image sequences ", Proc. of IEEE Conf. of Pattern Recognition and Image Processing, Dallas, Texas, Août 1981.
- Nag81b NAGEL H.-H. et NEUMANN B. " On 3-D reconstruction from two perspective views ", Int. Joint Conf. on Artificial Intelligence, Vancouver, Août 1981.
- Nag86 NAGEL H.-H. "Image Sequences -Ten (Octal) Years- from Phenomenology Towards a Theorical Foundation ", 8<sup>th</sup> Conf. Pat. Rec., 1986
- Nèg83 NEGRE R., SAMY R. et JOLY R., "Traitement d'images spatio-temporelles et son application à la poursuite par une caméra video ", *9e Colloque de Traitement du Signal et ses Applications*, Nice, Mai 1983.
- Neu80 NEUMANN B., "Motion analysis of image sequences for object grouping and reconstruction", *Proceedings of the International Conference on Pattern Recognition*, pp. 1261-1265, 1980.
- Neu86 NEUVO Y., "New Types of Edges Detectors Based on Linear and Median Operators", 8th Conf. on Pattern Recognition, 1986.
- Ono73 ONOE M., HAMANO N. et JAZUHIKO O., "Computer Analysis of Traffic Flow Observed by Subtractive Television", Computer Graphics and Image processing, Vol. 2, pp. 377-392, 1973.
- Orl85 ORLOWSKI M., "On the conditions for success Sklansky's convex Hull algorithm", *Pattern recognition*, Vol. 16, N°6, pp. 579-586, 1983.
- Orl85 ORLOWSKI M., "A convex hull algorithm for planar simple polygons ", Pat. Recognition, Vol. 5, 1985.
- Oro80 O'ROURKE J. et BADLER N.I. "Model-based image analysis of human motion using constraint propagation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 2, N°6, pp. 522-536, Novembre 1980.
- Oro81 O'ROURKE J. "Motion detection using Hough techniques", Proceedings of IEEE Conf. on Pattern Recognition and Image Processing, pp. 82-87, 1981.
- Osc82 OSCARSSON E., "TV-Camera Detecting Pedestrians for Traffic Light Control", Acta Imeko, 1982
- Oza86 OZAWA S. et ROSENFELD A., "Synthesis of a Road Image as Seen from a Vehicle", *Pattern Recognition*, Vol. 19, pp.123-145, 1986.

- Pos86 POSTAIRE J.G., STELMASZYK P. et BONNET P. "A visual surveillance system for traffic collision avoïdance", 5<sup>th</sup> IFAC/IFIP/IFORS International Conference on Control in Transportation Systems, Vienne, 1986.
- Pot75a POTTER J.L., "Velocity as a cue to segmentation", IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-5, pp. 390-394, Mai 1975.
- Pot75b POTTER J.L., "Scene segmentation by velocity measurement obtained with a cross-shaped template ", Proceedings of the 4<sup>th</sup> Int. Joint Conf. on Artificial Intelligence, Tbilissi, URSS, Septembre 1975.
- Pot77 POTTER J.L., "Scene segmentation using motion information", Computer Graphics Image Processing, Vol. 6, pp. 558-581, Décembre 1977.
- Pra79 PRAGER J.N, "Segmentation of static and dynamic scene ", *Technical Report*, 79-7, Dept. of Computer and Information Science, University of Massachusetts, Mai 1979.
- Pra78 PRATT W.K., Digital Image Processing, Wiley. Interscience, N.Y, 1978.
- Pra80 PRAZDNY K., "Egomotion and relative depht map from optical flow", *Biology* and Cybernetics, Vol. 36, pp. 97-102, 1980.
- Pra81b PRAZDNY K., "Determining the instaneous of motion from optical flow generated by a curvilinearly moving observer", *Computer Graphics and Image Processing*, Vol. 17, pp. 238-248, 1981.
- Pra81 PRAZDNY K., "A simple method for recovering relative depth map in the case of a translating sensor ", *Proc. 7th IJCAI*, Vancouver, Août 1981.
- Pre70 PREWITT J.M.S., "Object enhancement and extraction ", *Picture processing and psychopictorics.*, B.S. Lippking et A. Rosenfeld, Ed. Academic press, N.Y., 1970.
- Pri75 PRICE K.E. et REDDY R.," Change detection in multi-sensor images ", 10<sup>th</sup> Int. Symposium on Remote Sensing of Environment, A. Arbor, Michigan, Octobre 1975.
- Pri77 PRICE K.E. et REDDY R.," Change detection and analysis in multispectral images ", *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 619-625, 1977.
- Raj83 RAJALA S.A., "Application of the One-Dimensional Fourier Transform for Tracking Moving Objects in Noisy Environments", Comp. Vis. Graph. & Im. Proc., Vol. 21, 1983.
- Ras79 RASHID R.F., "LIGHTS: A study in motion", DARPA Image Understanding Workshop, pp.57-68, 1979.
- Ras80 RASHID R.F., "Towards a system for the interpretation of a moving light displays", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 2, N°6, pp. 574-581, Novembre 1980.
- Riv86 RIVES G. "Planar Partially Occluded Objects Scene Analysis", 8<sup>th</sup> Conf. on Pattern Recognition, 1986.

- Roa79 ROACH J.W., "Computer tracking of objects moving in space ", *IEEE PAMI*, vol 1, n°2, Avril 1979.
- Roa80 ROACH J.W. et AGGARWAL J.K., "Determining the movement of object from a sequence", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-2, N<sup>o</sup>6, pp. 554-562, Novembre 1980.
- Rob65 ROBERTS L.G., "Machine perception of 3-D solids. In optical and electro optical information processing", Tipet J.T. et coll. Ed. MIT Press, Cambridge, Massachusetts, USA, pp. 159-197, 1965.
- Rob77 ROBINSON G.S., "Edge Detection by compass gradient mask", Computer Graphics and Image Processing, Vol. 6, pp. 492-501, 1977.
- Ros80 ROSEN D., "On the areas and boundaries of quantized object", Computer Graphics and Image Processing, Vol. 13, N°1, pp. 94-98, Mai 1980.
- Ros71 ROSENFELD A. et THURSTON M., "Edge and curve detection for visual scene analysis", *IEEE Trans. on Computers*, Vol. C-20, pp. 562-569, 1971.
- Ros75 ROSENFELD A. et WESZKA J.S., "An improved method of angle detection on digital curves ", *IEEE Trans. on Computers*, Septembre 1975.
- Ros76 ROSENFELD A., "Scene labelling by relaxation operations ", IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-6, pp.420-433, 1976.
- Ros76b ROSENFELD A. et KAK A., Digital Picture Processing, Academic Press, N.Y., 1976.
- Sam85 SAMY R., "Real-time tracking of target moving on natural textured background", 2nd International Technical Symposium. on Optical and Electro-Optical Applied Science and Engineering, Cannes, Décembre 1985.
- Sam86 SAMY R. et BOZZO C.A., "Deux approches dans la segmentation de scènes dynamiques texturées ", 2<sup>e</sup> Colloque Image CESTA, Nice, Avril 86.
- Sch71 SCHMIDT R.A., " A study of the real-time control of a computer driven vehicle ", Ph.D. dissertation, Dep. Computer sciences, Stanford University, Stanford, CA, Août 1971.
- Sch82 SCHLUTMEYER et al., "Wide Area Detection System (WADS) ", JPL publ. 83-15, Final Report FHWA, Septembre 1982.
- Sha75 SHAMOS M., "Problems in computational geometry", Carnegie Mellon University Press, Pittsburgh, 1975.
- Sha78 SHAMOS M., "Computational geometry", PhD dissertation, YALE University, 1978.
- Sha86 SHARIAT H., "The motion problem : How to use more than two frames", Institute for Robotics and Intelligent Systems, L.A., California, Octobre 1986.
- Skl70 SKLANSKY J., "Recognition of convex blobs ", Pattern Recognition, Vol. 2, pp. 3-10, Janvier 1970.

- Skl72b SKLANSKY J., CHAZIN R.L. et HANSEN B.J., "Minimum perimeter polygons of digitized silhouettes", *IEEE Trans. on Computers*, C-21, pp. 260-268, Mars 1972.
- Skl72 SKLANSKY J., "Measuring concavity on a rectangular mosaic", *IEEE Trans* Comp., Décembre 1972.
- Smi72 SMITH S.P., "Automatic cloud tracking using precisely aligned digital ATS pictures ", *IEEE Transactions on Computer*, Vol. 21, N°7, pp. 715-729, Juin 1972.
- Sny80 SNYDER W.E., "Image modeling the continuity assumption and tracking", International Conference on Pattern Recognition, Vol. 1, Décembre 1980.
- Sny81 SNYDER W.E., "Computer analysis of time varying images ", *Computer*, Vol. 14, N°8, pp. 7-9, Août 1981.
- Sny83 SNYDER W.E.et RAJALA S.A., "Track Acquisition of Sub-Pixel Targets", Image Sequence Analysis, Ed. T.S.HUANG, p.503, 1983.
- Spo83 SPOER P., " Deplacement estimation for objects on moving background ", *Image sequence processing*, Ed. T.S. HUANG, 1983.
- Ste85b STELMASZYK P., BONNET P. et POSTAIRE J.G., "Analyse de scènes dynamiques par recherche des contours en mouvement", 5e Congrès AFCET/INRIA/ADI sur la Reconnaissanse des Formes et l'Intelligence Artificielle, Grenoble, 1985.
- Ste85 STELMASZYK P., "Analyse de scènes dynamiques par recherche des contours en mouvement. Application à la conduite automatique d'un tramway ", Thèse de Doctorat, Université des Sciences et Techniques de Lille, Novembre 1985.
- Stu80 STUCK F., "Réalisation d'un système adaptatif de traitement d'image pour l'identification et la localisation de pièces en robotique ", Thèse de 3e cycle, Toulouse, Décembre 1980.
- Tak82 TAKABA S. et al., "Measurement of Traffic Flow using Real-time Processing of Moving Pictures", Proceedings 32<sup>nd</sup> Conférence on Vehicular Technology, San Diego, CA, pp. 488-494, 1982.
- Tak83 TAKABA S. and KOISHI N., "Real-time Simulation of Road Traffic Flow", Summer Computer Conference, Vancouver, Canada, pp. 585-590, 1983.
- Tak84 TAKABA S. and OOYAMA, "Traffic flow measuring system with image sensors", Symposium on Road Research Program, OECD, Tokyo, Japan, Octobre 1984.
- Tan82 TANG I.S., "Extraction of Moving Objects in Textured Dynamic Scenes ", *IEEE Conf. Publ.*, Las Vegas, Juin 1982.
- The80 THEUWISSEN A., VITS A. et VERMEIREN J., "Analysis of traffic flow with a CCD-camera and a microprocessor", *seminar K, PTRC Summer Annual Meeting*, University of Warwick, pp. 283-287, 1980.

- Tho80 THOMPSON W.B., "Combining motion and contrast for segmentation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.2, n°6, 543-549, Janvier 1980.
- Tho81 THOMPSON W.B. et BARNARD S.T., "Lower-level estimation and interpretation of visual motion", *IEEE Computer*, Août 1981.
- Tou82 TOUSSAINT G.T., "On a convex hull algorithm for polygons and its application to triangulation problems ", *Pattern Recognition*, N°1, 1982.
- Tou87 TOUZANI A, "Classification automatique par détection des contours de modes des fonctions de densité, de probabilité multivariables et étiquetage probabiliste ", Thèse d'état, Université des Sciences et Techniques de Lille, Juin 1987.
- Tsa81 TSAI R.Y. et HUANG T.S., "Estimating 3-D motion parameters of a rigid planar patch", *IEEE Transactions on ASSP*, pp. 1147-1152, Décembre 1981.
- Tsu79 TSUGAWA S., YATABE T., HIROSE T. et WATANABE, "An automobile with artificial intelligence", 6<sup>th</sup> Int. Joint Conf. on Artificial Intelligence, 1979.
- Ull78 ULLMAN S., "Two dimensionality of the correspondence process in apparent motion", *Perception*, Vol. 7, pp.683-693, 1978.
- Ull79 ULLMAN S., "The interpretation of visual motion", Massachusetts Institut of Technology Press, Cambridge, 1979.
- Ull81 ULLMAN S., "The effect of similarity between line segments on the correspondence strenght in apparent motion", *Perception*, Vol. 9, pp. 617-622, 1981.
- Vie87 VIEREN C., STELMASZYK P., BONNET P. et POSTAIRE J.G., "Etude des possibilités offertes par le traitement d'images pour la détection d'obstacles devant un système de transport guidé ", Colloque Transports Guidés, Systèmes, Automatismes et Communications, GRPTIT/AFCET, Paris, Février 1987.
- Vie87b VIEREN C., STELMASZYK P., BONNET P. et POSTAIRE J.G., "Image processing techniques for security improvement in L.R.T. systems", *Int. Conf.* and Exhibition on Local Public Transport, Liverpool, Novembre 1987.
- Wal80b WALLACE T.P. et WINTZ P.A., "An efficient 3-D aircraft recognition algorithm using Fourier descriptors", Comp. Graph. Image Processing, Vol. 13, pp. 99-125, 1980.
- Wal80 WALLACE T.P. et MITCHELL O.R., "Analysis of 3-D movement using Fourier descriptors ", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-2, pp. 583-588, 1980.
- Wat84 WATERFALL R.C., DICKINSON K.W., "Image Processing Applied to Traffic: 2. Practical experience", *Traffic Engineering and Control*, Janvier 1984.
- Web81b WEBB J.A. et AGGARWAL J.K., "Visually interpreting the motion of objects in space ", *Computers*, Vol. 14, N°8, pp. 14-46, Août 81.



- Web81 WEBB J.A. et AGGARWAL J.K., "Structure from motion and jointed bodies", Proceedings of the 7th Int. Joint Conf. on Artificial Intelligence, Août 1981
- Wil80 WILLIAMS T.D., "Depth from camera motion in a real world scene ", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-2, Novembre 1980.
- Woh83 WOHN K., " Motion estimation based on multiple local contraints and nonlinear smoothing ", *Pat. Rec.*, Vol 16, N°6, 1983.
- Yac78 YACHIDA M., "Automatic motion analysing system of moving objects from the records of natural processes", *Proceedings of the International Joint Conference on Pattern Recognition*, Kyoto, Japan, pp.726-730, 1978.
- Yac83 YACHIDA M., "Determining velocity maps by spatio-temporal neighborhoods from image sequences", Computer Vision Graphics and Image Processing., Vol. 21, Février 1983.
- Yak78 YAKIMOVSKY Y. et CUNNINGHAM R. "A system for extracting threedimensional measurements from a stereo pair of TV cameras", *Computer Graphics and Image Processing*, Vol. 7, pp.195-210, Avril 1978.
- Yal80 YALAMANCHI S., "Difference Operations for the Segmentation of Moving Objects in Dynamic Scenes" 5th International Conference on Pattern Recognition, Miami, 1980.
- Yal82 YALAMANCHI S., "Extraction of Moving Object Descriptions via Differencing", Computer Graphics and Image Processing, Vol. 18, 1982.
- Yat78 YATABE T., HIROSE T., TSUGAWA S. et MATSUMOTO S., "Driving control method for automated vehicle with artificial intelligence", *Proceedings NECI* 1978-Industrial Aplications of Microprocesseurs, Mars 1978.
- Zah72 ZAHN C.T. et ROSKIES R.Z., "Fourier descriptors for plane closed curves," Visual Perception of Form, N.Y: Academic, pp. 206-211, 1972.

# RESUME

Dans cette thèse, on présente une méthode d'analyse de séquence d'images orientée vers l'interprétation de scènes dynamiques où évoluent des objets quelconques, sur fond non uniforme et sous éclairage non contrôlé.

L'approche proposée s'appuie sur une procédure de détection des contours en mouvement. En considérant les gradients des différences entre trois images successives prises deux à deux, on met en évidence les objets mobiles ainsi que les régions du fond découvertes ou recouvertes par ces objets pendant leurs mouvements. Les gradients de ces images différences sont ensuite combinées afin d'isoler uniquement les contours des objets en mouvement,même lorsque le fond de l'image n'est pas uniforme.

La procédure, implantée sur une machine de vision orientée temps réel a été testée sur des images réelles de trafic routier urbain. Les résultats obtenus montrent l'intérêt de l'approche proposée en vue de la surveillance et du contrôle de véhicules guidés en site non protégé. Ils mettent particulièrement en évidence la tobustesse des algorithmes qui, sans aucune information a priori sur le contenu des scènes permettent d'en extraire des modèles simples des objets mobiles.