

THESE

Présentée à l'Université des Sciences et Technologies de Lille

En vue de l'obtention du grade de

DOCTEUR DE L'UNIVERSITE

Spécialité

Instrumentation et Analyses Avancées

Par

Mohamed ZAYED



Véhicules Intelligents : Etude et développement d'un capteur intelligent de vision pour l'attelage virtuel

Soutenue publiquement le *12 juillet 2005*

Membres du jury :

M. P. MARCHE	Professeur, Ecole National Supérieure d'Ingénieurs de Bourges (Rapporteur).
M. J M. THIRIET	Maître de Conférence HDR, CRAN Université Henri Poincaré Nancy I (Rapporteur).
Mme M. BAYART	Professeur, LAGIS, USTL (Directrice de Thèse)
M. F. CABESTAING	Maître de Conférence HDR, LAGIS, USTL.
M. P. DESODT	Professeur, GIP, Ecole des Mines de Douai
M. J. BOONAERT	Maître de Conférence, GIP, Ecole des Mines de Douai (Co-directeur de Thèse)

Cette thèse a été préparée dans le cadre d'une collaboration entre le Laboratoire d'Automatique et d'Informatique Industrielle de Lille et le département du Génie Informatique et productive de l'Ecole des Mines de Douai.

Sommaire

INTRODUCTION GENERALE	5
1. CONTEXTE DE LA THÈSE	5
2. CADRE DE LA THÈSE.....	6
3. OBJECTIF DE LA THÈSE	11
4. PLAN DU MÉMOIRE	13
LES VEHICULES INTELLIGENTS : ETAT DE L'ART	14
I.1. INTRODUCTION	14
I.2. LA CONDUITE AUTOMATISÉE DES VÉHICULES.....	17
I.2.1. QUE FAUT-IL AMÉLIORER, LES VÉHICULES OU LES INFRASTRUCTURES ?.....	17
I.2.2. QUELLES SONT LES EXIGENCES QUE DOIT SATISFAIRE UNE APPLICATION ITS ?	18
I.2.3. QUEL TYPE DE CAPTEURS UTILISER ?.....	19
I.3. LA VISION DANS LE DOMAINE DES VÉHICULES INTELLIGENTS.....	20
I.4. LE SUIVI DE LA ROUTE.....	22
I.4.1. DÉTECTION DES MARQUAGES AU SOL	22
I.4.2. DÉTECTION DES OBSTACLES.....	25
I.5. LES TENDANCES RELATIVES A L'ÉVOLUTION DES MATÉRIELS.....	29
I.6. LES VÉHICULES PROTOTYPES.....	30
I.7. CONCLUSION.....	31
LES CAPTEURS INTELLIGENTS	33
II.1. INTRODUCTION	33
II.2. POURQUOI METTRE L'INTELLIGENCE AU SEIN DES CAPTEURS ?.....	34
II.2.1. L'ITINÉRAIRE DE L'INFORMATION.....	34
II.2.2. CE QU'ON ATTEND D'UN CAPTEUR INTELLIGENT	36
II.3.3. FONCTIONS ASSOCIÉES À « L'INTELLIGENCE » DES CAPTEURS	37
II.3.4. FONCTIONS DE L'INTELLIGENCE DANS LE TRAITEMENT DE L'INFORMATION.....	39
II.3. LES APPROCHES DE RÉALISATION DES CAPTEURS INTELLIGENTS.....	39
II.4. MODÉLISATION DU CONCEPT DU CAPTEUR INTELLIGENT :.....	40
II.5. LE CAPTEUR INTELLIGENT EN TANT QU'ÉQUIPEMENT INTELLIGENT	41
II.6. DOMAINES D'APPLICATIONS DES CAPTEURS INTELLIGENTS	42
II.7. LA CONCEPTION DES CAPTEURS INTELLIGENTS	43
II.7.1. APPROCHE FONCTIONNELLE	43
II.7.1.1. LA DÉCOMPOSITION FONCTIONNELLE.....	43
II.7.1.2. LA DÉCOMPOSITION PAR FLOTS DE DONNÉES.....	44
II.7.2. APPROCHE INFORMATIONNELLE	44
II.7.3. APPROCHE COMPORTEMENTALE	45
II.7.4. APPROCHE HYBRIDE	46
II.7.5. APPROCHE ORIENTÉE OBJET	46

II.8. LE CAPTEUR INTELLIGENT DÉDIÉ À L'ATTELAGE VIRTUEL	47
II.8.1. BESOINS SUPPLEMENTAIRES	48
II.8.2. SPÉCIFICATIONS DES BESOINS DU SYSTÈME D'ATTELAGE VIRTUEL	48
II.8.2.1. LES CONSOMMATEURS EXTERNES.....	49
II.8.2.2. LES SERVICES QUE DOIT FOURNIR LE SYSTÈME	50
II.8.3. ARCHITECTURE OPÉRATINNELLE DU SYSTÈME D'ATTELAGE VIRTUEL	63
II.9. CONCLUSION.....	65
APPLICATION : ATTELAGE VIRTUEL	66
III.1. INTRODUCTION	66
III.2. LES CONTRAINTES ET LES EXIGENCES DE L'APPLICATION.....	68
III.2.1. AU NIVEAU DE L'APPLICATION	68
III.2.2. AU NIVEAU MATÉRIEL	70
III.3. DÉCOMPOSITION FONCTIONNELLE DU CAPTEUR STÉRÉOSCOPIQUE DÉVELOPPÉ	70
III.3.1. ACQUISITION.....	72
III.3.1.1. LES DISPOSITIFS D'ACQUISITION D'IMAGES	73
III.3.1.2. LA TAILLE DU CAPTEUR.....	77
III.3.1.3. LE MODÈLE GÉOMÉTRIQUE DE LA CAMÉRA	77
III.3.1.4. CHAMPS DE VISION – VOLUME DE TRAVAIL	86
III.3.2. LE SYSTÈME DE STÉRÉOVISION.....	89
III.3.2.1. UNE CONFIGURATION GÉOMETRIQUE PARTICULIERE	89
III.3.2.2. DÉSCRIPTION DU SYSTÈME DE STÉRÉOVISION MIS EN PLACE.....	90
III.3.3. LA CALIBRATION	93
III.3.3.1. LA CALIBRATION PHOTOGRAMÉTRIQUE.....	95
III.3.3.2. LA CALIBRATION HYBRIDE	97
III.3.3.3. L'AUTOCALIBRATION (SELF-CALIBRATION).....	101
III.3.3.4. LA CALIBRATION D'UN CAPTEUR STÉRÉOSCOPIQUE	108
III.3.4. MISE EN ŒUVRE DE LA CALIBRATION HYBRIDE	110
III.3.4.1. BINARISATION DE L'IMAGE.....	112
III.3.4.2. DÉTECTION DES RECTANGLES ET EXTRACTION DES COINS DE LA MIRE	113
III.3.4.3. ORGANISATION DES COINS DÉTECTÉS.....	114
III.3.4.4. EVALUATION DES PERFORMANCES DE LA CALIBRATION DU SYSTÈME STÉRÉOSCOPIQUE.....	116
III.3.5. DÉTECTION DE LA PLAQUE D'IMMATRICULATION.....	128
III.3.5.1. LA SEGMENTATION STATISTIQUE	129
III.3.4.2. RÉSULTATS EXPERIMENTAUX.....	138
III.3.6. LECTURE DE PLAQUE D'IMMATRICULATION.....	140
III.3.6.1. SEGMENTATION DES CARACTÈRES	141
III.3.6.2. RECONNAISSANCE DES CARACTÈRES.....	146
III.3.6.3. EVALUATION DES PERFORMANCES DE LA PROCÉDURE DE LECTURE DE LA PLAQUE.....	155

III.3.7. L'ALGORITHME DE TRACKING.....	156
III.3.7.1.INTERET DE LA DEMARCHE	156
III.3.7.2.LA DÉMARCHE GLOBALE.....	156
III.3.7.3.MISE EN ŒUVRE DU FILTRE.....	159
III.3.8. EXTRACTION DES INFORMATIONS NÉCESSAIRES AU SUIVI DU VÉHICULE	160
III.3.8.1.LE MODE NORMAL DU FONCTIONNEMENT DU CAPTEUR STÉRÉOSCOPIQUE...	161
III.3.5.2.LE MODE DU FONCTIONNEMENT DÉGRADÉ DU CAPTEUR STÉRÉOSCOPIQUE.	174
III.4.CONCLUSION	186
CONCLUSION GÉNÉRALE	188
RÉFÉRENCES BIBLIOGRAPHIQUES.....	192
ANNEXE A.....	205
A. L'objectif des caméras.....	205
A.1. Paramètres caractéristiques d'un objectif.....	205
A.1.1. LA FOCALÉ :	205
A.1.2. LE DIAPHRAGME :	205
A.1.3. LE FOCUS :	206
A.2. La Qualité d'un objectif.....	206
A.2.1. LA QUALITÉ DU MATÉRIAU.....	206
A.2.2. LA RÉPONSE SPECTRALE.....	206
A.2.3. LA DISTORSION.....	206
ANNEXE B.....	207
B. Le transfert des images	207
B.1. Le transfert numérique.....	207
B.1.1. LA NORME RS – 422 OU EIA – 422	207
B.1.2. LA NORME RS – 644 OU EIA – 644	208
ANNEXE C.....	209
C Le modèle de la distorsion des lentilles	209
ANNEXE D.....	211
D.1. Le détecteur de Harris.....	211
ANNEXE E.....	213
E.1. Le calcul de la matrice fondamentale	213
ANNEXE F.....	215
F.1. Détail de l'extraction des paramètres intrinsèques de la caméra à partir de la matrice fondamentale ..	215
F.2 Extraction des équations de Kruppa à partir de la transformation épipolaire	216
ANNEXE G.....	221
G.1. Dimensions des plaques.....	221

INTRODUCTION GENERALE

1. CONTEXTE DE LA THESE

Si de nombreux aspects de notre vie sont devenus plus agréables grâce à l'utilisation de technologies avancées, il a fallu longtemps au secteur du transport pour combler son retard en la matière. Aujourd'hui, ces progrès sont devenus incontournables. Les véhicules deviennent de plus en plus sûrs, mais en contrepartie de plus en plus complexes; il est donc nécessaire d'avoir recours à l'automatisation pour prendre en charge les fonctions évoluées maintenant disponibles. A cela se rajoute un trafic plus intense et un environnement plus dense. Il faut pouvoir compter sur des systèmes rapides et robustes pour prolonger la perception humaine. Ces technologies, qui visent à accroître la sécurité, le rendement et la commodité de la conduite, sont actuellement en cours de conception et de développement. Ces systèmes doivent fonctionner en étroite collaboration avec le conducteur. Le concepteur d'un véhicule intelligent doit intégrer des technologies et des systèmes disparates pour créer une machine cohérente qui complète efficacement le conducteur humain.

Ainsi, le pari des chercheurs et des constructeurs automobiles est le développement d'un « véhicule intelligent », utilisant les avancées récentes en électronique et en informatique. Le véhicule intelligent doit pouvoir :

- Suivre la route,
- Se tenir dans la file correcte,
- Détecter les piétons,
- Gérer les inter-distances avec les autres véhicules,
- Adapter sa vitesse en fonction des conditions du trafic et des caractéristiques de la route,
- Changer la file pour suivre les véhicules et éviter les obstacles,
- Trouver l'itinéraire le plus court dans l'environnement urbain,
- Echanger les informations avec l'infrastructure et les autres véhicules,
- Etc !

Si le terme « véhicule intelligent » paraît encore un peu présomptueux, il traduit cependant la part déjà très importante mais toujours croissante des systèmes embarqués dans les véhicules d'aujourd'hui, puisque ceux-ci représentent environ 30% du coût total du véhicule [Har.04].

Certains des systèmes directement issus de travaux de laboratoire sont désormais disponibles sur les véhicules actuels. Parmi les plus célèbres, nous pouvons citer l'injection électronique, les airbag, l'ABS (Anti Bloking System), le régulateur de vitesse ou encore le système d'aide à la vision nocturne. D'autres dispositifs plus récents sont en cours de tests et concernent le concept des véhicules intelligents pour la ville appelés communément « Cybercars ».

Parallèlement aux efforts consentis pour améliorer l'agrément et la sécurité des véhicules, les recherches se concentrent sur les systèmes visant à accroître le « rendement » et les conditions de circulation propres aux infrastructures déjà en place. Ainsi, de nombreuses équipes de recherche travaillent sur le concept de train de véhicules, emprunté au monde ferroviaire, et basé essentiellement sur la notion « d'attelage virtuel ». Les buts poursuivis sont, entre autres, l'optimisation de la circulation routière et l'amélioration de la sécurité sur les autoroutes et les réseaux secondaires. Les retombées économiques potentielles sont nombreuses, que ce soit au niveau des constructeurs automobiles qu'au niveau des nouveaux services que l'utilisation des technologies associées aux véhicules intelligents permet de proposer aux usagers de la route.

Ceci explique en partie pourquoi l'amélioration ou l'optimisation de la circulation routière sont des problématiques qui ont été étudiées depuis longtemps. La possibilité de créer désormais des systèmes d'aide à la conduite accessibles au plus grand nombre ne fait qu'accroître l'intérêt que celles-ci suscitent auprès de nombreux chercheurs et de grands constructeurs automobiles.

2. CADRE DE LA THESE

Le travail présenté dans cette thèse est le résultat de la collaboration entre le département Génie Informatique et Productique de l'Ecole des Mines de Douai (EMD – GIP) et le Laboratoire d'Automatique et Génie Informatique et Signal (LAGIS). Il s'inscrit dans le cadre du programme AViVA (« Attelage Virtuel pour Véhicules Autonomes) du projet AutoRIS (« Automatique pour la Route Intelligente et Sûre ») mené par le Groupement de Recherche en Automatisation Intégrée et Systèmes Hommes-Machines (GRAISyHM). Ce projet, soutenu par 10 Laboratoires et organismes de recherche de la Région Nord pas de Calais (EMD – GIP, ERASM, ESTAS (INRETS), GEMTEX (ENSAIT), LASL (ULCO), LABOGP (FSA Bethune), LAGIS (EC Lille et USTL), LAMIH (UVHC), LEOST (INRETS)), concerne le développement des concepts de routes et de véhicules intelligents afin d'optimiser la circulation routière, améliorer la sécurité, préserver l'environnement et proposer de nouveaux services aux usagers.

Il s'agit principalement de concevoir et de développer des systèmes embarqués pour l'aide à la conduite et / ou l'automatisation de certaines tâches qui peuvent améliorer les conditions de circulation. Pour l'aide à la conduite, l'objectif est d'assister le conducteur dans des situations difficiles (détection d'obstacles, vision nocturne, régulation de vitesse, etc...)

ou encore d'automatiser la conduite pour effectuer des tâches spécifiques (dont l'attelage virtuel est un exemple). D'autres entités de recherche visent à développer des systèmes d'aide à la navigation. L'objectif de l'aide à la navigation est de proposer aux usagers des transports intelligents (véhicules touristiques, bus, etc...) une nouvelle gamme de services tels que les informations en temps réel sur les itinéraires, des informations touristiques, des services de réservations de chambres d'hôtels ou encore tout service susceptible d'intéresser le conducteur au cours de son déplacement.

Le travail présenté ici concerne l'aide à la conduite. Nous nous sommes intéressés plus particulièrement à la perception de l'environnement à l'avant du véhicule afin d'extraire les informations nécessaires à la mise en œuvre d'un attelage virtuel pour véhicules autonomes. Cette approche vise à remplacer les liaisons matérielles caractérisant les attelages mécaniques (dont un exemple est fourni dans la figure ci-dessous) par des chaînes informationnelles : le but est d'améliorer à la fois la sécurité et la flexibilité des véhicules en convoi.

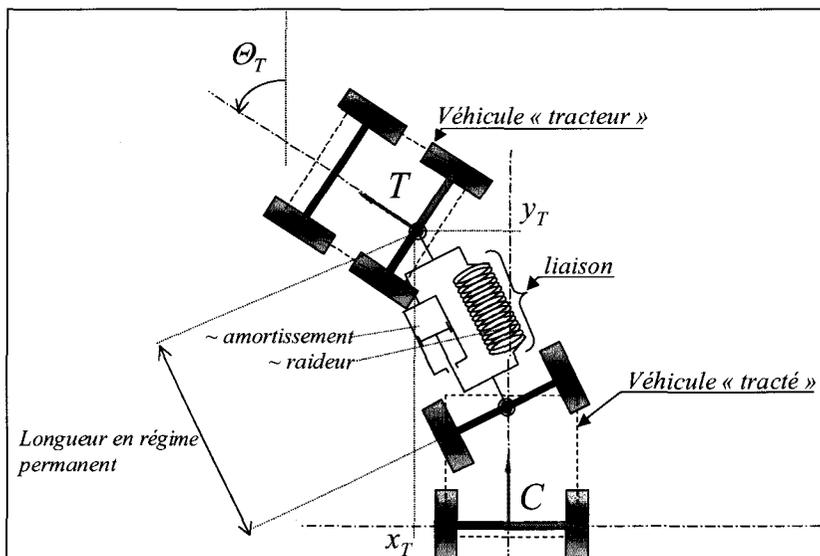


Figure 1 : Exemple de liaison « physique » simple.

L'équivalent virtuel de cette liaison doit donc exhiber des propriétés fonctionnelles équivalentes (la liste ci-dessous n'évoquant que celles qui nous intéressent le plus directement) :

- Assurer une certaine « interdistance » entre véhicules.
- Amener le véhicule tracté sur une trajectoire compatible avec un suivi correct du « tracteur ». Par « correct », on entend une trajectoire qui satisfasse les contraintes liées à la sécurité des personnes transportées et des autres usagers de la route, des matériels utilisés et des infrastructures.
- Pouvoir spécifier le comportement dynamique de la liaison (par le biais par exemple de paramètres similaires à l'amortissement et à la raideur). Ces paramètres pouvant être adaptés « en ligne » en fonction des conditions opératoires : conduite sur autoroute, conduite en ville, etc.

- Pouvoir adapter la distance intervéhiculaire en fonction de ces mêmes conditions opératoires. Il est en effet clair qu'en fonction de la vitesse absolue des véhicules impliqués dans la liaison, les distances de sécurité doivent être augmentées.

La mise en œuvre d'une telle liaison virtuelle suppose donc que le véhicule « tracté » soit capable par lui-même d'apprécier la « position » relative du véhicule « tracteur » (en fait la position relative en tant que telle, mais aussi l'orientation et la vitesse relatives, aussi bien linéaire qu'angulaire). L'entité responsable (le « pilote ») de la conduite du véhicule tracté doit donc, à partir des grandeurs mesurables associées aux éléments de l'attelage, extraire des informations exploitables et pertinentes, capables de le renseigner sur cette configuration relative. A partir de celles-ci, compte tenu de l'objectif assigné (distance intervéhiculaire principalement) et des paramètres de fonctionnement courant (« amortissement », erreur admise, etc.) le pilote agit sur les organes permettant la manœuvre du véhicule, à savoir :

- Le volant (angle de braquage).
- L'accélérateur.
- La commande de frein.

L'interaction entre les différents éléments impliqués dans la mise en œuvre d'une liaison virtuelle telle que nous venons de la décrire est représentée sur la figure suivante (figure 2).

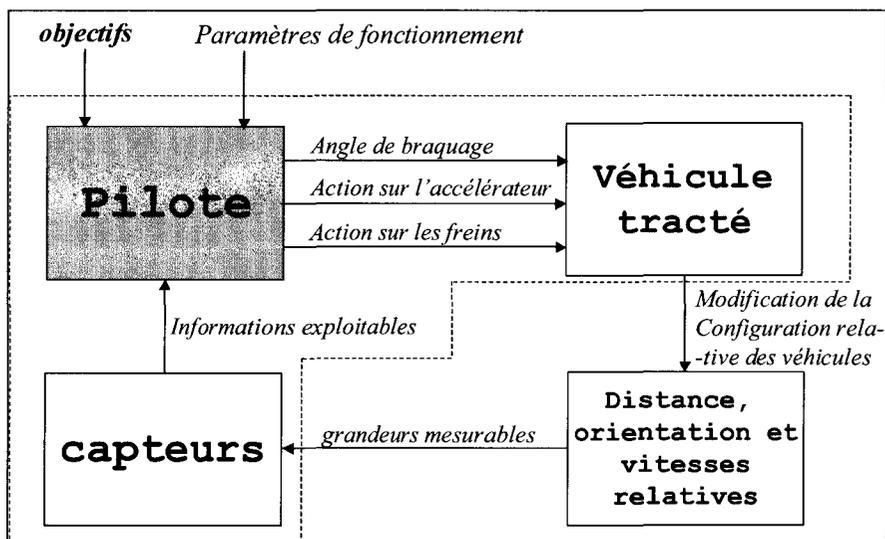


Figure 2 : Interactions entre les constituants de la liaison virtuelle.

On voit que ce « pilote » se trouve à l'intersection entre l'aspect « perception de l'environnement » et l'aspect « commande » du véhicule. Il est donc clair que la conception d'une liaison virtuelle mobilise les disciplines connexes à ces deux domaines. Ainsi, comme le montre la figure 3, l'extraction des informations pertinentes à partir des mesures recueillies sur le système sera du ressort de la partie « perception », tandis que l'élaboration d'une loi de commande s'appuyant sur ces données et tenant compte de l'objectif et des paramètres de fonctionnement, est dévolue à l'aspect « commande » du problème.

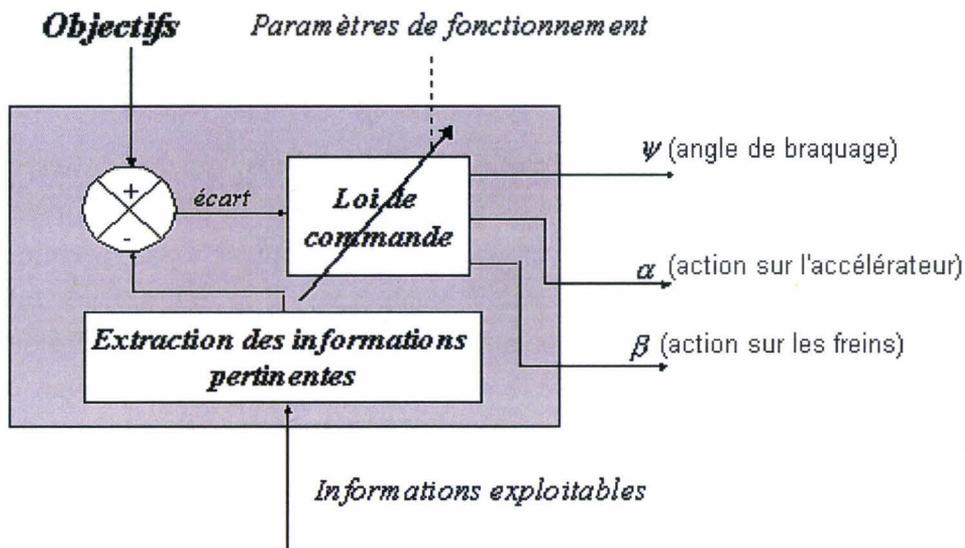


Figure 3 : Organisation interne du « pilote ».

Le travail exposé dans le présent mémoire est relatif à la partie « perception de l'environnement ». Il s'agit pour nous de définir et de mettre au point un capteur permettant de fournir les données nécessaires au contrôleur pour piloter le véhicule « tracté ».

Plusieurs approches peuvent alors être adoptées, qui différeront essentiellement par la technologie des capteurs utilisés :

- *Technologie « par émission-réception »* : Elle consiste à utiliser des capteurs mettant en œuvre une technique basée sur la réflexion d'une onde émise par le capteur lui-même. On peut ranger dans cette catégorie des dispositifs tels que les radars, sonars et lidars.
- *Technologie « par réception simple »* : Il s'agit de capteurs exploitant un signal présent de façon naturelle et « spontanée » dans l'environnement. Les caméras vidéos sont typiques de ce type de capteurs.

La technologie par émission-réception n'implique en général pas de traitements complexes, grâce à la sélectivité que procure les techniques employées, et les mesures obtenues sont directes. Cependant, dès lors que des performances élevées sont nécessaires, cette technologie peut s'avérer chère et présente dans tous les cas l'inconvénient majeur d'interférences entre capteurs du même type, dès lors qu'ils sont utilisés en grand nombre sur une surface réduite (par exemple par des véhicules sur une route).

A contrario, les technologies « par réception simple » nécessitent des traitements complexes afin d'extraire l'information pertinente du « bruit » et ses performances sont limitées dans certaines situations (par exemple dans le brouillard ou l'obscurité pour ce qui concerne les capteurs d'images). En dépit de ces difficultés, cette technologie présente deux avantages déterminants pour les transports intelligents.

Le premier d'entre eux est le coût, qui reste faible comparativement à la technologie par émission-réception, ceci en partie grâce aux récentes avancées techniques touchant les dispositifs impliqués dans ce premier type de capteur.

Le second est directement lié à la nature de ces capteurs, qui les préserve de façon intrinsèque des interférences entre dispositifs identiques. C'est en particulier pour ces deux raisons que nous avons opté pour le développement d'un capteur « par réception simple » répondant à des contraintes de coûts strictes. Compte-tenu de la richesse des informations qu'il est à même de procurer, celui-ci a pris la forme d'un capteur de vision stéréoscopique, se composant de deux caméras et d'une unité de traitement compact et économique, que ce soit financièrement et énergétiquement.

La stéréovision est une technique qui utilise deux caméras ou plus. Cette technique nécessite des traitements complexes et sensible aux mouvements du véhicule et aux erreurs de calibration. D'une manière générale, il est clair que toutes les techniques basées sur la vision sont sensibles à certaines conditions météorologiques et d'éclairage (nuit, brouillard...). Néanmoins, la stéréovision présente des avantages très intéressants (certains de ceux cités ci-dessous ne lui étant pas spécifiques):

- Une quantité d'informations à la fois très importante et très riche.
- La possibilité de pouvoir s'appliquer d'une façon générique (s'applique à tout objet).
- La reconstruction 3D sans requérir d'informations très précises sur la scène observée.
- Le Suivi des véhicules dans les courbes, là où le radar échoue.
- Une grande flexibilité, dans la mesure où un capteur de stéréovision permet l'implémentation des deux autres techniques en parallèle (ce qui représente une propriété intrinsèque très intéressante).
- La nature non intrusive de ce capteur (capteur passif), lui épargne les problèmes d'interférences.
- Le coût relativement « modique » d'un système de stéréovision basé sur des composants standards le rend attractif en comparaison de capteurs actifs dotés de performances adaptées à l'application (radars, lidars, etc.)

Ces avantages font du capteur de stéréovision un outil de choix pour la perception de l'environnement du véhicule et justifie son emploi dans le cadre de notre projet. Malgré tout, il est clair que son utilisation dans le domaine des transports n'est pas sans poser certains problèmes.

L'objectif principal du capteur stéréoscopique au sein de notre application est d'extraire l'information utile au contrôle du véhicule (vitesse et distance relatives, orientation, etc) à partir du couple d'images pris au même instant de deux points de vue différents.

Connaissant le modèle de projection de chacune des deux caméras, il s'agit de calculer les coordonnées 3D (par triangulation passive dans notre cas) d'un ensemble de points appartenant à la scène observée à partir de leurs projections sur les deux images. Pour y parvenir, nous devons faire face aux problèmes suivants :

- Estimer le modèle de projection de chacune des deux caméras, c'est à dire calibrer le capteur stéréoscopique.
- Résoudre le problème de la mise en correspondance. Ceci revient à savoir associer un pixel dans une image donnée à sa contrepartie dans l'autre image, correspondant à la projection du même point de la scène. Ce processus doit être effectué en respectant les contraintes temps réel et de fiabilité propre à l'application envisagée. Une technique de mise en correspondance efficace et rapide est présentée dans [Zay.03a].
- Effectuer la reconstruction tridimensionnelle. Les problèmes potentiels sont principalement liés à la calibration du capteur stéréoscopique et à la triangulation.

Ne dérogeant pas aux autres techniques mettant en œuvre une « réception simple », l'exploitation de la stéréovision dans le cadre des transports routiers nécessite la mise en place d'un certain nombre d'algorithmes de traitement et d'analyse d'images fortement « couplés » entre eux, devant être à la fois robustes et répondre à des contraintes temporelles strictes (surtout pour des application telles que le suivi automatique de véhicules). C'est en particulier l'objectif de cette thèse que d'aboutir à un système efficace et cohérent permettant de respecter de telles contraintes de performances et de coût, ceci grâce à l'emploi de méthodes adaptées aux conditions imposées par un usage routier et par l'exploitation de matériel compatible avec le prix de la voiture « de monsieur tout-le-monde »....

3. OBJECTIF DE LA THESE

La technique fondamentale mise en œuvre dans notre capteur, (à savoir la stéréovision) est un processus qui, à partir de deux ou plusieurs images d'une même scène prises de positions différentes, permet de retrouver la structure 3D de la scène observée. La reconstruction 3D d'une scène est assurée par l'exploitation conjointe des propriétés colorimétriques de la scène et des caractéristiques géométriques des caméras.

Dans son principe, l'attelage virtuel repose sur la capacité du capteur développé à fournir des informations concernant la position et l'orientation du véhicule précédant celui auquel il est lié. Si on désigne par « tracteur » le véhicule auquel on souhaite s'atteler, le capteur doit être capable d'exploiter une (voire des) entité(s) inhérente(s) au véhicule tracteur, qui permettrait au besoin de le singulariser parmi des véhicules strictement identiques par

ailleurs. Le choix de cette entité est une problématique en soi, puisque celle-ci doit exhiber plusieurs propriétés, à savoir :

- ✓ Cette entité doit être présente sur tous les véhicules tracteurs potentiels de façon naturelle (ou plutôt dès la sortie de l'usine), puisque nous ne voulons pas modifier l'apparence de ceux-ci.
- ✓ De façon (presque) antagoniste à la propriété précédente, l'entité choisie doit présenter des variations de sorte à pouvoir singulariser un véhicule précis de façon univoque.
- ✓ Le véhicule suiveur se trouvant nécessairement derrière le véhicule tracteur, l'entité choisie doit être obligatoirement à l'arrière du véhicule tracteur.
- ✓ Elle doit être « facile » à distinguer de son environnement et doit présenter des propriétés géométriques similaires pour tous les véhicules, afin que les algorithmes développés soient « génériques » à tous les véhicules tracteurs potentiels.



Figure 4 : Exemple de scène observée par le système de localisation.

Les propriétés énoncées ci-dessus permettent de dresser une partie des spécifications de notre capteur stéréoscopique, compte tenu de la nécessaire adéquation devant exister entre l'entité à détecter et le couple algorithme / architecture afin de respecter nos contraintes de temps de réponse et de robustesse. Les autres spécifications du capteur (dimensionnelles, optiques, etc..) découlent directement des conditions d'exploitation de celui-ci.

Dans cette thèse, nous proposons donc un capteur stéréoscopique intelligent - de part sa structure et l'unité de traitement qui lui y associée - qui permet l'extraction des informations nécessaires à la tâche d'attelage virtuel. Sa disposition géométrique particulière lui permet également de proposer un fonctionnement adéquat dans le cas de défaillance d'une seule caméra. Les lignes directrices de cette thèse étaient de proposer une solution matérielle bon marché, flexible et embarquable (compacité, disponibilité, basse consommation électrique, solidité, etc...), mettant en œuvre une solution logicielle modulaire, robuste et rapide, accessible à tous et permettant une évolution continue en toute simplicité.

4. PLAN DU MEMOIRE

L'organisation retenue pour ce mémoire est la suivante : après une présentation détaillée de la problématique de l'attelage virtuel et la justification des choix effectués en matière de capteurs, le chapitre 1 dresse l'état de l'art des véhicules intelligents puis présente les approches les plus communément adoptées dans le cadre de cette thématique, en insistant plus particulièrement sur la « conduite automatique de véhicule ». Il passe en revue les solutions expérimentales les plus prometteuses et les prototypes développés à travers le monde. Les exemples présentés ont pour point commun de s'appuyer sur la vision par ordinateur afin de percevoir l'environnement du véhicule. Bien entendu, le panorama présenté n'a pas prétention à l'exhaustivité car seules n'ont été retenues ici les approches jugées les plus intéressantes du point de vue de notre problématique.

Dans le chapitre 2, nous introduisons la notion de « capteurs intelligents » et présentons les différentes définitions à la base de ce concept. Le caractère « utile » du capteur intelligent passe nécessairement par l'analyse de l'acheminement de l'information, raison pour laquelle une attention particulière est accordée à ce point. Nous présentons ensuite les différentes architectures de capteurs intelligents et le rôle de « l'intelligence » dans ces derniers. Nous abordons également les différentes méthodes permettant d'intégrer cette « intelligence » au capteur. Enfin, nous passerons en revue les approches de conception des capteurs intelligents que nous mettrons en application pour identifier les différents services et fonctionnalités que doit approprier le capteur stéréoscopique intelligent pour réaliser la tâche de l'attelage virtuel.

C'est le chapitre 3 qui constitue le « cœur » de ce travail, puisqu'il s'intéresse à la réalisation et à la mise en œuvre de capteur stéréoscopique intelligent destiné à permettre la mise en œuvre d'un attelage virtuel. L'une des particularités de ce travail est de se baser uniquement sur les caractéristiques visuelles d'un véhicule banal, sans qu'il soit nécessaire d'instrumenter le véhicule suivi (le « tracteur »). Une démarche bien particulière sera détaillée tout au long de ce chapitre. Les résultats obtenus montrent la faisabilité et l'efficacité d'une telle approche.

Nous détaillerons les différentes phases de l'extraction des informations nécessaires à l'attelage virtuel, depuis l'acquisition des images stéréo en passant par la calibration des caméras, la détection de l'entité d'intérêt, la « certification » de celle-ci, le suivi 2D et 3D et pour finir l'extraction des informations de position et d'orientation intervéhiculaire. Nous présentons par ailleurs dans ce chapitre une technique permettant le suivi du véhicule tracteur avec une seule caméra dans le but de pallier les défaillances que peut subir notre système au cours de son fonctionnement.

LES VEHICULES INTELLIGENTS :

ETAT DE L'ART

A ctuellement, l'automatisation de la conduite automobile est sujette à un grand intérêt. En effet, ses potentialités sont nombreuses et séduisantes, qu'il s'agisse de l'optimisation des infrastructures, de l'amélioration de la mobilité, de la minimisation des risques et du temps de parcours et, par voie de conséquence, de la consommation d'énergie.

Ce chapitre présente les approches les plus communément utilisées dans le cadre de la thématique « véhicules intelligents », et plus spécifiquement dans le contexte de la « conduite automatique de véhicule ». Il passe en revue les solutions expérimentales les plus prometteuses et les prototypes développés à travers le monde. Ceux-ci se basent essentiellement sur des techniques d'Intelligence Artificielle et sur la Vision par ordinateur afin de percevoir l'environnement du véhicule.

N'ont été retenues ici que les approches jugées les plus intéressantes du point de vue de notre problématique. La fin du chapitre est l'occasion d'évoquer les perspectives d'évolution pour les décennies à venir de ce domaine en constante mutation que constitue les « véhicules intelligents ».

I.1. INTRODUCTION

A u cours des dernières décennies une grande importance a été donnée aux systèmes de transport afin d'améliorer les conditions de sécurité, d'optimiser l'exploitation des réseaux routiers, de diminuer la consommation d'énergie et de préserver l'environnement.

Les efforts consentis pour résoudre ces problèmes ont ouvert un nouveau domaine de recherche et d'application, à savoir, *la conduite automatisée de véhicule*, dans lequel des nouvelles techniques sont exploitées pour automatiser entièrement ou partiellement les tâches

de la conduite.

Les différentes tâches en question sont reprises dans la liste ci-dessous (qui n'a pas prétention à l'exhaustivité) :

- Suivre la route,
- Se tenir dans la file correcte,
- Se maintenir une distance sûre avec les autres véhicules,
- Détection des piétons,
- Adapter la vitesse du véhicule en fonction des conditions de trafic et des caractéristiques de la route,
- Changer de file pour suivre les véhicules et éviter les obstacles,
- Trouver l'itinéraire le plus court vers une destination donnée,
- Se déplacer et stationner dans l'environnement urbain.

L'intérêt pour les technologies des systèmes de transport intelligents (ITS) est apparu depuis une vingtaine d'années, quand le problème de la mobilité des personnes et des marchandises, commençait à se poser avec une plus grande acuité, favorisant ainsi la recherche de solutions alternatives.

Des institutions gouvernementales ont activé cette phase initiale de prospection au moyen de divers projets, impliquant un grand nombre d'unités de recherche qui ont travaillé dans une voie coopérative afin de produire plusieurs prototypes et de mettre en œuvre les solutions pressenties, toutes basées sur des approches assez différentes les unes des autres.

En Europe, le projet PROMETHEUS¹ a commencé cette étape exploratoire en 1986. Le projet a impliqué plus de 13 constructeurs automobiles et plusieurs unités de recherches gouvernementales et universitaires de 19 pays européens. Dans ce cadre, un certain nombre d'approches différentes des ITS² ont été conçues et mises en application.

Aux Etats-Unis la majeure partie des initiatives a été lancée afin de traiter le problème de la mobilité, impliquant plusieurs universités, centres de recherches et compagnies automobiles. Après cette phase pilote, le gouvernement américain a établi le NAHSC³ [Bis.97] en 1995.

¹ PROgram for an European Traffic with Highest Efficiency and Unprecedented Safety

² Intelligent Transportation Systems, comme indiqué précédemment.

³ NAHSC: National Automated Highway System Consortium

Au Japon, où le problème de mobilité est beaucoup plus important, quelques prototypes de véhicules ont été développés dans le cadre de différents projets. A l'instar de ce qui s'est produit aux USA, l'Advanced Cruise-Assist Highway System Research Association (AHSRA) a été fondée en 1996 de manière à impliquer un grand nombre d'industriels de l'automobile et de centres de recherches [Tok.97], qui ont conjointement développé différentes approches relatives à la conduite automatisée de véhicule.

Nous pouvons également citer le projet coopératif international ATLANTIC⁴ qui a débuté en 2001, impliquant plusieurs laboratoires de recherches européens et Nord Américains. Ce projet se décompose en 8 thématiques⁵ de recherches distinctes sur les systèmes de transports intelligents [Gin.04].

Les résultats principaux de cette première étape se résument d'une part à une analyse profonde du problème et d'autre part au développement d'une étude de faisabilité, indispensable à la compréhension des exigences et des implications de ce qu'il est convenu d'appeler dorénavant la « technologie ITS ».

Le domaine de l'ITS entame maintenant sa deuxième phase, qu'on peut caractériser par une maturité dans ses approches et par l'avènement de nouvelles possibilités technologiques qui permettent le développement des premiers produits expérimentaux. Un certain nombre de prototypes de véhicules intelligents, ont été conçus, mis en oeuvre et testés sur route. La conception de ces prototypes résulte aussi bien de l'analyse critique des résultats issus directement de la recherche sous-jacente que de l'apport d'autres disciplines, parmi lesquelles la Robotique, l'Intelligence Artificielle, l'Informatique, les Télécommunications, l'Automatique, le Traitement du Signal, pour ne citer que celles-ci.

Initialement, les dispositifs technologiques les plus adaptés au problème de la perception dans le cadre du problème posé (caméras infrarouges, radars, sonars, etc.) étaient dérivés d'applications militaires très coûteuses. Cependant, grâce à l'effet conjugué de l'engouement croissant pour les applications ITS et du progrès de la production industrielle, la technologie d'aujourd'hui offre des capteurs et des systèmes de traitement à des prix très compétitifs.

Pour évaluer et comparer les différentes approches, ces prototypes de véhicules automatiques sont équipés d'un grand nombre de capteurs et de calculateurs.

L'Europe, de son côté, finance depuis de nombreuses années des projets sur l'électronique embarquée et les véhicules intelligents. Aujourd'hui, au-delà des développements industriels potentiels pour l'aide à la conduite ou la conduite automatisée, la vocation de ces projets est clairement de fédérer les recherches et les développements des différents acteurs européens. Parmi les différents projets en cours, Carsense [Lan.01] réunit 12

⁴ ATLANTIC: A Thematic Long-term Approach to Networking for the Telematics and ITS Community

⁵ <http://www.atlan-tic.net/>

partenaires⁶ depuis janvier 2000 : Des constructeurs automobiles, des équipementiers et des instituts de recherche. L'objectif de ce projet est de proposer un système d'aide à la conduite à faible vitesse (de l'ordre de 30 km/h) aussi complet que possible, adapté à des situations complexes comme la conduite en ville. Le concept clé de ce système repose sur les performances de l'ensemble des capteurs équipant le véhicule : radars, lasers, caméras, capteurs à ultrasons, etc.. . Le but est d'en extraire suffisamment d'informations pour bien évaluer l'environnement du véhicule et être capable de prendre une décision tout en respectant les contraintes temporelles imposées par ce type d'application. Deux autres projets majeurs, financés par la commission européenne ont débuté en 2001 pour une durée de trois ans : CyberCars⁷ [Par.03] lancé en août 2001 et CyberMove⁸ [Ram.04] lancé en décembre 2001. Les deux projets réunissent treize partenaires⁹. Il convient par ailleurs de citer le projet ARCOS¹⁰ [Arc.03] qui a débuté au mois de janvier 2002.

Comme nous l'avons précisé au début de ce mémoire, dans le cadre de notre projet de recherche sur les véhicules intelligents et les ITS, nous nous intéressons plus particulièrement aux systèmes basés sur la vision par ordinateur. Avant toute chose, il est nécessaire de revenir de façon plus détaillée sur les motivations qui sont à la base du développement de tels dispositifs, en précisant quelles sont leurs particularités et les contraintes qui s'y appliquent. Une attention toute particulière sera consacrée aux approches les plus représentatives ou au moins les plus « applicables », au problème de l'attelage virtuel tel que nous l'avons introduit. Pour terminer, nous tenterons de tracer les grandes lignes des perspectives d'évolution des véhicules intelligents.

1.2. LA CONDUITE AUTOMATISEE DES VEHICULES

Avant d'adopter une stratégie d'automatisation de la conduite, il est légitime d'explorer les différentes façons d'aborder le problème. C'est le but des sections qui suivent.

1.2.1. QUE FAUT-IL AMELIORER, LES VEHICULES OU LES INFRASTRUCTURES ?

L'automatisation de la conduite automobile peut se baser sur des modifications de la structure routière ou (de façon non exclusive) sur des modifications des véhicules eux-mêmes. Selon les spécificités de l'application, chacune de ces deux alternatives possède des avantages

⁶ **Partenaires du projet Carsense** : Autocruise, BMW, CRF, Thales, INRETS-LEOST, Renault, IBEO, Jena-Optronik, INRIA, LIVIC, TRW Automotive, ENSMP.

⁷ Cybernetic technologies for the car in the city.

⁸ Cybernetic transportation systems for the cities.

⁹ **Partenaires du projet CyberCars** : INRIA (France), RUF (Danemark), FIAT research (Italie), Université de Bristol (GB), FROG Navigation Systems (Pays-bas), Yamaha (Japon), Robosoft (France), Technion (Israël), TNO (Pays-bas), Université de Southampton (GB), Serpentine (Suisse), Université de Coimbra (Portugal), Autos et Energies (Suisse).

Partenaires du projet CyberMove : Les 13 partenaires cités ci-dessus ainsi que GEA (Suisse).

et des inconvénients. L'amélioration des infrastructures routières peut apporter des bénéfices aux types de transport basés sur des itinéraires pré-définis, comme c'est le cas pour les transports en commun et la robotique industrielle (pour exemple la « transitique » basée sur un filoguidage ou un dispositif assimilable). Cependant, cette approche exige une organisation et un entretien complexes et étendus qui peuvent devenir encombrants et extrêmement chers dans le cas des réseaux routiers de grandes dimensions. Une restructuration de l'environnement ne peut être raisonnablement appliquée qu'à un sous-ensemble réduit du réseau routier (par exemple une route entièrement automatisée que seuls des véhicules automatiques, publics ou privés, pourraient emprunter). Pour cette raison, les systèmes qui se réaliseront à court terme doivent pouvoir tirer profit des infrastructures déjà existantes, ce qui autorise de substantielles économies en matière d'infrastructure en comparaison de la première solution évoquée. Par conséquent, les travaux des chercheurs s'orientent plus essentiellement sur l'amélioration des véhicules.

I.2.2. QUELLES SONT LES EXIGENCES QUE DOIT SATISFAIRE UNE APPLICATION ITS ?

Comme tous les autres, les systèmes de transport intelligents disposent de certaines spécificités qui sont directement liées à leur domaine d'application. A ces spécificités s'associent des contraintes particulières.

Ainsi, pour [Ber.00], n'importe quel système embarqué pour une application ITS doit satisfaire certaines exigences vitales :

- Le système final, doit être assez robuste pour s'adapter aux différents états et changements d'environnement, de route, de trafic, d'éclairage, ainsi qu'à des conditions météorologiques variables.
- Les systèmes embarqués propres aux applications ITS sont des systèmes de sûreté critique pour lesquels un haut degré de fiabilité est exigé. Par conséquent, le projet doit être approfondi et rigoureux pendant toutes ses phases, de la spécification des exigences à la conception et la mise en oeuvre. Une phase intensive de mise à l'épreuve et de validation est donc d'une importance capitale.
- Afin de rendre la « technologie ITS » accessible au plus grand nombre, et par conséquent rendre cette dernière attractive pour les industriels du secteur, la conception d'un équipement ITS doit intégrer dès le départ des contraintes de coût stricts¹¹. Les dépenses de tout ordre (comme la consommation électrique) doivent être tenues aussi basses que possible, puisque les performances du véhicule (qu'elles soient techniques ou économiques) ne doivent pas être affectées négativement par l'utilisation du système ITS.

¹¹ Le coût du système embarqué ne doit pas dépasser 10 % du prix de véhicule.

- Le matériel et les capteurs du système doivent être compacts et discrets. Ils ne doivent par ailleurs pas « détonner » vis-à-vis du style du véhicule.
- Enfin, puisque les systèmes ITS doivent être déclenchés et contrôlés par un opérateur humain, ils ont besoin d'une interface homme-machine conviviale, simple, réactive et efficace.

I.2.3 QUEL TYPE DE CAPTEURS UTILISER ?

L'ensemble des capteurs utilisés dans le cadre de la Robotique Mobile « d'intérieur » (Indoor) ne s'applique pas avec le même bonheur lorsqu'il s'agit d'équiper un système ITS, destiné à affronter le monde extérieur dans des conditions très variables. Ainsi, les capteurs tactiles et les capteurs acoustiques sont inappropriés pour les applications impliquant des véhicules à moteur évoluant à vitesse élevée [Ber.00], du fait de la proximité que nécessitent les premiers et de la sensibilité au mouvement relatif à l'air qui affecte les seconds.

Les capteurs à base de laser et les radars exploitent la réflexion d'un signal émis par le dispositif sur un obstacle. La mesure elle-même repose principalement sur la détermination d'un temps de vol ou par la caractérisation d'une « onde de battement » (issu du produit entre signal émis et signal réfléchi). Etant à l'origine du signal permettant la mesure, « nous les classifions en tant que capteur par émission-réception ». Leurs principaux inconvénients sont la basse résolution spatiale (en général) et leur temps d'acquisition relativement important.

Les capteurs basés sur la vision, n'étant pas à l'origine du signal permettant la mesure, sont *a contrario* définis en tant que capteurs « par émission simple » et ont un avantage intrinsèque par rapport à leurs homologues par émission réception qui est (par essence) la faculté d'acquérir les données sans modifier l'environnement (capteur non intrusif). Malheureusement, de par leur nature même, les capteurs de vision sont davantage perturbés que les radars par le brouillard, le manque ou l'excès de lumière, etc. En somme, même si les effets peuvent être exacerbés ou, au contraire atténués, les capteurs de vision sont sensibles aux mêmes perturbations que la vision humaine elle-même.

En outre, les capteurs par émission-réception possèdent quelques particularités qui leurs donnent des avantages sur la vision dans certaines applications spécifiques : ils effectuent certaines mesures (comme celle du mouvement) d'une façon plus directe que la Vision en exigeant moins de ressources en calcul car devant acquérir une quantité considérablement inférieure de données pour parvenir à l'estimation voulue.

En dépit de ces indéniables qualités, le problème principal des capteurs par émission-réception est l'interférence entre capteurs du même type, qui peut devenir critique lorsqu'un grand nombre de véhicules se déplacent simultanément dans le même environnement (comme dans le cas de véhicules automatiques roulant sur des autoroutes intelligentes.).

Il est clair qu'un tel problème ne se pose pas dans le cadre de l'utilisation de capteurs par réception simple, tels que des caméras. Ceci constitue donc un atout majeur pour l'application envisagée.

Le tableau 1 regroupe de façon synthétique les forces et faiblesses de chacune de ces deux familles de capteur :

Remarque : Evidemment, la vision n'augmente pas nécessairement les capacités de détection au-delà des possibilités humaines (en présence d'un important brouillard ou la nuit en l'absence d'éclairage spécifique, par exemple), mais elle peut par contre aider le conducteur en cas de défaillance (manque de concentration, fatigue, etc.)

Les capteurs	Avantages	Inconvénients
Par Emission – Réception (dits ici « Actifs » ¹²)	<ul style="list-style-type: none"> - Mesures directes, - Exigent comparativement moins de ressources de calcul, - Robustesse vis-à-vis de certaines conditions de fonctionnement dégradées. 	<ul style="list-style-type: none"> - Interférences entre les capteurs du même type.
Par Réception simple (dits ici « Passifs » ¹²)	<ul style="list-style-type: none"> - Très peu (voire pas du tout) intrusifs. 	<ul style="list-style-type: none"> - Par nature, sensibles aux conditions d'acquisition.

Tableau I.1 : Comparaison entre capteurs « actifs »¹² et capteurs « passifs »¹².

La nature non intrusive des capteurs passifs en général et la richesse que procure la vision en particulier fait de cette dernière une approche incontournable dans le domaine des véhicules intelligents.

I.3. LA VISION DANS LE DOMAINE DES VEHICULES INTELLIGENTS

Des questions importantes doivent être soigneusement traitées durant la phase de conception des systèmes de vision dédiés aux applications automobiles. Les paragraphes qui suivent traitent de ces dernières :

¹² G. Hash, entre autres, définit un capteur actif par sa faculté à produire une forme d'énergie (en général sous forme électrique) lorsqu'il est excité par le signal à mesurer, tandis qu'un capteur passif modifie certaines de ses caractéristiques (par exemple sa résistance) sous l'influence de ce signal – Ici, le terme « actif » désigne des capteurs devant « exciter » leur environnement pour produire un signal exploitable- alors que le terme « passifs » se rapporte à des capteurs utilisant un signal « spontanément » disponible.

- En premier lieu, les systèmes ITS sont plus exigeants que d'autres applications en terme des contraintes temporelles (temps-réel). En effet, le temps de réponse du système se doit d'être en accord avec la vitesse de déplacement du véhicule, qui peut être élevée. De façon antagoniste, les applications de Vision nécessitent le traitement d'un flux important de données. Le défi que posent les techniques de Vision artificielle dans le contexte de ce type d'application est donc d'être capable de traiter une grande quantité de données de façon fiable et en temps-réel. Des architectures et des techniques spécifiques à la satisfaction de cette double contrainte sont, par conséquent, un domaine de Recherche très actif [Cab.99, Har.04, Iss.04, Lab.04, Per.05].
- Deuxièmement, dans le domaine de l'automobile, aucune supposition ne peut être faite sur les paramètres principaux (comme par exemple l'éclairage de la scène observée ou le contraste des tons) influant directement sur les informations prises en compte par le capteur de vision. Par conséquent, le traitement ultérieur doit être assez robuste pour s'adapter à des conditions environnementales différentes (comme le soleil, la pluie, le brouillard) et à leurs changements s'effectuant avec des dynamiques très variables, tels que les transitions entre le soleil et l'ombre ou l'entrée ou la sortie d'un tunnel.
- Pour finir, d'autres problèmes-clés, tels que la tolérance aux mouvements propres du véhicule (par exemple roulis, tangage, etc.) et les dérives dans la calibration des caméras, doivent aussi être traités.

Les avancées récentes en informatique et dans les technologies des capteurs ont favorisé l'utilisation de la vision également dans le domaine des véhicules intelligents. Les progrès du matériel informatique, en particulier au niveau du degré d'intégration et de la diminution de la tension d'alimentation, permettent de produire des machines de vision plus puissantes, moins dispendieuses en énergie (et donc plus facilement « embarquables »), à des prix accessibles.

De nombreux autres facteurs¹³ doivent être pris en compte lors de la conception et du choix d'un dispositif d'acquisition d'images. Les choix peuvent se porter sur de la Vision monoculaire (une seule caméra), ou binoculaire (stéréovision), voire « multicaulaire » (approche multi – caméras). D'autres paramètres importants sont liés à l'angle de prise de vue, qui se doit d'être adapté à l'application envisagée (surveillance globale des abords du véhicule ou accostage, par exemple), de même que la résolution et la profondeur¹⁴ des images. Les contraintes fixées doivent par ailleurs permettre de faire un choix entre l'exploitation d'images en couleurs ou en niveaux de gris.

¹³ Si nombreux que l'énumération ci-dessous est nécessairement incomplète !

¹⁴ **Profondeur** signifie le nombre de bits par pixel

Au-delà des considérations sur la structure du capteur et la nature des données retenues, d'autres caractéristiques, intrinsèques au capteur, doivent être prises en considération, tels que le temps d'intégration et la dynamique du capteur.

Conclusion : Bien que fortement exigeante et extrêmement complexe à analyser ; la richesse des informations fournies par la Vision (dont il a été évalué qu'elle constitue 90% des données exploitées par un conducteur dans sa tâche de pilotage) en font un des moyens les plus puissants pour percevoir l'environnement et a, de ce fait, été largement utilisée pour traiter un grand nombre de problèmes dans le domaine de la conduite automatique, [Cou.02, Ste.02, Lab.02].

I.4. LE SUIVI DE LA ROUTE

Le *suiti automatique de la route* est sans doute l'une des approches ayant suscité le plus de travaux ces dernières décennies. Sa mise en œuvre repose sur deux modules complémentaires assurant les tâches suivantes :

- La détection des lignes séparant les différentes files de la chaussée, la modélisation de la route, la détermination de la position relative du véhicule par rapport à la voie et l'analyse de la direction des autres véhicules.
- La détection et la localisation des obstacles se trouvant éventuellement sur la trajectoire actuelle du véhicule.

En effet, la détection et la localisation des obstacles est une partie intégrante de la tâche du suivi automatique de la route. De ce fait, ces deux thématiques de recherche dans le domaine des transports intelligents se sont forgés une destinée commune. Dans ce qui suit, nous évoquons plus en détails la détection des marquages au sol et la détection des obstacles.

I.4.1. DETECTION DES MARQUAGES AU SOL

Dans la plupart des prototypes de véhicules intelligents développés à travers le monde, le suivi de la route est divisé en deux étapes, du type « perception » et « action » : initialement, la position relative du véhicule est calculée par rapport à la voie, et ensuite les actionneurs le maintiennent dans la configuration correcte. Quelques exemples des stratégies qui peuvent être adoptées pour résoudre le problème du contrôle latéral sont présentés dans [Tay.99].

Cependant, certains des premiers systèmes n'étaient pas basés sur la détection préliminaire de la position du véhicule par rapport à la route, mais déterminaient directement les commandes des actionneurs (steering wheel angles) à partir des images acquises. Ainsi,

dans le cadre du projet ALVINN¹⁵ [Joc.93a], le système est basé sur une approche par réseaux de neurones : il est capable de suivre la route après une phase d'analyse d'un grand nombre d'images. Quoiqu'il en soit, la position des marquages sur la route étant utile à d'autres sous-systèmes de dispositifs comparables, la localisation de ceux-ci est de toute façon généralement exécutée.

Bien entendu, les systèmes tels que ceux précédemment cités ne sont pleinement efficaces que lorsque les files sont matérialisées par des lignes au sol. D'autres réalisations, telles que le SCARF [Cri.91] et PVR¹⁶ III [Kim.95], ont été conçues pour exploiter des routes peu structurées. Ces systèmes sont basés sur l'utilisation de caméras couleur et exploitent l'hypothèse qu'une route est colorée de façon homogène afin de détecter celle-ci dans les séquences d'images.

Dans plusieurs cas, la détection de la route est réduite à la détection de certaines structures spécifiques à cette dernière. Ces méthodes simplifient la détection, mais doivent néanmoins faire face à deux problèmes majeurs qui sont :

- *La présence des ombres*, qui produit de « pseudo-objets » sur la surface de la route et en modifie localement la texture. La plupart des groupes de recherche font face à ce problème en utilisant des algorithmes de filtrage d'images très sophistiqués généralement appliqués à des images en niveau de gris. Dans de plus rares cas, des images en couleur sont utilisées. Le véhicule intelligent MOSFET¹⁷ [Mic.97] utilise par exemple un algorithme de segmentation couleur maximisant le contraste entre le marquage des lignes et la route.
- *Les marquages sur la route occultés par d'autres véhicules*. Pour faire face à ce problème, certains systèmes traitent seulement une région de la route où l'absence d'autres véhicules peut être raisonnablement supposée. Ainsi, les véhicules intelligents LAKE et SAVE traitent la partie d'image correspondant aux 12 mètres les plus proches devant le véhicule. Des travaux tels que [Cod.97] ont montré que cette approche permet de manœuvrer le véhicule sur les autoroutes sans risque. S'appuyant sur une technique tout autre, le système RALPH¹⁸ [Pom.96] réduit la partie de l'image traitée grâce aux résultats d'un module de détection d'obstacle à base de radar. Le groupe de recherche du Laboratoire Régional Des Ponts et Chaussées de Strasbourg exploite l'hypothèse qu'il doit toujours y avoir un contraste chromatique entre la route et son environnement (ou les obstacles), au moins dans

¹⁵ ALVINN: Autonomous Land Vehicle In a Neural Net

¹⁶ SCARF: Supervised Classification Applied to Road Following

PVR: POSTECH Vehicle Road

¹⁷ MOSFET: Michigan Off-road Sensor Fusing Experimental Testbed.

¹⁸ RALPH: Rapidly Adapting Lateral Position Handler

une composante de couleur. En s'appuyant sur cette hypothèse, le concept de saturation chromatique [Cha.98] est utilisé pour séparer les composantes.

En raison des contraintes physiques et de continuité, le traitement de l'image entière peut être réduit à l'analyse d'une *région d'intérêt*. C'est la stratégie généralement adoptée en exploitant des connaissances *a priori* sur l'environnement de la route. Par exemple, le système développé par le groupe de recherche de Robert Bosch GmbH [Gol.98] utilise un modèle dynamique de la route et du véhicule afin de déterminer la partie de la route où il est susceptible de trouver les marquages au sol.

L'hypothèse réaliste d'une largeur fixe de la route (en première approximation) permet de perfectionner le critère de recherche, en limitant cette dernière aux marquages presque parallèles. Le véhicule PVR III [Kim.95] utilise les réseaux de neurones et des algorithmes de vision afin d'accomplir cette tâche. De façon analogue, le système RALPH [Pom.96] traite la partie de l'image correspondant à 20 – 70 m devant le véhicule. L'effet de perspective est contrebalancé par calcul sur cette zone de l'image, la courbure de la route est ensuite estimée en comparant cette zone à un certain nombre de modèles comportant des marquages parallèles.

La reconstruction géométrique de la route est simplifiée par des suppositions concernant sa forme. Les groupes de recherche de l'Universitat der Bundeswehr [Lüt.98] et Daimler-Benz [Fra.98] détectent la route en se basant sur un modèle spécifique: les marquages au sol sont modélisés sous la forme de *clothoïdes*¹⁹. L'avantage d'une telle modélisation est que la connaissance de deux paramètres est suffisante pour définir complètement la localisation du marquage au sol et permet de déterminer d'autres grandeurs telles que la compensation latérale et l'angle de direction. D'autres groupes de recherche utilisent une représentation polynomiale des marquages au sol. Ainsi, le véhicule intelligent MOSFET [Mic.97] exploite pour ce faire des paraboles, une transformation simplifiée de Hough étant ensuite employée pour accomplir la procédure d'adaptation.

D'autres systèmes s'appuient quant à eux sur un modèle générique de la route. Par exemple, le système ROMA [Tok.97] utilise une approche de la Vision à base de contours. Un modèle dynamique permet le traitement de petites régions en temps réel. Les contours formés par des pixels qui ont une valeur de direction de gradient significative font l'objet d'une procédure de « *tracking* » (poursuite). D'une manière analogue, un modèle triangulaire générique de la route a été développé pour le véhicule expérimental MOB-LAB par les groupes de recherche de Di Parma Università [Bro.95] et Istituto ElctroTecnico Nazionale [Den.94] "G.Ferraris".

La connaissance *a priori* des paramètres de calibration du système de vision et certaines hypothèses sur l'état de la route (route plate en particulier) simplifient la mise en

¹⁹ Dans une *clothoïde* la courbure dépend linéairement de la référence curviligne.

correspondance entre les points des images d'une part et les points des images et leurs coordonnées dans un repère lié à la scène observée d'autre part.

Remarque: Pour déterminer la distance avec un obstacle ou encore la courbure de la route, la grande partie des systèmes présentés précédemment supposent que la route devant le véhicule est plate.

Le système GOLD²⁰ [Ber.98] implémenté sur le véhicule intelligent ARGO, ainsi que le système RALPH, exploitent l'hypothèse que la route est localement plane. La corrélation temporelle entre les images successives est généralement utilisée pour valider le résultat.

Hypothèse	Avantages	Inconvénients
Région d'intérêt	- Analyse temps-réel,	- Le choix de la région d'intérêt est critique.
Largeur de la route fixe	- Amélioration des critères de recherche, - Robustesse vis-à-vis des phénomènes de type « ombres ».	- N'est pas adaptée à des routes dont la largeur est variable.
Forme de la route	- Robustesse par rapport aux ombres, - Simplification de la phase de reconstruction de la géométrie de la route, - Contrôle simplifié.	- Nécessité d'une grande puissance de calcul, - Nécessite une bonne adéquation route réelle / modèle.
Les connaissances a priori de la route surface / inclinaison	- Simplification de la mise en correspondance entre les pixels des images et les coordonnées réelles (détermination des distances avec les obstacles et de la courbure de la route).	- Les hypothèses (par exemple, route plate) ne sont en général satisfaites qu'en première approximation, - la recalibration du système stéréoscopique est souvent nécessaire.

Tableau I.1 : Avantages et inconvénients des hypothèses classiques faites dans le cadre de la détection de routes.

I.4.2. DETECTION DES OBSTACLES

Certains systèmes limitent la détection des obstacles à la localisation des autres véhicules, ils s'appuient alors principalement sur la recherche de modèles spécifiques dans l'image, procédé qui peut être utilement renforcé par des considérations de forme et de symétrie des objets recherchés (les véhicules).

²⁰ GOLD: Generic Obstacle and Lane Detection.

L'algorithme de détection d'obstacles développé à l'Université der Bundeswehr [Lüt.98] se base sur la détection des contours et la modélisation des obstacles ; le système est capable de détecter et « traquer » jusqu'à 12 objets autour du véhicule. Les variables continuellement mises à jour et associées aux obstacles sont : la distance, la vitesse relative, l'accélération relative, la position latérale, la vitesse latérale et la taille.

Dans l'exemple cité précédemment, la détection des obstacles exploite des modèles spécifiques, de ce fait, le traitement d'une seule image suffit, mais l'approche manque d'efficacité dans le cas où l'obstacle ne correspondrait pas suffisamment aux modèles.

Une définition plus générale, qui compliquera évidemment les algorithmes, considère que tout objet se trouvant sur le chemin du véhicule est un obstacle. Autrement dit, toute déformation de la route correspondra à un obstacle potentiel. Dans ce cas, la détection des obstacles est réduite à l'identification de l'espace libre devant le véhicule :

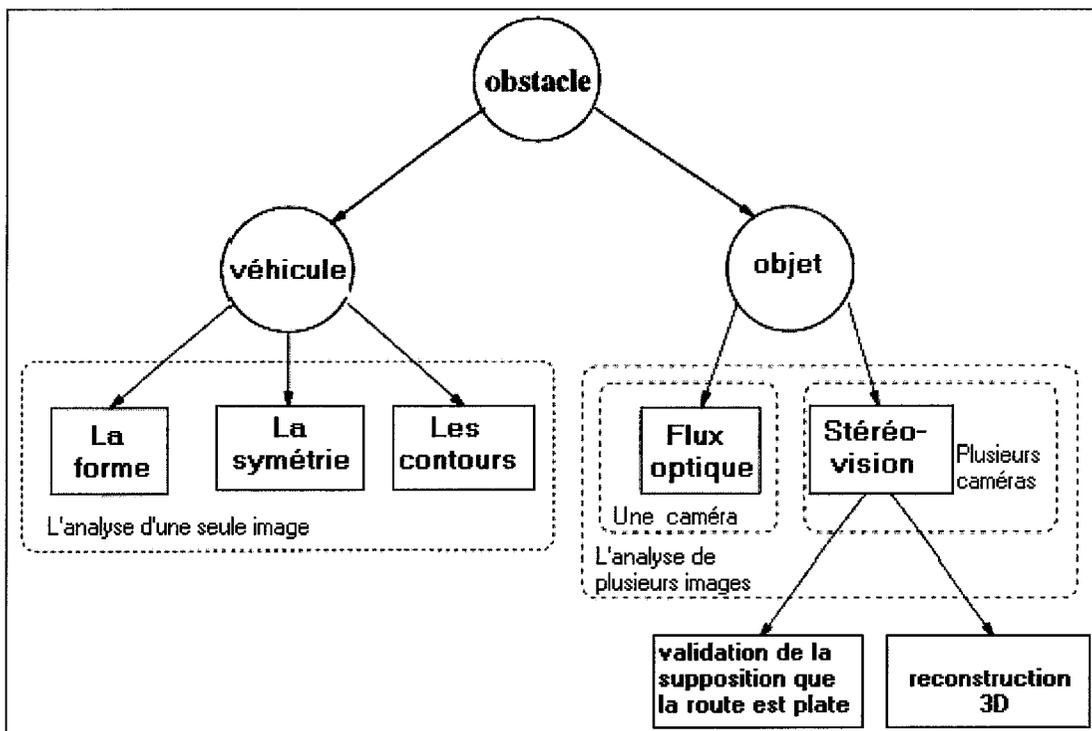


Figure I.1 : Les techniques de détection utilisées pour les différentes définitions des obstacles considérées[Ber.00].

L'adoption de cette définition exige l'utilisation de techniques de détection plus complexes ; les plus connues sont :

- L'analyse de type **flux optique**.
- Le traitement d'images **non monoculaires**.

Dans le premier cas plusieurs images sont acquises par le même capteur à des instants différents. Dans le deuxième cas, plusieurs caméras acquièrent des images simultanément de

différents points de vue. En plus de leur complexité intrinsèque causée par la quantité de données à traiter, ces techniques doivent être suffisamment robustes vis-à-vis des imprécisions, incertitudes et indéterminations causées par le mouvement du véhicule et les dérives dans la phase de calibration des caméras.

La technique du *flux optique* exige l'analyse d'une séquence de deux images au moins. Un vecteur 2D des mouvements horizontaux et verticaux de chaque pixel est alors calculé dans l'espace image. Le résultat est utilisé pour estimer le mouvement. Certains systèmes extraient le déplacement directement de l'odomètre ; les obstacles sont détectés en analysant la différence entre les domaines de mouvements prédits et réels. Le système ROMA [Kru.95] intègre un module de détection d'obstacles qui est basé sur la technique du flux optique et sur l'analyse des données fournies par un odomètre. De la même façon, ASSET-2 ²¹ est un système complètement temps réel pour la segmentation et la détection des objets qui se déplacent de façon autonome. Sa particularité essentielle réside dans le fait qu'il n'exige pas la calibration de la caméra. Il traque les objets et il est capable de traiter correctement les occultations des obstacles. Chaque nouvel objet « automoteur » qui entre dans la scène initie automatiquement une nouvelle procédure de « *tracking* ». Initialement, ASSET-2 construit un domaine de flux clairsemé et segmente ensuite l'image obtenue en régions présentant des variations de flux homogènes. La corrélation temporelle [Smi.95] est utilisée pour filtrer le résultat, dont on améliore ainsi l'exactitude.

En ce qui concerne le traitement des images non-monoculaires, ce dernier exige la mise en correspondance des pixels des différentes images : deux images pour la stéréovision binoculaire et trois dans le cas de la stéréovision trinoculaire. L'avantage d'analyser des stéréos images (deux ou trois images) au lieu d'une séquence d'images monoculaires se résume dans la possibilité de détecter directement la présence d'obstacles (par exemple grâce à l'analyse des disparités), ce qui n'est pas possible pour la vision monoculaire. De plus, dans des conditions limites où le véhicule et les obstacles ont des vitesses faibles ou nulles, l'approche à base de flux optique ne détecte pas les obstacles, tandis que la stéréovision, elle, n'est pas prise en défaut.

Le projet UTA²² du groupe de recherche Daimler-Benz visait à développer un système intelligent (Stop & go) à l'intérieur de la ville en adoptant la stéréovision pour extraire l'information 3D en temps réel. Le prototype développé [Fra.98] est par ailleurs capable de reconnaître les signes du trafic, les balises de signalisation, les piétons, les passages protégés et les lignes d'arrêt.

Le groupe de recherche de MIT²³, a développé un système de stéréovision pour la détection tridimensionnelle des routes et la visualisation du trafic, qui peut tout aussi bien être

²¹ ASSET-2: A Scene Segmentation Establishing Tracking v2

²² UTA: Urban Traffic Assistant.

²³ MIT: Massachusetts Institute of Technology.

adapté à d'autres applications. Le système est capable de séparer les véhicules partiellement chevauchés et de les distinguer depuis leurs ombres.

En outre, pour diminuer la complexité intrinsèque de la stéréovision, des contraintes spécifiques au domaine sont généralement adoptées.

Dans le système GOLD, la suppression de l'effet de perspective dans les images stéréo [Bro.99a], permet d'obtenir deux images qui ne peuvent différer l'une de l'autre qu'en présence d'un ou plusieurs obstacles ou si l'hypothèse de planéité de la route n'est pas respectée. Dans le cas contraire, il s'agit alors d'un système détectant l'espace libre devant le véhicule.

D'une façon similaire, l'unité de recherche de l'Université de Californie [Kol.95], a développé un algorithme qui rectifie l'image gauche en utilisant le point de vue de l'image droite, détectant ainsi les obstacles à partir de l'analyse des disparités. Dans cette optique, un filtre de Kalman est utilisé pour suivre les obstacles.

Des algorithmes de détection d'obstacles plus récents ont été développés dernièrement. Nous pouvons citer à titre d'exemple l'algorithme de détection d'obstacles avec un stéréoscope composé de deux caméras matricielles dans le cadre du projet ARCOS [Lab.04] ou encore la technique de détection d'obstacles rapide basée sur un stéréoscope composé de deux caméras linéaire (voir [Har.04]), ces techniques ont le mérite d'exploiter uniquement les propriétés géométriques de la projection de la scène sur les caméras pour pouvoir détecter un obstacle.

Le tableau I.3 compare les points forts et les points faibles des différentes approches relatives au problème de la détection d'obstacles.

Les approches adoptées	Avantages	Inconvénients
L'analyse d'une seule image	<ul style="list-style-type: none"> - des algorithmes simples, - traitement rapide, - pas d'influence du mouvement du véhicule. 	<ul style="list-style-type: none"> - perte d'informations concernant la profondeur des scènes si des suppositions spécifiques ne sont pas faites, - n'aboutit pas si le modèle ne correspond pas.
Flux optique	<ul style="list-style-type: none"> - détection générique des objets, - permet le calcul du déplacement et de la vitesse relative des obstacles. 	<ul style="list-style-type: none"> - traitements complexes, - sensible au mouvement du véhicule et aux erreurs de calibrage, - n'aboutit pas si le véhicule ou l'obstacle a une vitesse nulle.

La stéréovision	<ul style="list-style-type: none"> - détection générique des objets, - permet la reconstruction 3D. 	<ul style="list-style-type: none"> - traitements complexes (des contraintes spécifiques peuvent réduire la complexité), - sensible au mouvement du véhicule et aux erreurs de calibrage.
------------------------	---	--

Tableau I.3 : Comparaison entre les différentes approches de détection d'obstacles.

Comme mentionné ci-dessus, différentes techniques ont été proposées dans la littérature et évaluées sur des prototypes afin de résoudre le problème du suivi de route, mais seulement quelques-unes d'entre elles fournissent une solution intégrée, (comme par exemple la détection des routes et la détection d'obstacle) justifiant la poursuite de recherches destinées à l'amélioration des résultats et à la mise au point des traitements plus rapides et plus efficaces.

I.5. LES TENDANCES RELATIVES A L'EVOLUTION DES MATERIELS

Dans les premières années des applications ITS un grand nombre de solutions ont été proposées et beaucoup d'entre elles étaient basées sur des matériels spécialisés. Ce choix était justifié par le fait que le matériel disponible sur le marché à un prix raisonnable n'était pas assez puissant pour pouvoir traiter les images en temps réel. Par exemple, les chercheurs de l'Universität Der Bundeswehr [Dic.95, Joc.93b, Joc.98, Bro.97] ont développé leur propre architecture du système.

En plus du choix des capteurs appropriés et du développement des algorithmes spécifiques, un grand pourcentage de cette première étape de recherche était consacré à la conception, la mise en oeuvre et aux essais des nouvelles plates-formes matérielles. En fait, quand une nouvelle architecture d'ordinateur fait son apparition, il faut prendre en considération non seulement le matériel et les aspects architecturaux (tel que le jeu d'instructions, les intercommunications, les Entrées / Sortie, etc.) mais aussi les solutions logicielles applicables.

Ces dernières années, l'évolution technologique a tout changé: presque tous les groupes de recherche utilisent des composants directement disponibles sur le marché pour leurs systèmes.

Par exemple, le NavLab 5 (véhicule du Carnegie Mellon University) et ARGO (le véhicule de Di Parma Università) ont utilisé des systèmes basés sur des processeurs universels. Grâce à la disponibilité des réseaux ultra-rapides, quelques solutions MIMD²⁴ ont

²⁴ MIMD: Multiple Instruction Multiple Data.

même été explorées. Ces dernières sont composées d'un petit nombre de processeurs performants et interconnectés, comme dans le cas du véhicule VaMoRs-P [Lüt.98] de l'Universität der Bundeswehr où l'ancien système a été remplacé par trois PCs (dual Pentium II) connectés *via* un réseau rapide basé sur Ethernet.

Les tendances actuelles [Ber.00], cependant, semblent converger vers une architecture hybride, dans laquelle le processeur universel puissant est « épaulé » par du matériel spécifique comme par exemple des cartes et des circuits intégrés mettant en oeuvre le calcul de flux optique, l'appariement des formes, la convolution et des filtres morphologiques (liste non-exhaustive).

De plus, quelques capacités SIMD²⁵ sont maintenant transférées dans le jeu d'instructions des dernières générations de CPUs, ceci afin de faciliter la mise en oeuvre d'une forme de parallélisme intrinsèque au traitement de données visuelles et audio.

En conclusion, *il est important de souligner que bien que les systèmes de nouvelle génération soient tous basés sur du matériel du commerce, le développement de matériel personnalisé n'a pas perdu sa signification. Au contraire, cette démarche bénéficie d'un regain d'intérêt pour ce qui touche à la conception de systèmes embarqués. Une fois que les prototypes matériels et logiciels ont été conçus, réalisés et largement testés, leurs fonctionnalités doivent ensuite être intégrées dans un système embarqué optimisé avant sa commercialisation. C'est dans cette étape du projet que le développement du matériel personnalisé joue toujours un rôle fondamental, justifiant les dépenses qu'il entraîne durant la phase de développement par les économies qu'il assure lors de la mise en oeuvre.*

I.6. LES VEHICULES PROTOTYPES

Plusieurs groupes de recherche sur les applications ITS ont intégré leurs approches et solutions les plus prometteuses pour la conduite automatisée dans des véhicules prototypes. Plusieurs de ces résultats expérimentaux ont été présentés et éprouvés pendant des événements internationaux importants tels que la réunion finale du projet PROMETHEUS à Paris (1994), la démonstration de NAHSC à San Diego, CA (août 1997), et Automated Vehicle Guidance Demo'98 à Rijnwoude, au Pays-Bas (juin 1998), ainsi que Demo'99, OH (juillet 1999) et dernièrement la démonstration du véhicule prototype de l'école des mines de Paris [Ste.02], et du véhicule prototype du projet européen ARCOS [Lab.02], à Versailles, dans le cadre du Symposium des Véhicules Intelligents (juin 2002).

D'une façon générale, les véhicules prototypes ont subi plusieurs essais en environnements « contrôlés », tels que des pistes fermées. Mais, seulement dans certains cas, des essais étendus ont été effectués sur la voie publique. On peut citer à cet égard :

²⁵ SIMD: Single Instruction Multiple Data.

- Le prototype VaMP, qui a été testé sur l'itinéraire de Munich (Allemagne) vers Odense (Danemark) [Mau.96] en 1995,
- Le système RALPH, a été testé sur NavLab 5 par un voyage (*No Hands Accross America* [Smi.95]) de Pittsburgh, PA, vers San Diego, CA en 1995.
- Le véhicule ARGO, conduit par le système GOLD pendant presque 2000 km sur l'ensemble de l'Italie pendant le *MilleMiglia in Automatico Tour* [Bro.99a] en 1998.

Ces tests ont permis d'expérimenter les véhicules dans différentes conditions et ils ont mis en évidence leurs points forts et leurs points faibles. Les résultats encourageants obtenus par tous les groupes témoignent de la maturité atteinte par la recherche dans ce domaine (voir, [Bro.99b] pour plus de détails).

I.7. CONCLUSION

Des conclusions communes peuvent être tirées des trois expériences décrites : toutes ont réussi à atteindre un pourcentage de conduite automatisée assez élevé. Des expériences plus récentes ont profité des dernières avancées technologiques et ont utilisé uniquement du matériel du commerce.

Les principaux problèmes rencontrés pendant les démonstrations, étaient dus aux réflexions de la lumière et aux conditions imparfaites pour l'acquisition d'image (par exemple le fait que la route soit mouillée, les brutales variations d'illumination lors de l'entrée ou la sortie dans les tunnels, la présence des ombres des ponts ou autres infrastructures routières, etc.).

Les résultats prometteurs obtenus dans ces premières étapes de la recherche dans le domaine des véhicules intelligents démontrent qu'une pleine automatisation de trafic (au moins sur des autoroutes ou des routes suffisamment structurées) est techniquement faisable. Néanmoins, au-delà des problèmes techniques, d'autres aspects se doivent d'être soigneusement pris en considération, comme par exemple les aspects légaux liés à la responsabilité en cas de faute ou suite à un comportement incorrect du système, de même que l'influence de la conduite automatique sur les passagers humains.

Par conséquent, une longue période d'essais approfondis doit précéder la mise à disposition de ces systèmes sur le marché. Un système routier entièrement automatisé avec des véhicules intelligents pratiquant l'échange d'informations semble difficilement envisageable avant une dizaine d'années.

Pour l'instant, l'automatisation complète sera limitée aux infrastructures spéciales, comme dans le cadre de certaines applications industrielles ou de moyens de transports en commun déterminés. Partant de là, la technologie des véhicules intelligents pourra être graduellement étendue à d'autres secteurs de transport. On peut citer à ce titre le transport de

marchandises par camions (véhicule onéreux) où le coût du système de conduite automatisée sera négligeable par rapport à celui du véhicule et du service fourni.

Finalement, une fois que la technologie sera maîtrisée et que des algorithmes puissants et fiables auront passé avec succès les épreuves d'une expérimentation en vraie grandeur, une intégration massive de ces systèmes sera pensable sur les véhicules privés, mais cela reste encore un objectif pour le moyen terme. Dans cette optique, l'approche « attelage virtuel », parce qu'elle cible un problème *a priori* moins vaste, permet d'obtenir plus rapidement un résultat proche d'une automatisation complète et fiable de l'activité de conduite automobile. Qui plus est, cette faculté de suivi automatique doit être considérée non pas comme concurrente de l'approche « traditionnelle », mais elle est bien comme complémentaire. Investir dans cette voie conserve donc tout son sens.

La suite de ce mémoire présente donc le capteur intelligent basé sur la vision que nous avons défini, développé et mis au point dans le contexte de l'attelage virtuel. Avant toute chose, nous développons dans le chapitre suivant la notion de « capteur intelligent » qui nous permet de formaliser les différentes propriétés qu'un tel dispositif se doit de présenter.

*C*e chapitre introduit la notion de « capteurs intelligents » et présente les différentes définitions à la base de ce concept. Le caractère « utile » du capteur intelligent se manifeste clairement dès lors qu'on s'intéresse au cheminement de l'information. Dans cette optique une attention particulière lui y accordée. Ensuite, nous présentons les différentes architectures de capteurs intelligents et le rôle de « l'intelligence » dans ces derniers. Nous abordons également les différentes méthodes permettant d'intégrer l'intelligence au capteur et les perspectives d'évolution concernant ce domaine. Enfin, nous passerons en revue les différentes approches de conception des capteurs intelligents que nous mettrons en application pour identifier les différents services et fonctionnalités que doit s'approprier le capteur stéréoscopique intelligent pour contribuer à la réalisation de l'attelage virtuel.

II.1. INTRODUCTION

L'intégration des fonctions de traitement du signal au sein même des capteurs est représentative de la notion de « capteur intelligent ». Les incessants progrès dans ce domaine sont un archétype d'innovation technologique. L'incorporation des traitements dédiés au traitement du signal permet d'augmenter la flexibilité de conception des dispositifs de perception (les capteurs) et de réaliser de nouvelles fonctions complexes. L'autre conséquence est une réduction de la charge sur les unités centrales de traitement et les dispositifs de transmission de l'information obtenue, en distribuant les opérations dans les systèmes de mesure eux-mêmes [Gia.86, Bri.86].

Les progrès rapides dans les technologies connexes à la perception et à la commande impactent très largement le domaine. L'accroissement de la puissance de calcul disponible et l'amélioration des capteurs, tant en performances qu'en coût, augmentent le nombre des systèmes automatisés, offrant par ailleurs une qualité de service accrue. Ces systèmes

autorisent et soutiennent à leur tour la progression des capteurs intelligents, alimentant une sorte de « cercle vertueux ».

Les domaines d'application de la perception et de la commande se sont rapidement diversifiés. Les exemples typiques des domaines nouvellement développés sont la Robotique, la mesure environnementale et les secteurs biomédicaux. La diversité des applications de la

Robotique intelligente exige des capteurs très avancés afin d'égaliser ou de dépasser la « perception naturelle » des opérateurs humains. Le secteur biomédical et les systèmes de diagnostic sont prometteurs, compte tenu de l'existence d'un fort besoin en la matière. Par ailleurs, l'instrumentation environnementale employant des systèmes de télédétection nous informe des crises environnementales pouvant potentiellement frapper la Terre ou de l'étendue des dégâts en cas de désastre avéré.

La réalisation de fonctions à la fois utiles et attrayantes au sein de dispositifs avancés représente le fruit de l'émergence du concept des capteurs intelligents. Dans ce qui suit, nous abordons les principes de fonctionnement de ces derniers.

II.2. POURQUOI METTRE L'INTELLIGENCE AU SEIN DES CAPTEURS ?

La réponse à cette question passe nécessairement par l'analyse de l'acheminement de l'information entre les différents « mondes » (physique, logique et le monde « intellectuel » de l'homme). Ceci nous permettra d'établir la problématique associée aux capteurs intelligents.

II.2.1. L'ITINERAIRE DE L'INFORMATION

Les concepts de mesures, de commande et de communication homme-machine permettent de dresser un tableau d'ensemble de ce qu'est un capteur intelligent. D'une façon générale pour ce qui touche aux systèmes de perception et de commande, l'information circule non seulement entre les objets et les systèmes de perception, mais aussi entre ces entités et l'opérateur (communications homme-machine). Ce flux peut être représenté comme une communication entre trois « mondes » différents, (voir Figure II.1). On peut facilement comprendre qu'un échange régulier et efficace de l'information est essentiel et facilite l'exploitation du système par l'utilisateur [Yam.91]. Les trois mondes dont il est question plus haut sont définis de la façon suivante:

- 1) **Le monde physique** représente les objets mesurables et contrôlables. Les lois naturelles dominant ce monde dans lequel la causalité est strictement établie. L'information est transmise en tant que signal physique.
- 2) **Le monde logique** représente le système de traitement de l'information, de la mesure et de la commande. Des règles rationnelles règnent sur ce monde dans lequel l'information est décrite par des codes logiques.
- 3) **Le monde intellectuel humain** c'est le monde mental interne du cerveau humain. L'information est traduite en connaissances et concepts. Les lois dominant ce monde « restent à établir de façon formelle et sont l'objet d'actives et nombreuses recherches ».

L'information dans le monde physique est extraite par des techniques de perception et est transférée au monde logique. Elle y est alors traitée selon les objectifs des systèmes de perception et de commande et est finalement interprétée par l'homme au travers des interfaces homme-machine, qui exhibent des données mesurées et traitées. De cette manière, l'information est à la fois acheminée vers le monde intellectuel humain et « formatée » pour pouvoir s'insérer dans le cadre de la connaissance et des concepts de celui-ci. Ainsi, les différentes informations obtenues sont structurées et deviennent une partie de la connaissance. A leur tour, ces connaissances peuvent s'assembler en concepts qui peuvent eux-mêmes s'agréger en domaines de la science et de la technologie.

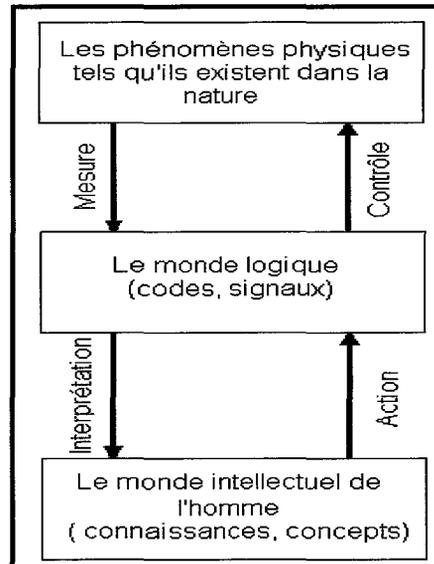


Figure II.1 : L'échange de l'information entre les trois mondes.

Le comportement humain est généralement basé sur l'information et la connaissance préalablement acquise. Les intentions sont exprimées en tant que commandes et elles sont transmises au système par des actions à travers les interfaces homme-machine. Le système de contrôle logique régule ses propres objets au travers d'actionneurs dans le monde physique, les actions entreprises étant basées sur la commande reçue. Nous pouvons détailler la chaîne « perception et contrôle » par un procédé de type perception, identification, action et contrôle impliquant trois mondes différents dans lesquels l'objectif humain est atteint dans le monde physique. Entre le monde physique et le monde logique, l'information s'échange entre les systèmes de perception et les systèmes de contrôle. Les capteurs et les actionneurs agissent donc en tant qu'interfaces. Plus précisément, les capteurs sont les dispositifs d'entrée des systèmes de perception. Ils extraient l'information issue des objets du monde physique et transmettent le signal correspondant.

Entre le monde logique et le monde humain, les communications sont assurées par une interface homme-machine. Un capteur intelligent permet de rapprocher la frontière entre monde physique et monde logique de celle séparant monde logique et monde humain. Cela signifie en pratique qu'une partie des traitements de l'information effectuée généralement

dans le monde logique est remplacée par un traitement au sein du capteur lui-même. Une interface homme-machine intelligente comporte un système dans lequel la frontière entre le monde humain et le monde logique est décalée vers le monde humain. Ainsi, une partie de l'analyse et de l'interprétation des données fournie est effectuée par l'interface homme-machine elle-même.

II.2.2. CE QU'ON ATTEND D'UN CAPTEUR INTELLIGENT

Un capteur est dit intelligent lorsqu'il est capable d'élaborer localement sa mesure, en prenant éventuellement en compte les grandeurs d'influence et en vérifiant la cohérence des informations. Il est également capable d'envoyer sa mesure à la demande ou de manière systématique à destination du système qui devra l'exploiter. L'interface de communication permet également au capteur intelligent de recevoir les informations du système nécessaires à l'élaboration de sa mesure et à sa validation. Elle peut également être utilisée dans les phases de calibration et de mise en service de l'équipement dans son environnement de travail (pour plus d'information voir [Sta.94]).

Les technologies associées aux capteurs ont évolué au rythme des besoins. Il est ainsi possible de décrire les grands champs d'investigation actuels et futurs associés aux capteurs intelligents en dressant le panorama des attentes placées dans ces derniers. Il est ainsi possible de dégager les axes majeurs suivants :

- (1) Nous pouvons percevoir les paramètres physiques d'objets dans leur état normal avec à la fois une très grande précision et une grande sensibilité. Cependant, la détection des anomalies et des dysfonctionnements est assez peu développée. La capacité à détecter des défauts et à prévoir des défaillances constitue une fonctionnalité de plus en plus recherchée.
- (2) Les technologies de perception actuelles peuvent mesurer avec précision des quantités physiques ou chimiques en un seul point. Cependant, cette mesure se complique dès lors que la ou les grandeurs qu'on cherche à caractériser ne sont plus « localisées » mais « réparties » (que se soit spatialement ou temporellement) dans un espace multidimensionnel, dont notre environnement fourni un bon exemple.
- (3) Des grandeurs physiques « élémentaires » (masse, température, vitesse, etc.) peuvent être mesurées avec précision et avec une très grande sensibilité. Cependant, certaines informations considérées comme « élémentaires » du point de vue de la perception humaine, comme par exemple le goût ou l'odeur, nécessitent la mise en œuvre de capteurs complexes auxquels doivent être associés de traitements performants.

Les trois points abordés ci-avant introduisent une problématique commune qui est celle de la définition de l'objet de la mesure. Si nous pouvons le définir clairement et établir son modèle précis, il est possible de lui associer un ensemble de grandeurs caractérisant son état sans ambiguïté et par conséquent de choisir les capteurs les plus appropriés à la mesure de

celui-ci. La mise en œuvre d'une telle démarche, classique dans le domaine de l'automatisation des processus, s'avère délicate à mettre en œuvre dès lors qu'est recherché une certaine « universalité » du dispositif.

Dans ce contexte, cette approche « systématique » issue de l'Automatique présente (d'un point de vue purement pragmatique) des performances moindres que celles que réalise quotidiennement chaque être humain. Ainsi, notre expérience nous permet-elle de caractériser l'état normal ou non d'un objet familier ou de prévoir le dysfonctionnement de celui-ci grâce (entre autre) aux connaissances collectées sur ce dernier. Les hommes et les animaux peuvent par ailleurs exploiter leurs capteurs naturels suivant plusieurs modes, permettant d'aboutir à une collecte et une interprétation des données en rapport avec la situation rencontrée. Ainsi sommes-nous capables de reconnaître une voix connue dans le « brouhaha » ou de tendre l'oreille pour percevoir un son ténu dans un silence presque complet. L'une des forces du système de perception humain est sa faculté à exploiter conjointement et simultanément plusieurs sortes de capteurs avec sa « connaissance sensorielle », sans nécessiter de modèle *a priori*. Nous sommes donc capables de traiter des problèmes de perception complexes grâce à ces mécanismes de « fusion » entre capteurs et entre « données capteur » et connaissances. Ces mécanismes représentent une approche performante pour les capteurs intelligents.

II.3.3. FONCTIONS ASSOCIEES A « L'INTELLIGENCE » DES CAPTEURS

Ce qu'il est convenu d'appeler « l'intelligence » du capteur réalise un traitement distribué des signaux dans la couche inférieure de la hiérarchie du système de perception. Le rôle de la fonction de traitement des signaux dans les capteurs intelligents peut être récapitulé comme suit :

- 1) Pallier les caractéristiques « fâcheuses » inhérentes au dispositif de la perception.
- 2) Améliorer la qualité du signal en vue de l'extraction des caractéristiques utiles des objets.

L'opération la plus connue pour pallier les caractéristiques inhérentes au capteur est la compensation. Elle consiste à supprimer l'influence des variables indésirables sur le mesurande. Pour améliorer la qualité du signal, divers traitements sont adoptés afin d'extraire les caractéristiques utiles. Les buts du traitement du signal sont d'éliminer le bruit afin d'extraire la caractéristique la moins bruitée possible, c'est à dire la plus riche en informations pertinentes compte tenu du phénomène à caractériser.

Les techniques de traitement du signal utilisent généralement les différences entre les réponses dynamiques dues au signal « utile » d'une part et au bruit d'autre part. Elles sont divisées en trois types différents :

- a) Le traitement dans le *domaine fréquentiel*.

- b) Le traitement dans le *domaine temporel*.
- c) Les traitements dans le *domaine spatial*.

Ces traitements constituent une approche dynamique pour l'amélioration de la sélectivité du signal (Figure.II.2).

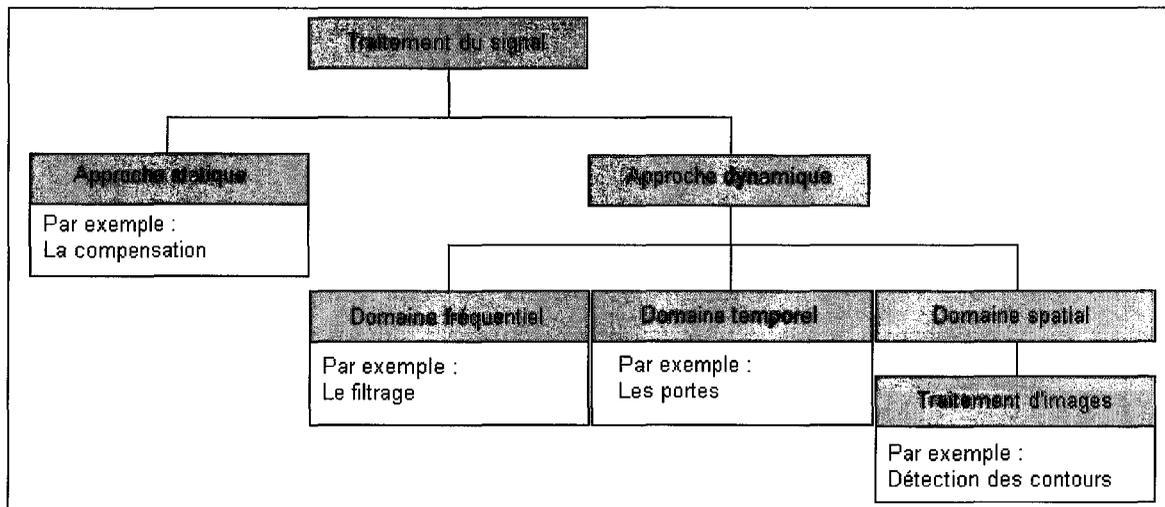


Figure II.2 : La classification des traitements du signal lié au capteur intelligent.

L'archétype du traitement de signal embarqué sur un capteur intelligent est le filtrage du bruit en temps réel. Cette tâche est effectuée par des filtres passe-bas, passe-haut ou passe-bande dans le domaine fréquentiel. Cette fonction peut être prise en charge par des circuits analogiques, bien qu'ils n'offrent pas la même souplesse que leurs homologues numériques. Tout aussi important que le filtrage fréquentiel, le filtrage spatial dans le cadre du traitement d'images permet par exemple de rendre les signaux visuels plus « signifiants » : Elimination du bruit, calcul de valeurs moyennes (équivalent à un filtrage passe bas), rehausse des contours ou encore extraction d'objets dans l'image traitée.

Au passage, la quantité d'informations issues des capteurs d'images étant énorme, le traitement d'image basique devient une nécessité afin de réduire la charge de travail induite sur les couches plus élevées.

Nous pouvons résumer les fonctions associées à l'intelligence embarquée sur les capteurs en disant que la plus importante d'entre elles est d'améliorer la sélectivité du signal des différents dispositifs de perception dans le monde physique. Ceci inclut les opérations simples effectuées sur les signaux de sortie des différents capteurs, telles que les traitements basiques des données (filtrage, compensation, etc...). Cependant, ceci n'inclut pas l'optimisation des paramètres des dispositifs ou l'intégration des signaux issus des différents capteurs, car ceci exige la connaissance des autres dispositifs de perception et d'informations associées aux objets observés.

II.3.4. FONCTIONS DE L'INTELLIGENCE DANS LE TRAITEMENT DE L'INFORMATION

Le rôle de l'intelligence dans la couche médiane est d'organiser les sorties multiples de la couche basse et de produire des sorties intermédiaires. Le rôle le plus important est d'extraire les caractéristiques essentielles de l'objet. Dans la couche médiane du système de traitement, les signaux de sortie des multiples capteurs sont combinés ou intégrés. Ainsi, les éléments extraits sont alors utilisés par l'intelligence de la couche supérieure pour identifier la situation. Ce traitement est effectué dans le monde logique.

Nous pouvons considérer l'intégration ou la fusion des signaux des capteurs comme une architecture de base pour concevoir un système de mesure intelligent adaptatif. Les signaux du capteur pour les différents mesurandes sont combinés dans la couche médiane et les résultats fournissent une nouvelle information utile. L'ambiguïté ou l'imperfection du signal d'un mesurande peut être compensée par un autre mesurande. Ce traitement crée une nouvelle phase d'information.

Une autre fonction importante de la couche médiane est l'adaptation des paramètres des capteurs afin d'optimiser les performances du système dans sa globalité. Cette adaptation est conduite sur la base des caractéristiques extraites et la connaissance *a priori* du signal traité. La connaissance provient de la couche supérieure sous la forme d'un algorithme d'optimisation.

II.3. LES APPROCHES DE REALISATION DES CAPTEURS INTELLIGENTS

Selon Yamasaki [Yam.84], il y a trois approches différentes pour mettre en oeuvre l'intelligence d'un capteur (Figure. II.3).

- 1) Intégration du capteur avec les ordinateurs, équivalente à une *intégration intelligente*.
- 2) Utilisation des matériaux fonctionnels spécifiques ou *matériaux intelligents*.
- 3) Utilisation de structures géométriques fonctionnelles équivalentes à une *structure intelligente*.

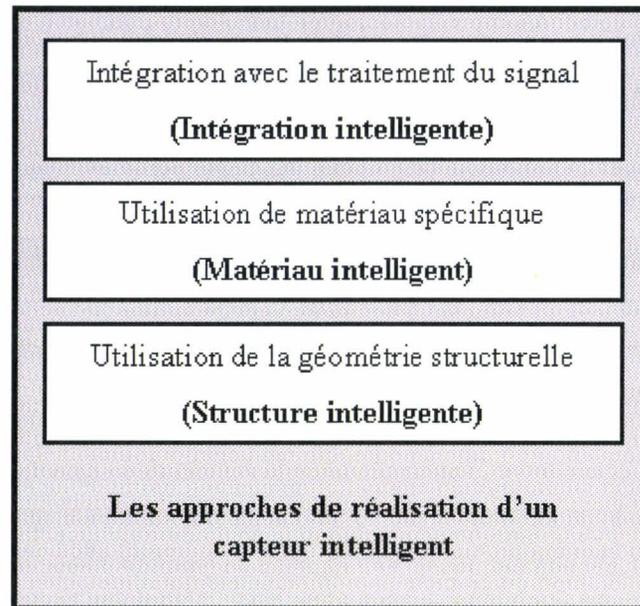


Figure II.3 : Trois approches différentes pour mettre en œuvre l'intelligence du capteur.

La première approche est la plus répandue. Un exemple typique de l'intégration intelligente est l'association de la fonction « mesurer » avec la fonction de traitement de l'information. C'est clairement ce qui est réalisé lorsque le dispositif de perception intègre un microprocesseur. L'algorithme de traitement est dans ce cas programmable et peut être modifié ultérieurement (mise au point, adaptation, « redesign », etc.).

Le traitement de l'information consiste en l'optimisation, l'extraction des simples caractéristiques, etc. Quelques opérations sont effectuées en temps réel et en parallèle avec des circuits analogiques ou à travers le réseau. Nous appelons une telle intégration un *capteur – calculateur* [Van.92].

Dans les deuxième et troisième approches, les propriétés du capteur (liées au matériau ou à la structure suivant le cas) permettent à elles seules de discriminer le signal utile du bruit. La sélectivité du signal est ainsi améliorée.

Dans la deuxième approche plus particulièrement, la combinaison unique des caractéristiques de l'objet et de celles du capteur contribue à la réalisation d'une sélectivité presque idéale du signal. Dans la troisième approche, la fonction de traitement des signaux est réalisée par la structure géométrique ou mécanique du capteur (c'est le cas par exemple de certains dispositifs optiques ou ultrasonores.).

II.4. MODELISATION DU CONCEPT DU CAPTEUR INTELLIGENT :

Le développement d'un système, d'un produit, d'un composant ou d'un service, ou plus exactement d'une « fourniture » telle que la norme le définit [Afn.91], passe par la description

des activités complémentaires permettant, à partir de l'expression du besoin initial, d'aboutir à une fourniture opérationnelle (dans notre cas un capteur).

Le cycle de développement a été présenté de différentes manières. On trouve principalement, le cycle de développement en cascade [Roy.70], le cycle de développement en V [Afn.96], le cycle de développement en spirale [Boe.88] et le cycle de développement en X qui est une extension du cycle de développement en V [Hod.91] et qui fait apparaître le concept de réutilisation des composants matériels et logiciels.

II.5. LE CAPTEUR INTELLIGENT EN TANT QU'EQUIPEMENT INTELLIGENT

Les systèmes développés pour les transports intelligents sont en réalité des équipements intelligents et font ainsi partie de l'ensemble des systèmes temps-réel réactifs. Après une définition des qualificatifs relatifs au système étudié (i.e. temps-réel et réactif), on s'attardera sur les domaines d'application plus particuliers des équipements intelligents.

1. Temps-réel :

La définition de l'aspect « temps-réel » d'un système est un problème à part entière, compte-tenu du fait que cette notion est très fortement « couplée » à la dynamique de celui-ci. Le problème qui nous concerne ici est le suivant :

Suivant quels critères qualifions-nous un système de « temps réel » ? Plusieurs points de vue peuvent être adoptés. Le premier point de vue consiste à qualifier tout système dont les données et les résultats sont relatifs à des événements « en cours » de « système temps-réel ». Ceci revient à dire que les informations collectées doivent encore avoir un rapport avec l'état du système au moment où celles-ci sont prises en compte. Nous aboutissons à l'aphorisme habituellement énoncé dans les systèmes temps-réel [Lem.99] :

« Tout résultat juste mais hors délai est un résultat faux »

Le délai dont il est question est fonction de la dynamique du processus piloté par le système qualifié de temps-réel. Ainsi plus celle-ci sera rapide, plus les contraintes « temps-réel » seront strictes et difficiles à respecter matériellement. On peut avoir plusieurs ordres de grandeurs du temps de réponse maximal toléré par rapport au processus considéré :

- De l'ordre de la milliseconde pour les systèmes embarqués (avion, voiture, missiles),
- De l'ordre de la seconde pour les systèmes de détection d'obstacles,
- Jusqu'à plusieurs heures pour certains réacteurs chimiques.

D'une façon générale, tout système automatisé, et a plus forte raison tout équipement intelligent, se doit de respecter des contraintes de temps plus ou moins sévères suivant le domaine d'application considéré. Ces contraintes peuvent être imposées dans le cahier des charges (par exemple en terme d'échéance maximale) ou non. Par ailleurs, les données temporelles apparaîtront inévitablement au cours du cycle de développement ne serait-ce que pour ordonner les différentes actions du système automatisé.

1. Réactif :

Le terme de système réactif a été introduit par [Har.85]. Un système est dit réactif lorsque son comportement dépend directement des informations qu'il échange avec son environnement. Ces informations arrivent au système sous forme de stimuli. Un stimulus est l'occurrence d'un événement en provenance de l'environnement, de l'application vers le système. Cette occurrence est furtive, i.e, si elle n'est pas traitée et/ou mémorisée à l'instant de son occurrence, elle est perdue [Del.95]. Le système doit alors réagir de manière instantanée pour pouvoir prendre en considération l'information contenue dans le stimulus. La rapidité de la réaction dans les systèmes réactifs est un critère déterminant afin d'assurer un fonctionnement adéquat et sûr. De ce fait les systèmes réactifs sont considérés comme étant une classe particulière des systèmes « temps-réel ». Le comportement réactif de tout équipement intelligent est caractérisé suivant la nature de son environnement, à événements discrets ou hybrides. Généralement, les systèmes automatisés intègrent les deux types de comportement à la fois.

II.6. DOMAINES D'APPLICATIONS DES CAPTEURS INTELLIGENTS

On dénombre quatre grands domaines d'application dans lesquels les capteurs intelligents sont plus particulièrement usités :

- ✓ L'industrie manufacturière, caractérisée par des contraintes temps-réel fortes et des contraintes environnementales variables suivant le type d'application.
- ✓ L'industrie des « process continus », caractérisée par des contraintes temps-réel plutôt faibles mais qui doit satisfaire des contraintes environnementales fortes (processus chimiques, thermiques, ...).
- ✓ Les systèmes embarqués, qui doivent également satisfaire des contraintes temporelles très variables suivant les applications, mais des contraintes relatives à l'environnement très fortes,
- ✓ Enfin, le tertiaire, qui est le plus souvent le moins contraignant tant au niveau temps-réel qu'environnemental.

Les deux premiers domaines d'applications concernent plus particulièrement les capteurs et les actionneurs intelligents. Les systèmes embarqués sont similaires aux deux précédentes, et sont en plus assujettis à des contraintes d'encombrement, et de consommation énergétique très fortes (espace disponible, poids, puissance, consommation, ...). Le tertiaire ne vise pas un type d'applications bien identifiées, tout au contraire des trois premiers. En effet, le tertiaire est très varié. On peut y trouver par exemple tout ce qui concerne les problèmes de détection utilitaire (incendie, intrusion ...etc.).

II.7. LA CONCEPTION DES CAPTEURS INTELLIGENTS

La problématique de base dans la conception des instruments intelligents en général et des capteurs intelligents en particulier consiste à identifier la meilleure technique pour le développement d'un capteur adapté à un problème donné.

En effet, le processus de développement est étroitement lié à l'approche adoptée pour la conception, qui est à son tour issue de la manière avec laquelle nous percevons les choses. Généralement, les approches de conception font partie de l'un de ces trois points de vue :

- **FAIRE** : On modélise ici ce que le système **fait**. Ce sont les approches orientées fonctions ou traitements.
- **ETRE** : On modélise alors ce que le système **est**. Ce sont les approches orientées données.
- **DEVENIR** : On modélise plutôt ce que le système **devient**. Ce sont les approches orientées comportement ou état.

Les points de vue abordés ci-dessus séparément, ne présentent pas le seul point de vue qu'on peut adopter pour la conception des équipements intelligents. Des approches dites « combinées » peuvent être adoptées pour prendre en compte un maximum d'aspects de l'équipement intelligent. Dans ce qui suit, nous allons décrire ce que recouvrent les termes précédemment introduits.

II.7.1. APPROCHE FONCTIONNELLE

L'approche orientée fonctions s'articule autour de ce que fait l'équipement sans s'intéresser très finement à comment il le réalise exactement. Ce qui intéresse le concepteur c'est la sortie d'une « boîte noire ». Ce point de vue laisse le choix au concepteur-réalisateur d'adopter la technologie la mieux adaptée en terme de performances attendues et de coût.

II.7.1.1. LA DECOMPOSITION FONCTIONNELLE

L'idée principale de cette approche est de décomposer le problème de conception en plusieurs fonctions, sous-fonctions et interfaces entre ces fonctions. On associe à chaque fonction déterminée un ou plusieurs traitements de bases. En effet, on se penche sur la

spécification des différents traitements et interfaces nécessaires pour répondre aux exigences du cahier des charges de l'équipement en question (capteur, actionneur, ...etc.).

Néanmoins, il est difficile d'estimer si la spécification apportée recouvre tous les aspects de ce que la nouvelle application doit faire, puisque l'approche par décomposition fonctionnelle n'offre que peu de correspondance entre la fonctionnalité et le domaine d'application. De ce fait, la complétude du problème n'est donc pas atteinte si on néglige certaines parties qui n'apparaissent pas explicitement dans l'énoncé du cahier des charges. Par exemple, le calcul de la vitesse relative, nécessite deux types de données : la distance relative et le temps. Pour pouvoir s'assurer de la complétude de notre spécification, nous devons soit définir une nouvelle fonction permettant de fournir la ou les données manquantes, soit celles-ci doivent être fournies par l'environnement de l'équipement intelligent : il faut alors l'explicitier. Par conséquent, la décomposition fonctionnelle seule ne couvre pas tous les aspects du problème à traiter. C'est pour cette raison que celle-ci doit être associée à des décompositions par flots de données.

II.7.1.2. LA DECOMPOSITION PAR FLOTS DE DONNEES

Cette approche, en complémentarité à la décomposition fonctionnelle, permet de regrouper les flots de données et de contrôle en entrée et en sortie de l'équipement intelligent (capteur, actionneur, ...etc.).

Les raffinements successifs des différents traitements permettent d'exprimer les traitements en termes de flots de données. De ce fait, cette méthode exige qu'on ait une bonne compréhension des traitements utilisés pour obtenir les différentes données, ou des besoins en terme de données de sorties relatifs aux traitements ainsi que les flots de contrôle générés.

En effet, cette approche permet au concepteur – réalisateur de représenter directement le monde réel à partir du cahier des charges. Ce dernier, comme le révèle la décomposition fonctionnelle, donne les différents flots de données demandées générés par la spécification en fonctions. Ces fonctions, sont des traitements qui transforment et consomment des données, elles-mêmes issues d'autres fonctions. Par conséquent, cette approche nécessite une bonne analyse des flots de données de la part du concepteur et du demandeur de service.

La décomposition fonctionnelle par flots de données est à la base de plusieurs langages de conception très connus tels que : S.A (Structured Analysis) et son extension aux systèmes temps-réels incluant des problèmes temporels dans la spécification réalisée par Ward (S.A.R.T) [War.86, Hat.90].

II.7.2. APPROCHE INFORMATIONNELLE

Cette approche incite à s'intéresser à ce qu'est l'équipement, dans son ensemble et, en particulier, à ses propriétés intrinsèques. L'équipement est alors décrit minutieusement. Cette particularité de l'approche orientée donnée implique qu'on s'intéresse au type d'équipement et à la technologie utilisée, i.e. comment est réalisé ce que l'équipement fait. Cette méthode

permet de modéliser tous les aspects liés à l'information. On y retrouve par exemple les diagrammes entités-relations définies par Peter Chen : Objet + attributs + Relations, voir [Che.76]. Ces types de diagrammes sont particulièrement utilisés pour structurer des systèmes d'information de grande taille. Ainsi, tous les langages de description formelle appartiennent à la catégorie des approches orientées données.

Néanmoins, les informations ainsi modélisées ne traitent pas certains concepts comme le service, en termes de comportement et de structure, de liens généralisation-spécialisation et de composés-composants. Ainsi, les approches orientées données ne suffisent pas en elles-mêmes à l'entière conception d'un équipement intelligent mais elles restent incontournables pour prendre en charge l'aspect informationnel attaché à celui-ci.

II.7.3. APPROCHE COMPORTEMENTALE

Les approches orientées comportement s'intéressent plus à la dynamique du système et à son évolution dans le temps qu'à son état statique qui est plus représenté par les approches orientées fonctions (ou traitements). Ainsi, il est courant de joindre au modèle obtenu soit par une approche orientée données soit par une approche orientée fonctions, un modèle de comportement « dynamique ». En effet, les approches orientées comportement s'appuient sur le modèle développé par l'une des deux approches précédentes afin de déterminer les changements du système dans le temps, puisque le comportement du système est étroitement lié de qu'il fait et/ou sur quoi il le fait.

Parmi les approches orientées comportement les plus connues, on trouve :

- Approches par systèmes de transitions,
- Approches par Réseaux de Petri,
- Approches par GRAFCET,
- StateCharts,
- Langages asynchrones : ELECTRE,

Il existe un nombre important de langages permettant d'étudier le comportement dynamique des systèmes sans autant pouvoir exclure une analyse « statique préalable » afin de modéliser correctement le comportement. La multitude d'approches s'explique par le fait qu'il existe plusieurs moyens d'aborder le comportement d'un système : les notions de séquençement de tâches, parallélisme, d'exclusion, de préemption, de synchronisation, ...etc. ne sont pas toutes prises en charge par les modèles proposés. Pour plus de détail sur les différentes approches orientées comportement, le lecteur peut se référer à [Lem.99].

II.7.4. APPROCHE HYBRIDE

La constatation dominante lors de l'analyse des différentes approches introduites plus haut, nous laisse penser qu'une seule approche n'est pas suffisante pour prendre en compte tous les aspects d'un système : Données, traitements, et comportement.

De ce fait, la combinaison des différentes approches peut apporter un plus considérable à la tâche de conception d'un capteur intelligent. On trouve principalement des méthodes qui allient à la fois le côté « Données » et le côté « Fonctions » car elles permettent d'avoir une bonne vision du système tant au point de vue fonctionnel qu'au point de vue structurel. Parmi les approches hybrides les plus connues, la méthode S.A.D.T (Structured Analysis and Design Technique) utilise des actigrammes inspirés d'une représentation par flots de données et de l'analyse structurée [Igl.88, Geh.94].

Afin d'apporter une vision incluant l'aspect comportemental aux aspects données et traitements, les approches orientées objets ont fait leur apparition dans la modélisation.

II.7.5. APPROCHE ORIENTEE OBJET

Cette approche vise à décomposer le système à concevoir en objets interagissant les uns avec les autres. L'apport principal de cette méthode est son adéquation avec les propriétés intrinsèques liées à tout objet, puisque celui-ci se définit par des informations le caractérisant, un fonctionnement et également un comportement. Le concept de l'approche orientée objet se base sur trois principes fondamentaux [Aub.91] :

- a) **ABSTRACTION** : Chaque objet dispose d'un état relatif à ses attributs et d'un ensemble d'opérations qui fournissent des services aux autres objets qui sont clients de l'objet.
- b) **ENCAPSULATION** : L'encapsulation d'un objet permet de séparer ses aspects externes, accessibles aux autres objets, de ses aspects internes concernant l'implémentation et donc généralement invisibles aux autres objets.
- c) **HERITAGE** : L'héritage est le mécanisme qui définit entre les classes une relation dans laquelle une classe partage la structure ou le comportement défini dans une ou plusieurs classes. La relation d'héritage organise hiérarchiquement les classes en graphe d'héritage, et permet d'hériter d'une ou de plusieurs classes à la fois [Lut.97].

Afin, d'utiliser pleinement les capacités du concept objet une nouvelle génération des langages de programmation a été développée. On peut citer parmi eux les plus connus : SmallTalk, C++, CLOS ou Eiffel, Java, et dernièrement le C#. Il existe également un certain nombre de méthodes utilisées pour la conception. Parmi elles, on peut citer les méthodes, O.O.A. (Object Oriented Analysis) [Coa.91], la méthode H.O.O.D (Hierarchical Object Oriented Design) [Esa.89], la méthode O.M.T (Object Modeling Technique), et dernièrement

U.M.L (Unified Modelisation Language) [Boo.00]. L'inconvénient principal de la modélisation objet est sa difficulté d'intégration sur des composants électroniques par exemple.

Les techniques de modélisation sont nombreuses. Malheureusement, la plupart d'entre elles n'ont qu'une vision partielle du domaine, voire deux aspects du système. Pour prendre un système en son entier il faut utiliser plusieurs approches, qu'il faut adapter les unes aux autres.

II.8. LE CAPTEUR INTELLIGENT DEDIE A L'ATTELAGE VIRTUEL

Le système de l'attelage virtuel tel qu'il a été introduit, est un équipement d'aide à la conduite permettant au conducteur de suivre automatiquement un véhicule, que nous appelons ici « tracteur », le long d'un trajet. Quand le véhicule suiveur se trouve derrière le véhicule tracteur, le conducteur peut engager le système d'attelage virtuel. Le cahier des charges décrivant le problème de l'attelage virtuel se partage en plusieurs points :

- La vitesse du véhicule est contrôlée par une action sur un potentiomètre (la commande d'accélération dans notre cas) qui commande la puissance fournie au moteur électrique.
- La consigne de vitesse du système d'attelage est la grandeur réglante permettant d'assujettir la distance intervéhiculaire à une consigne imposée par les conditions d'utilisation.
- La distance intervéhiculaire est déterminée par un système de vision stéréoscopique.
- Quand le système d'attelage virtuel est engagé, le conducteur doit rester, à tout instant, maître de son véhicule. Il peut ainsi engager des actions qui modifient la vitesse de son véhicule. Il peut le faire de deux manières différentes en ce qui concerne la vitesse : en agissant sur l'accélérateur ou le frein.
 - En appuyant sur l'accélérateur, la vitesse du véhicule augmente et le système d'attelage virtuel est désactivé.
 - En appuyant sur le frein, le système de l'attelage est désactivé également.

Dans les deux cas précédents, le conducteur peut activer le système d'attelage virtuel, pour suivre le véhicule se trouvant devant lui et sur la même ligne, à tout moment en appuyant par exemple sur un bouton de type « Réengager ».

Les besoins du système d'attelage virtuel ont été ici posés en quelques lignes. Il s'agit essentiellement du besoin du conducteur du véhicule. Des attentes supplémentaires sont exposées dans ce qui suit.

II.8.1. BESOINS SUPPLEMENTAIRES

Le conducteur possède plusieurs autres besoins que nous allons définir ci-dessous. Ces besoins ne sont pas directement en rapport avec le problème de l'attelage virtuel mais peuvent interférer avec lui.

- A la mise en service et à la demande ultérieure du conducteur, une procédure de calibration des caméras (du capteur stéréoscopique) détermine les paramètres intrinsèques et extrinsèques du système de vision. Cette procédure permet au système d'exprimer la relation qui relie le capteur au monde extérieur (son environnement).
- Le conducteur peut enregistrer les paramètres issus de la calibration, pour une utilisation ultérieure, ainsi une procédure de vérification de l'exactitude de ces paramètres est lancée avant une utilisation effective.
- De la même manière, le conducteur peut fixer la distance minimum qui doit séparer son véhicule du véhicule tracteur pour les faibles vitesses (dans les embouteillages par exemple) une alarme peut ainsi être déclenchée automatiquement si la distance n'est pas respectée, le système de l'attelage fournissant de cette façon un autre service d'aide à la conduite.
- Une alarme doit signaler la défaillance du système, au cas où un problème survient au niveau matériel ou logiciel.

Ces quatre fonctionnalités supplémentaires sont-elles aussi à prendre en compte lors de la spécification des besoins.

Cette présentation « en langage naturel » du cahier des charges n'est évidemment pas complète. D'une part, les informations ne sont pas formalisées ce qui les rend difficiles à exploiter en l'état et d'autre part, les informations relatives aux aspects comportementaux ne sont pas prises en compte. Ces informations n'étant pas connues *a priori*, elles seront à définir au cours de l'étude.

II.8.2. SPECIFICATIONS DES BESOINS DU SYSTEME D'ATTELAGE VIRTUEL

Dans le but de bien cerner le système d'attelage virtuel en tant qu'équipement devant fournir différents services, nous allons expliciter rigoureusement les différents besoins des utilisateurs en accord avec ce qui a été exprimé précédemment. Nous commençons tous d'abord par déterminer les différents consommateurs externes du système d'attelage virtuel et nous déterminons ensuite les différents services que doit fournir le système pour satisfaire les besoins des consommateurs ainsi que ceux nécessaires à son propre fonctionnement. Cette démarche de spécification est basée sur le modèle externe développé au L.A.I.L.²⁶ [Sta.94,

²⁶ L.A.I.L.: Laboratoire d'Automatique et d'Informatique Industrielle de Lille

Geh.94, Bay.95, Lem.99] et qui s'inspire de la méthode S.A.D.T²⁷ (combinaison d'une approche orientée donnée et d'une approche orientée fonctions) à laquelle on ajoute la partie des problèmes temporels et comportementaux que l'application fait ressortir.

II.8.2.1. LES CONSOMMATEURS EXTERNES

On recense trois grands consommateurs : *le superviseur* (le conducteur humain), ce que nous appellerons *le moteur*²⁸ et *le volant* (la direction), qui représentent l'environnement physique de l'ensemble « Système d'attelage virtuel ».

Les besoins exprimés par ces trois consommateurs sont les suivant, ils sont classés selon leur importance par rapport au système d'attelage :

Services Indispensables

- la calibration des caméras **Calibration**,
- la Commande Manuelle du Véhicule **CMV**,
- la Commande Automatique du Véhicule **CAV** (en agissant sur le frein, accélérateur et volant),
- l'estimation de la distance relative entre le véhicule suiveur et le véhicule tracteur **Distance_Relative**,
- l'estimation de la direction relative du véhicule tracteur par rapport au véhicule suiveur **Direction_Relative**
- l'estimation de la Vitesse Relative du véhicule tracteur par rapport au véhicule suiveur **VR**,

Services Optionnels

- le calcul de la distance parcourue, en utilisant le système d'attelage virtuel **Distance_Parcourue**,
- le calcul de la Vitesse Moyenne sur un parcours **VM**,

Les fonctions de consommation associées à chacun des consommateurs sont alors les suivantes :

$$\begin{aligned}
 f_c(\text{Conducteur}) &= \{ \text{Distance_Parcourue}, \text{VR}, \text{VM}, \text{Distance_Relative} \} \\
 f_c(\text{Moteur}) &= \{ \text{Calibration}, \text{CMV}, \text{CAV} \} \\
 f_c(\text{Volant}) &= \{ \text{Direction_Relative} \}
 \end{aligned}$$

²⁷ S.A.D.T : Structured Analysis and Design Technique

²⁸ Au sens « puissance motrice » à laquelle contribuent également les freins.

II.8.2.2. LES SERVICES QUE DOIT FOURNIR LE SYSTEME

On a recensé également les besoins en termes de services, que nous désignons de la façon suivante :

- *Calibration,*
- *Commande_Manuelle,*
- *Commande_Automatique,*
- *Elaboration_Distance,*
- *Elaboration_Distance_Relative,*
- *Elaboration_Direction_Relative,*
- *Elaboration_Vitesse_Relative,*
- *Elaboration_Vitesse_Moyenne,*
- *Saisie_Vitesse_Max,*
- *Saisie_Distance_Min,*
- *Détection_Entité_Accrochage*
- *Suivi_Entité_Accrochage*

La description de chaque service exprimé est ensuite détaillée et nous présentons schématiquement chacun d'entre eux sous forme de boîtes SADT selon un modèle classique [Igl.88, Geh.94]. Dans ce qui suit, nous indiquerons exclusivement les ressources impliquées directement dans l'attelage virtuel. Cette description ne concerne donc pas les ressources nécessaires au bon fonctionnement du véhicule (batterie, moteur, roues, ...etc.). Par ailleurs, les traitements spécifiques aux services définis ici seront détaillés dans le troisième chapitre de ce mémoire.

1. Service Calibration

Ce service permet d'extraire les paramètres intrinsèques et extrinsèques des caméras utilisées par le capteur stéréoscopique et de définir la géométrie du capteur. Il doit s'exécuter au commencement de toute utilisation du système d'attelage virtuel ou suite à une modification affectant l'une des caméras. Ce service s'arrête lorsqu'on lui signale que le nombre d'images nécessaires pour la calibration a été pris. Cet événement déclenche alors la mise à jour du paramètre **Calib_Params**.

Ce service a donc besoin d'un compteur du nombre d'images acquises et d'une donnée **Stop_Calib_IMAGES** qui devra indiquer que le nombre d'images « étalons » nécessaire à la calibration a été pris. Au niveau des conditions d'activations, il faut que le système d'attelage soit validé, cette donnée sera fournie par l'opérateur *via* la variable booléenne **Attelage_ON**, et le service sera activé sur la requête **Start_Calib** au libre choix de l'opérateur.

D'un point de vue comportemental, la non-détection d'une mire de calibration dans les images acquises implique l'abandon du service. Le processeur est ici une ressource partagée et vitale.

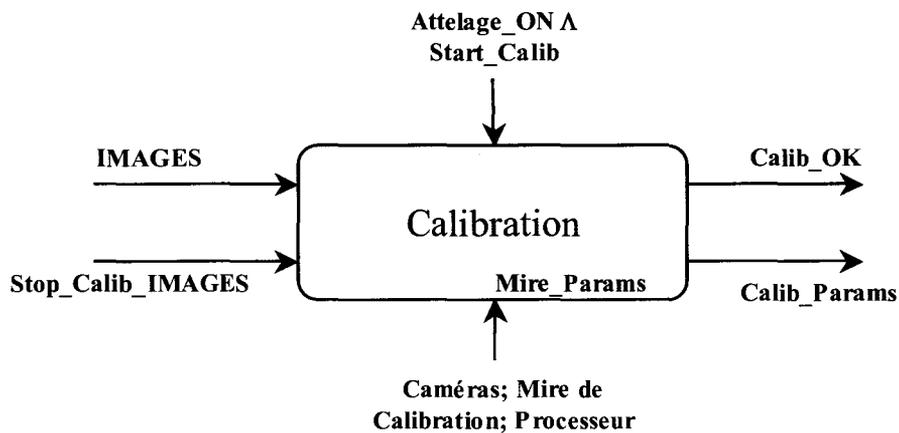


Figure II.4 : Service Calibration.

Entrées : (IMAGES, Stop_Calib_IMAGES)

Sorties : (Calib_Params ; Calib_OK)

Paramètres : Mire_Params

Ressources : Caméras ; Mire de Calibration ; Processeur.

Conditions d'activation : Attelage_ON & Start_Calib

Méthode : « Calibration du capteur stéréoscopique », voir chapitre 3.

Priorité : 1

Temps d'exécution au plus tôt : 0

Périodicité : Non

Préemptabilité : Oui

2. Service Commande_Manuelle

Ce service est le service primordial au véhicule. On supposera d'ailleurs qu'il ne tombe que très rarement en panne. Ce service permet de donner la commande **CMV** en appuyant sur l'accélérateur et/ou sur le frein. Il possède ainsi deux entrées **ACC** et **FREIN** qui indiquent respectivement la position de l'accélérateur et du frein. Ces deux données sont fournies par le même acteur : le conducteur. Ce service n'est pas paramétrable en cours de fonctionnement. Au niveau des conditions d'activation, ce service est « actif » dès que le moteur est en marche. Donc, la condition d'activation est caractérisée par une donnée booléenne unique que nous nommons **M_E_M** qui sera évidemment fournie par le moteur. Les deux ressources essentielles et vitales à ce service sont les pédales d'accélérateur et de frein : **Ped_ACC** et **Ped_FREIN**. D'un point de vue comportemental, ce service est assujéti à la position relative de l'accélérateur et/ou du frein. Il est donc toujours actif tant que le moteur est en marche et ne dépend que du processeur. Il en résulte que ce service n'est ni périodique ni préemptif.

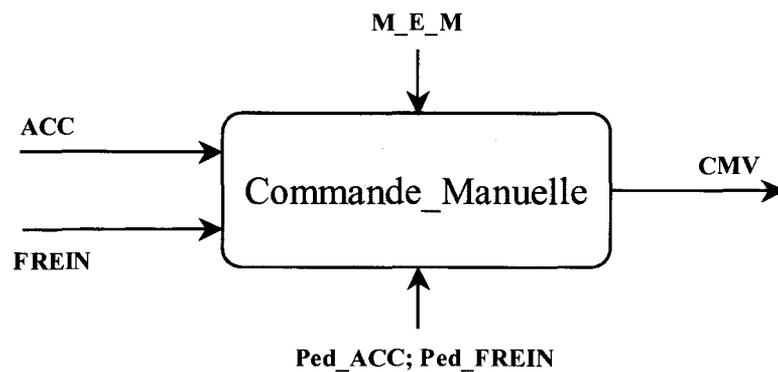


Figure II.5 : Service Commande Manuelle.

Entrées : (ACC, FREIN)

Sorties : (CMV)

Paramètres : \emptyset

Ressources: (Ped_ACC ; Ped_FREIN)

Conditions d'activation : M_E_M

Méthode : « Commande directe par rapport à la position des pédales »

Priorité : 1

Temps d'exécution au plus tôt : 0

Périodicité : Non

Préemptabilité : Non

3. Service Elaboration_Distance

Ce service permet d'élaborer la distance D que le véhicule a parcourue depuis sa mise en service. Il n'a besoin pour son fonctionnement que d'une seule entrée : les données fournies par le capteur de distance (codeur incrémental associé aux roues, par exemple). Sa condition d'activation est que le suivi est en cours de fonctionnement. Cette condition sera symbolisée par la variable booléenne **Suivi_En_Cours_OK**.

Ce service a un comportement non périodique : dans le cas de l'exploitation d'un codeur incrémental la mise à jour de la variable distance D dépend uniquement du nombre d'impulsions prises en compte lorsque le système d'attelage est dans le mode « suivi ». L'absence du capteur de distance sur le véhicule rend le service indisponible et le processeur est la ressource partagée qui sert à mettre à jour la distance parcourue **Distance_Parcourue**.

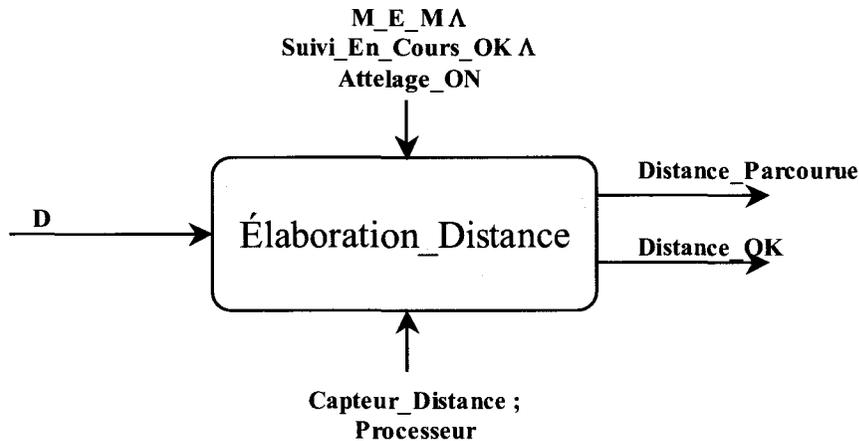


Figure II.6 : Service Elaboration de la Distance

Entrées : (D)

Sorties : (Distance_Parcourue ; Distance_OK)

Paramètres : \emptyset

Ressources : (Capteur_Distance ; Processeur)

Conditions d'activation : M_E_M & Suivi_En_Cours_OK & Attelage_ON

Méthode : « Comptage – Mise à jour »

Priorité : 4

Temps d'exécution au plus tôt : 0

Périodicité : Non

Préemptabilité : Oui

4. Service Elaboration_Distance_Relative

Ce service est un service primordial de la spécification puisque c'est sur lui que repose pratiquement tout le problème de l'attelage virtuel. Ce service a besoin en entrée des coordonnées 2D d'un point caractéristique (au moins) du véhicule tracteur sur les deux caméras « 2D_Entité_Coord_Camera1, 2D_Entité_Coord_Camera2 », et des paramètres de calibration du système de stéréovision issu de la calibration **Calib_Params**. Au niveau des conditions d'activation, on dénombre deux conditions nécessaires, la première est que le système de stéréovision doit avoir été calibré, la seconde est que l'entité suivie doit être également détectée. Ces conditions sont représentées par deux variables booléennes nommées respectivement **Calib_OK** et **Entité_Détectée**. Le service est activé sur la requête **Attelage_ON** au libre choix du conducteur. D'un point de vue comportemental, l'absence des informations issues des caméras implique l'abandon immédiat de ce service. Néanmoins une version dégradée de celui-ci est envisageable si on dispose des connaissances *a priori* sur la cible associée au véhicule tracteur et des images issues d'une seule caméra. Le processeur est ici une ressource partagée vitale.

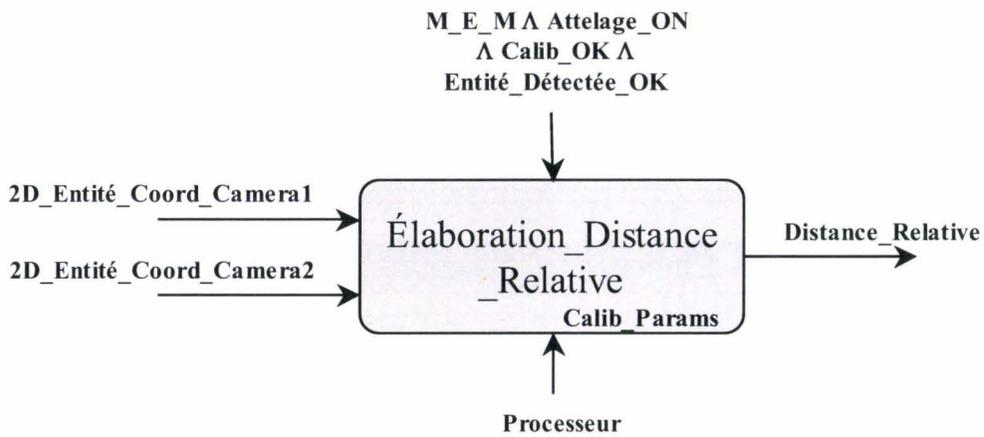


Figure II.7 : Service Elaboration de la Distance Relative

Entrées : (2D_Entité_Coord_Camera1 ; 2D_Entité_Coord_Camera2)

Sorties : (Distance_Relative)

Paramètres : Calib_Params

Ressources : (Processeur)

Conditions d'activation : M_E_M & Attelage_ON & Calib_OK & Entité_DéTECTÉE_OK

Méthode : « Reconstruction tridimensionnelle, Calcul – Mise à jour »

Priorité : 2

Temps d'exécution au plus tôt : 0

Périodicité : Non

Préemptabilité : Oui

5. Service Elaboration_Vitesse_Relative

Ce service est aussi un service primordial de la spécification puisqu'il fournit une information directement liée à la mise en place du suivi. D'un point de vue fonctionnel, le service **Elaboration_Vitesse_Relative** ressemble beaucoup au service **Elaboration_Distance_Relative**, puisqu'ils ont les mêmes ressources et les mêmes conditions d'activation. Ils diffèrent au point de vue des entrées et du comportement.

Le calcul de la vitesse relative représente l'évolution temporelle de la distance relative. La distance relative est la sortie du service précédent. Il faut cependant une base de temps qui sera caractérisée par une date d'estimation « estampille horaire » correspondant à la date à laquelle le véhicule « tracteur se trouvait à la distance estimée (détection de l'entité suivie). Cette « estampille horaire » doit être fournie par l'environnement du système d'attelage virtuel (horloge de l'unité de traitement par exemple).

D'un point de vue comportemental, la procédure d'élaboration de la vitesse relative se fait lorsqu'une nouvelle estimation de la distance relative est disponible. L'absence d'une des ressources nécessaires pour le service de l'élaboration de la distance relative rend

indisponible ce dernier. Le processeur représente la seule ressource vitale puisqu'il permet de mettre à jour la vitesse relative à chaque estimation de la distance relative.

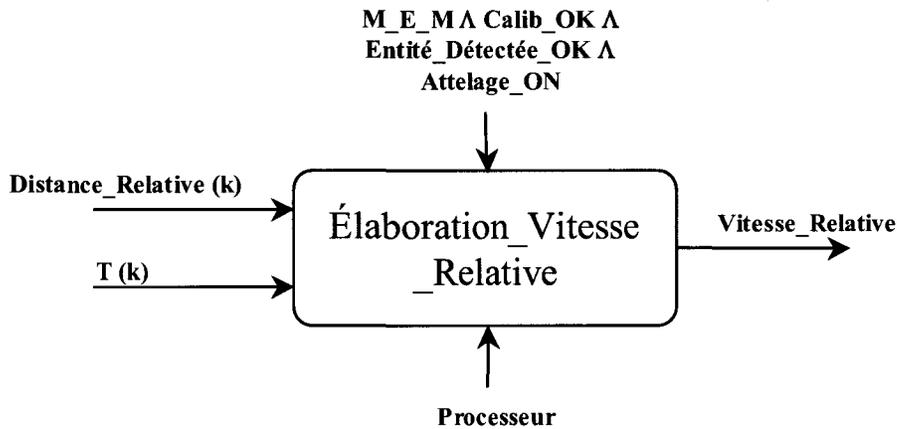


Figure II.8 : Service Elaboration de la Vitesse Relative

Entrées : (Distance_Relative(k) ; T(k))

Sorties : (Vitesse_Relative)

Paramètres : \emptyset

Ressources : (Processeur)

Conditions d'activation : M_E_M & Calib_OK & Entité_Déetectée_OK & Attelage_ON

Méthode : « Calcul – Mise à jour »

Priorité : 2

Temps d'exécution au plus tôt : 0

Périodicité : Oui

Préemptabilité : Oui

6. Service Commande_Automatique

Le service de la commande automatique du véhicule possède des similarités avec le service Commande_Manuelle. Ce service est actif lorsqu'une demande d'attelage a été faite par la requête du conducteur **Attelage_ON** et ceci tant que l'attelage n'est pas désactivé par le conducteur ou à cause de la perte complète de la cible suivie.

Pour donner la consigne de la commande automatique du véhicule (constituée par une consigne de vitesse et une consigne de direction) ce service a besoin de : la vitesse actuelle du véhicule **V**, et de la vitesse, la distance et la direction relatives entre les deux véhicules. Au niveau des conditions d'activation, il existe des conditions d'utilisation nécessaires à l'activation du service : il faut que le moteur soit en marche **M_E_M**, que la pédale de frein ne soit pas pressée (**FREIN = 0**) , la pédale d'accélération non plus (**ACC = 0**) et que le service suivi de l'entité, ait été effectué correctement **Suivi_En_Cours_OK**. Il y a ensuite deux manières d'activer le service : à partir de la requête **Attelage_ON** ou à partir de la

requête **ReEngager** avec la condition supplémentaire que la calibration soit encore valide (**Calib_OK**). Les deux requêtes **Attelage_ON** et **ReEngager** sont toutes deux fournies par le conducteur. De plus, on définit un évènement qui force l'abandon immédiat du service : **Attelage_OFF**.

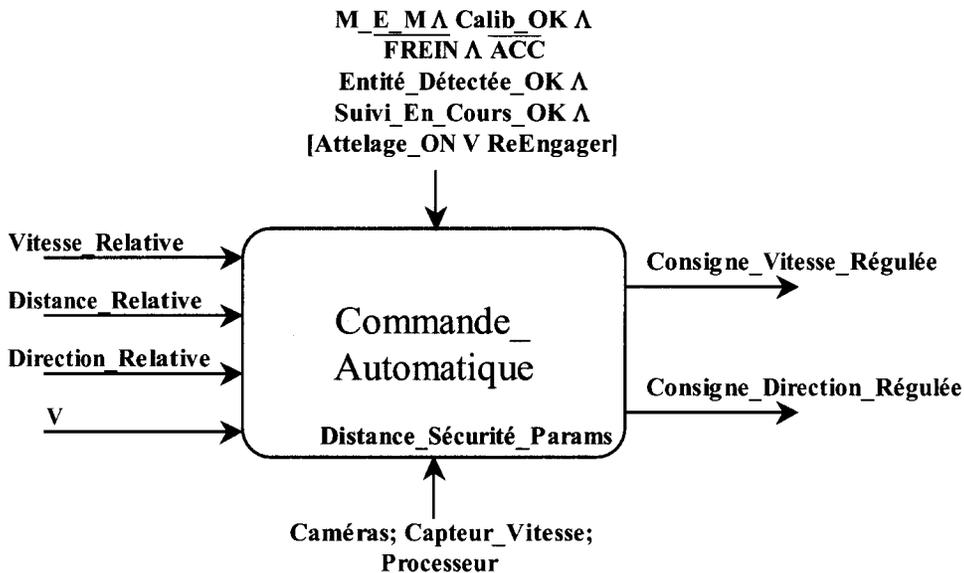


Figure II.9 : Service Elaboration de la Commande Automatique.

Entrées : (Vitesse_Relative ; Distance_Relative ; Direction_Relative ; V)

Sorties : (Consigne_Vitesse_Relative ; Consigne_Direction_Régulée)

Paramètres : Calib_Params

Ressources : (Caméras ; Processeur)

Conditions d'activation : $M_E_M \& \overline{FREIN} \& \overline{ACC} \& Entité_Détectée_OK \& Suivi_En_Cours_OK \& [Attelage_ON \parallel ReEngager]$

Méthode : « Calcul – Mise à jour »

Priorité : 3

Temps d'exécution au plus tôt : 0

Périodicité : Oui

Préemptabilité : Oui

7. Service Elaboration_Vitesse_Moyenne

Ce service propose au conducteur la vitesse moyenne **VM** sur un trajet où il a utilisé le système d'attelage virtuel. Ainsi, ce service a besoin de deux données principales qui sont la distance parcourue **D** et le temps de parcours. Le temps écoulé depuis le début du trajet peut être déduit en utilisant la date et l'heure courante **Date_Heure** ainsi que celle correspondant au début du trajet. Les entrées du service sont alors les données **D** et **Date_Heure**.

Les conditions d'activation du service sont assez simples : il faut que le moteur soit en marche, l'attelage soit en cours, que le service élaboration de la distance soit fonctionnel et que le conducteur demande la vitesse moyenne. La requête de demande de vitesse moyenne est nommée **DVM** son acteur associé étant le conducteur. Le comportement du service est alors asynchrone déclenché sur demande du conducteur. Il a besoin de plusieurs ressources : le processeur qui est une ressource partagée et l'horloge système qui fournit l'objet **Date_Heure** et qui constitue une ressource vitale.

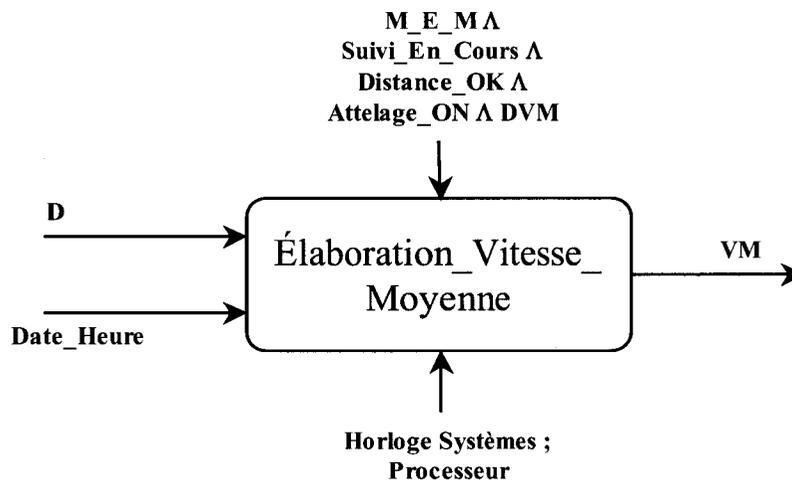


Figure II.10 : Service Elaboration de la Vitesse Moyenne.

Entrées : (D ; Date_Heure)

Sorties : (VM)

Paramètres :

Ressources : (Horloge système ; Processeur)

Conditions d'activation : M_E_M & Suivi_En_Cours & Distance_OK & Attelage_ON & DVM

Méthode : « Calcul – Mise à jour »

Priorité : 4

Temps d'exécution au plus tôt : 0

Périodicité : Non

Préemptabilité : Oui

8. Service Saisie_Vitesse_Maximum

Ce service est associé à la prise en compte de la commande automatique de la vitesse dès l'apparition de la requête **Attelage_ON**. Ce service a pour but de mémoriser la vitesse maximum et de la fournir à son consommateur, le service **Commande_Automatique**.

Le service **Saisie_Vitesse_Maximum** a besoin qu'une seule entrée **V** pour obtenir ses sorties **Consigne_Vitesse_Maximum** et **Consigne_Vitesse_OK** et il n'est pas paramétrable.

Ce service n'enregistre la vitesse maximum que si celle-ci est inférieure à 90 km/h. Le fait que la vitesse soit inférieure à 90 km/h est explicité par la variable booléenne **Consigne_Vitesse_OK**.

En ce qui concerne le comportement, ce service est présent dès et tant que le système d'attelage est mis en marche **Attelage_ON** et influence la vitesse du suivi du véhicule puisque celle-ci ne pourra pas dépasser **Consigne_Vitesse_Maximum**. Pour cette raison, ce service n'est ni périodique ni préemptif par rapport au processeur. Ce service est interrompu dès que **Attelage_OFF** apparaît.

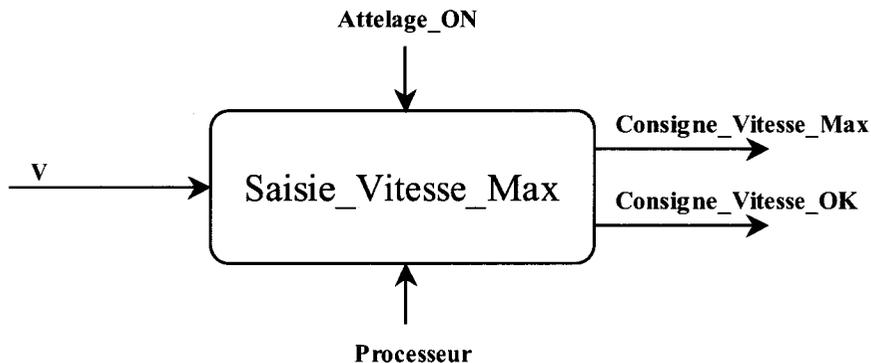


Figure II.11 : Service de la Saisie de la Vitesse Maximum.

Entrées : (V)

Sorties : (Consigne_Vitesse_Max ; Consigne_Vitesse_OK)

Paramètres : \emptyset

Ressources : (Processeur)

Conditions d'activation : Attelage_ON

Méthode : « Enregistrement »

Priorité : 2

Temps d'exécution au plus tôt : 0

Périodicité : Non

Préemptabilité : Non

9. Service Saisie_Distance_Minimum

Ce service comme, celui de la saisie de la vitesse maximum, est associé à la prise en compte de la commande automatique de la vitesse dès l'apparition de la requête **Attelage_ON**. Ce service a pour objectif de mémoriser la distance minimum qui doit séparer le tracteur de son suiveur et de la fournir à son consommateur, le service **Commande_Automatique**.

Le service **Saisie_Distance_Minimum** n'a besoin que d'une seule entrée **D** pour obtenir ses sorties **Distance_Relative_Minimum** et **Distance_Relative_OK** et il n'est pas

paramétrable. Le fait que la distance minimum respecte la distance de sécurité minimum est explicité par la variable booléenne **Distance_Relative_OK**.

De point de vue comportement, ce service est présent dès et tant que le système d'attelage virtuel est en marche **Attelage_ON** et influence **Distance_Relative** puisque celle-ci ne pourra pas être inférieure à **Distance_Minimum**. Pour cette raison, ce service n'est ni périodique ni préemptif par rapport au processeur. Ce service est interrompu dès que **Attelage_OFF** apparaît.

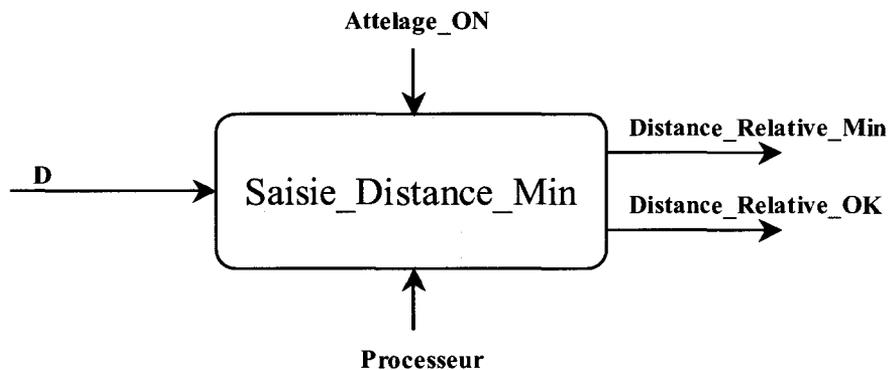


Figure II.12 : Service de la Saisie de la Distance Minimum

Entrées : (D)

Sorties : (Consigne_Relative_Min ; Distance_Relative_OK)

Paramètres : ∅

Ressources : (Processeur)

Conditions d'activation : Attelage_ON

Méthode : « Enregistrement »

Priorité : 2

Temps d'exécution au plus tôt : 0

Périodicité : Non

Préemptabilité : Non

10. Service Elaboration_Direction_Relative

Ce service est un service primordial de la spécification puisqu'il représente, comme le service d'élaboration de la vitesse relative et de la distance relative, un maillon très important dans la chaîne des services que doit fournir le système afin de mener à bien sa tâche principale qui est l'attelage automatique. Ce service a besoin en entrée des coordonnées $2D$ d'un point caractéristique du véhicule tracteur sur les deux caméras « **2D_Entité_Coord_Camera1**, **2D_Entité_Coord_Camera2** », et des paramètres de calibration du système de stéréovision issus de la calibration **Calib_Params**. Au niveau des conditions d'activation, on dénombre deux conditions nécessaires : la première est que le

système de stéréovision doit avoir été calibré, la seconde est que l'entité suivie doit être également détectée. Ces conditions sont présentées par deux variables booléennes nommées respectivement **Calib_OK** et **Entité_Détectée**, le service est activé sur la requête **Attelage_ON** au libre choix du conducteur. D'un point de vue comportemental, l'absence des informations issues des caméras implique l'abandon immédiat de ce service. Le processeur est ici une ressource partagée vitale.

Compte-tenu de la similarité des entrées de ce service et du précédent et de la complémentarité des informations fournies, ces deux services peuvent être regroupés sous le nom d'un même service qu'on appellera la reconstruction 3D.

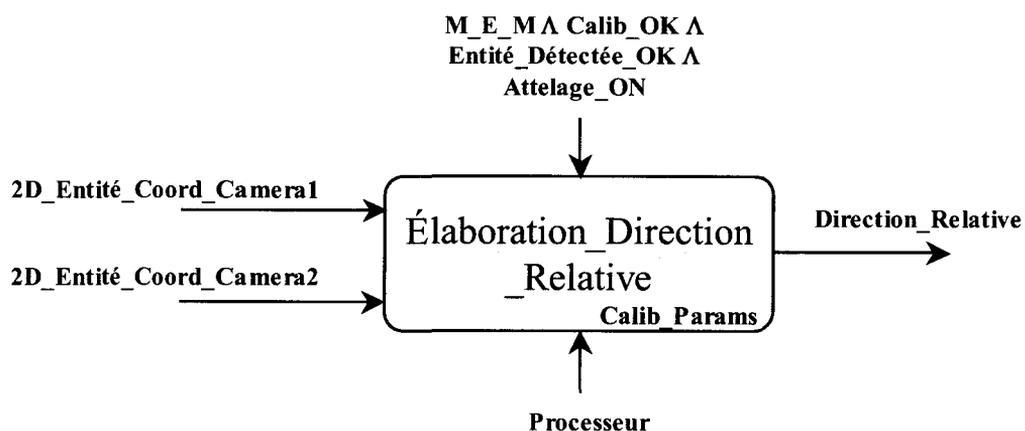


Figure II.13 : Service Elaboration de la Direction Relative.

Entrées : (2D_Entité_Coord_Camera1 ; 2D_Entité_Coord_Camera2)

Sorties : (Distance_Relative)

Paramètres : Calib_Params

Ressources : (Processeur)

Conditions d'activation : M_E_M & Attelage_ON & Calib_OK & Entité_Détectée_OK

Méthode : « Reconstruction tridimensionnelle, Calcul – Mise à jour »

Priorité : 2

Temps d'exécution au plus tôt : 0

Périodicité : Non

Préemptabilité : Oui

11. Service Détection_Entité_Accrochage

Ce service est très important pour la suite des traitements, puisque ses sorties représentées respectivement par : les coordonnées 2D des points caractéristiques de l'entité d'accrochage **2D_Entité_Camera_Coord** et une valeur booléenne indiquant la réussite de l'opération de la détection **Entité_Détectée_OK**, seront consommées par les services

d'élaboration de la distance et de la direction relatives, qui représentent à leur tour le cœur du capteur stéréoscopique dédié à la tâche de l'attelage virtuel.

Le service Détection de l'entité d'accrochage **Détection_Entité_Accrochage** a besoin d'une seule entrée qui est l'**Image** et est activable sous les conditions que le système d'attelage **Attelage_ON** et le moteur **M_E_M** sont tous deux affirmés. Ce service n'exige pas que la calibration du capteur stéréoscopique soit faite puisque les traitements effectués au sein de ce service sont du ressort du traitement d'images pur et non de la vision par ordinateur. Par ailleurs, ce service est paramétrable selon les propriétés géométriques et colorimétriques de l'entité à détecter.

D'un point de vue comportemental, l'absence de l'image issue de la caméra (défaillance de caméra « connectée » au service) implique l'abandon du service. Le processeur est ici une ressource partagée et vitale.

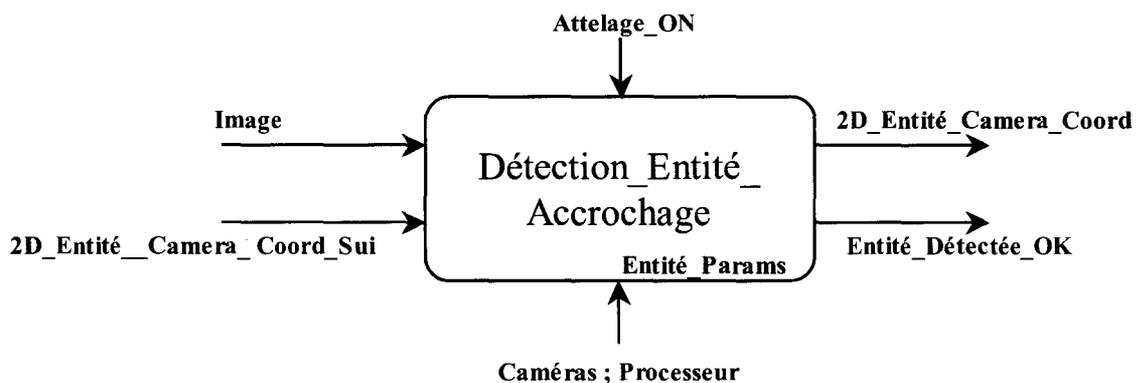


Figure II.14 : Service de Détection de l'Entité d'Accrochage.

Entrées : (Image)

Sorties : (2D_Entité_Camera_Coord ; Entité_DéTECTÉE_OK)

Paramètres : Entité_Params

Ressources : (Caméras ; Processeur)

Conditions d'activation : M_E_M & Attelage_ON

Méthode : « Traitement d'images, Calcul – Mise à jour »

Priorité : 2

Temps d'exécution au plus tôt : 0

Périodicité : Non

Préemptabilité : Oui

12. Service Suivi_Entité_Accrochage

Ce service est aussi très important pour la suite des traitements, puisque ses sorties représentées respectivement par : la prédiction des coordonnées 2D suivantes des points

caractéristiques de l'entité d'accrochage (**2D_Entité_Camera_Coord_Sui**) et une valeur booléenne indiquant que le suivi 2D de l'entité est bien en cours (**Suivi_En_Cours_OK**), seront consommées par les services « Détection de l'entité d'accrochage » et « Elaboration de la vitesse et la distance relatives » ainsi que la commande automatique du véhicule qui sont les services principaux du système de l'attelage virtuel.

Le service Suivi de l'entité d'accrochage **Suivi_Entité_Accrochage** n'a besoin qu'une seule entrée qui est **2D_Entité_Camera_Coord_Act** et nécessite pour son activation que le système d'attelage **Attelage_ON** et le moteur **M_E_M** soient tous deux mis en marche, que la calibration ait bien été effectuée (**Calib_OK**) et que l'entité d'accrochage ait bien été détectée (**Entité_Détectée_OK**). Par ailleurs, ce service est paramétrable selon le modèle du déplacement du véhicule tracteur. Le processeur représente pour ce service une ressource partagée et vitale.

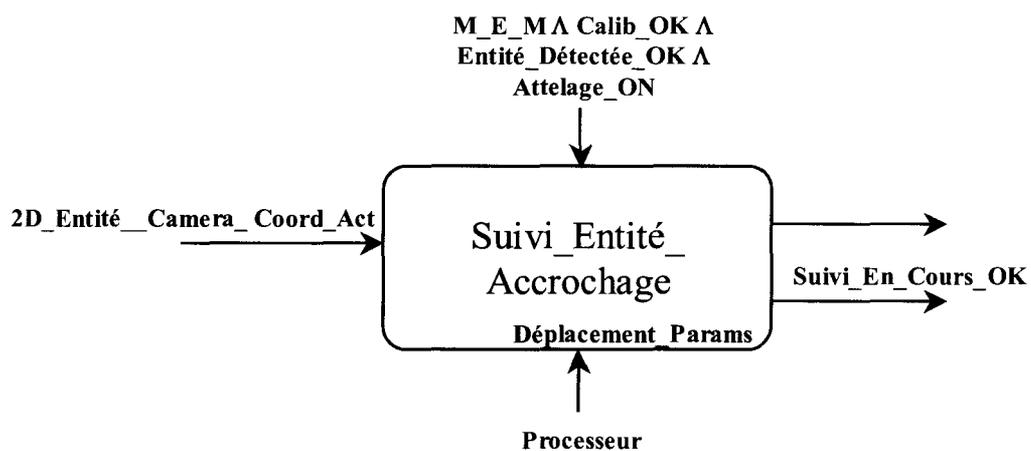


Figure II.15 : Service du Suivi de l'Entité d'Accrochage

Entrées : (2D_Entité_Camera_Coord_Act)

Sorties : (2D_Entité_Camera_Coord_Sui ; Suivi_En_Cours_OK)

Paramètres : Déplacement_Params

Ressources : (Processeur)

Conditions d'activation : M_E_M & Calib_OK & Entité_Détectée_OK & Attelage_ON

Méthode : « Calcul – Mise à jour »

Priorité : 2

Temps d'exécution au plus tôt : 0

Périodicité : Non

Préemptabilité : Oui

II.8.3. ARCHITECTURE OPERATINNELLE DU SYSTEME D'ATTELAGE VIRTUEL

Dans la partie ci-dessus nous avons détaillé les différents services qui constituent un système d'attelage virtuel dans sa globalité et cela sans séparer la partie capteur de la partie actionneur (autrement dit, le pilote comme il a été souligné dans l'introduction). La figure II.16 représente l'architecture opérationnelle complète du système d'attelage virtuel, tout en mettant en évidence la partie capteur de la partie actionneur du système global.

En effet, le schéma ci-dessous fait apparaître justement deux blocs indépendants, le premier a besoin en entrée principalement des images et fourni en sortie, une estimation de la distance intervéhiculaire, le cap et la vitesse relative. Le second bloc quant à lui récupère en entrée les sorties du bloc précédent pour pouvoir générer la commande du véhicule automatisé en l'occurrence le véhicule suiveur. L'interconnexion des deux système tel que l'on peut voir sur la figure II.16 constitue le système final qui a pour but l'attelage virtuel.

Dans le cadre de cette thèse nous nous sommes intéressés particulièrement à la mise en place et le développement de la partie « capteur », représentée par un capteur stéréoscopique dans le but est de percevoir l'environnement du véhicule suiveur et fournir en conséquence, la distance intervéhiculaire, le cap et la vitesse relative qui sont des informations nécessaires pour l'automatisation de la conduite du véhicule, qui est l'objet d'une autre thèse en cours au sein de notre laboratoire (voir, [Tou.04])

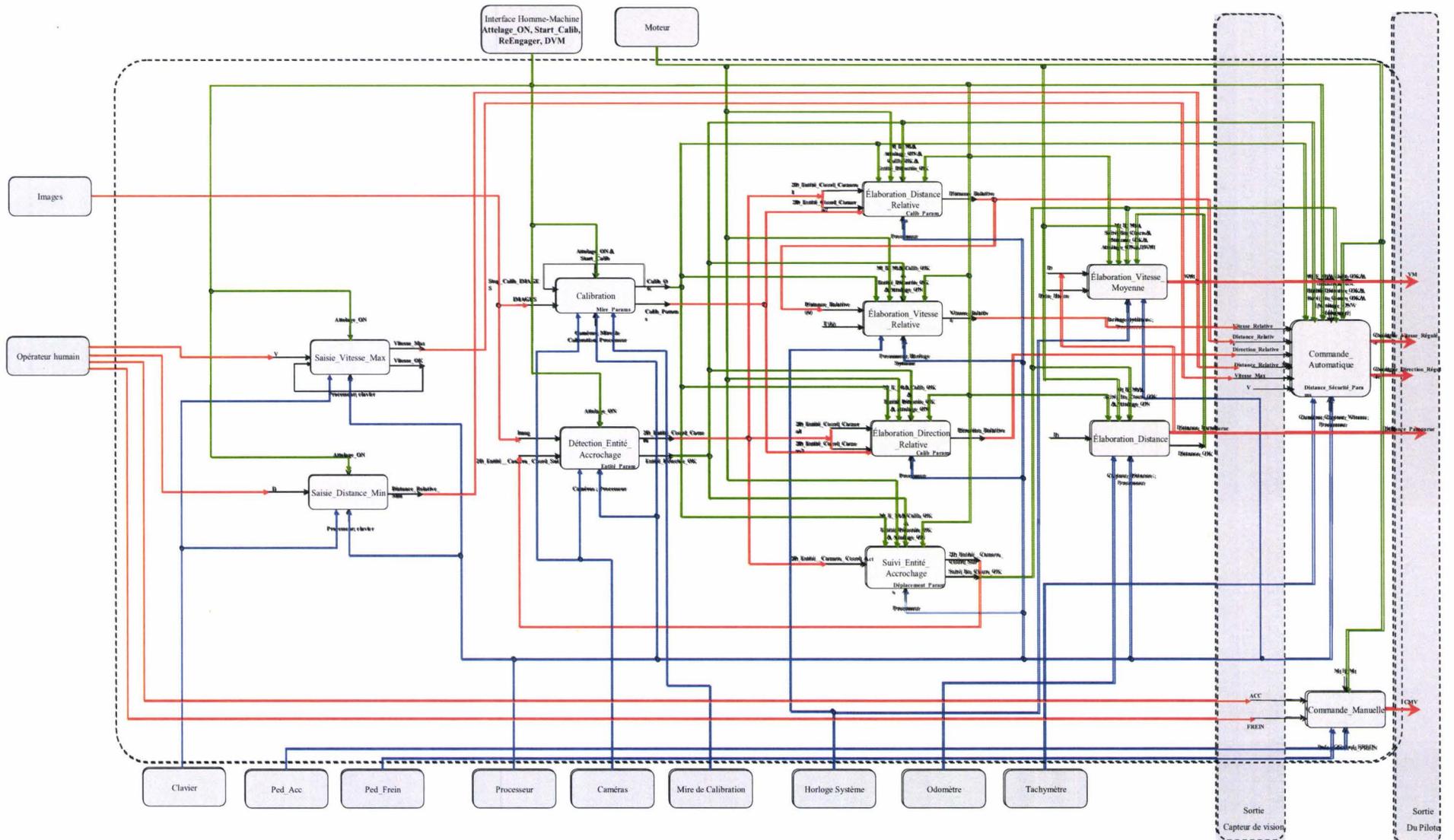


Figure II.16 : Architecture Opérationnelle globale du Système d'attelage virtuel

II.9. CONCLUSION

Dans ce chapitre nous avons donné une vision assez globale de ce qui est un capteur et avons précisé les différentes architectures associées à un système de mesure intelligent. Différentes approches relatives à la réalisation de tels systèmes en adéquation avec leurs fonctions ont été présentées.

Dans un deuxième temps, nous avons abordé la problématique de la modélisation du concept du capteur dit intelligent et les différents domaines d'applications qui lui sont associés. Les différentes approches de conception des capteurs intelligents ont été également introduites.

Enfin, nous avons spécifié les besoins du capteur intelligent dédié à la problématique de l'attelage virtuel, ce qui nous a permis de mettre en évidence les différents services et fonctionnalités que celui-ci se doit de fournir. Cette spécification nous permettra par la suite de mettre en place les traitements spécifiques aux services identifiés dans ce chapitre.

Dans ce qui suit nous allons aborder les problématiques liées à la réalisation des services un à un et de mettre en place des algorithmes des domaines conjoints du traitement d'images et de la vision par ordinateur afin de mener à bien la tâche principale de notre capteur stéréoscopique intelligent, qui consiste à fournir les informations nécessaires pour effectuer l'attelage virtuel.

Le troisième chapitre, a pour rôle de décrire en détail le travail qui en a découlé tout au long de cette thèse, dans le but de mettre en œuvre un capteur stéréoscopique intelligent dédié à la tâche de l'attelage virtuel.

APPLICATION : ATTELAGE

VIRTUEL

Ce chapitre s'intéresse à la réalisation et à la mise en œuvre d'un capteur stéréoscopique intelligent dédié à l'attelage virtuel, se basant uniquement sur les caractéristiques d'un véhicule banal, le véhicule suivi ne nécessitant pas d'être instrumenté. Les résultats obtenus montrent la faisabilité et l'efficacité d'une telle approche.

III.1. INTRODUCTION

Comme nous l'avons évoqué dans le premier chapitre, l'engouement constaté depuis maintenant une vingtaine d'année relativement aux systèmes de transport intelligents (ou I.T.S) trouve sa source aux multiples problèmes que pose le transport des biens et des personnes. Les champs d'investigations fleurissent à mesure qu'apparaissent les technologies disposant du potentiel suffisant pour les explorer. Ainsi, les équipements qui furent parmi les premiers à être expérimentés (caméras infrarouges, radars, capteurs ultrasonores, etc.) furent-ils directement empruntés à des applications militaires extrêmement coûteuses. Cependant, grâce aux progrès remarquables accomplis au niveau des procédés de fabrication de ces matériels, nous assistons actuellement à l'émergence de capteurs et de systèmes de traitement à la fois bon marché et performants [Ber.98].

C'est donc sur la base de ces nouvelles possibilités que se dessine l'avenir de la technologie automobile, en particulier en matière de contrôle automatique, dont les bénéfices attendus sont variés : utilisation plus rationnelle des infrastructures déjà en place, amélioration de la mobilité, minimisation des risques, meilleure maîtrise des durées de transport, économies d'énergie, etc. Le chapitre 1 a été l'occasion de décrire plusieurs approches exploitant l'intelligence artificielle et une très vaste variété de capteurs, dans le but d'aboutir à un véhicule complètement autonome en matière de conduite. Même si les résultats obtenus sont prometteurs, il n'en reste pas moins que les performances de ces systèmes nécessitent d'être encore renforcées, ne serait-ce que parce qu'aucun d'entre eux n'atteint 100 % de temps de conduite en automatique. Tous requièrent donc, à un moment où à un autre (ne pas savoir lequel étant déjà en soi un problème), la présence d'un opérateur humain. Comme nous avons déjà pu le préciser, l'approche « attelage virtuel » développée ici se veut complémentaire plus qu'alternative à la conduite automatique. Ces travaux s'inscrivent dans le cadre du projet de recherche régional AutoRIS (Automatique pour la Route Intelligente et Sûre) et plus exactement au sein du programme AViVA (Attelage Virtuel pour Véhicules

Autonomes). Rappelons que l'objectif d'AViVA est de permettre à terme la formation de « train de véhicules », dont seul celui de tête aurait besoin d'être « activement » conduit. La liaison entre les différents véhicules constitutifs de cet attelage doit être « immatérielle », c'est à dire ne s'appuyant que sur des mesures issues de différents capteurs et des lois de commande appropriées. Les bénéfices possibles de la technique d'attelage virtuel sont nombreux. Pour commencer, cette notion de "trains virtuels routiers" qu'elle autorise pourrait permettre d'associer une "cohérence" de trafic habituellement propre au réseau ferroviaire à une souplesse d'utilisation qui a fait le succès du transport routier. Parmi les points les plus précis que nous pouvons citer, l'attelage virtuel conduit à une optimisation des infrastructures car (potentiellement) les véhicules pourraient se suivre de plus près. Ceci parce qu'entre autres choses, un système automatisé présente par rapport à un opérateur humain un "degré de vigilance" constant. Les améliorations induites peuvent aussi porter sur la consommation (de par l'exploitation de lois de commandes performantes et la structure même de "train"). Un impact positif sur l'environnement est lui aussi à envisager, ne serait-ce que comme retombées de la rationalisation de la gestion de l'infrastructure routière et de la maîtrise de la consommation d'énergie.

Comme nous l'avons expliqué dans l'introduction générale de ce mémoire, la problématique de l'attelage virtuel se ramène à asservir un véhicule en distance et en cap par rapport à celui qui le précède. L'une des originalités de ce travail est de tenter d'y parvenir en prenant la vision pour seule source de données extéroceptives. L'asservissement proprement dit peut se concevoir de deux points de vue différent, à savoir :

- *Un asservissement en situation* : Dans ce cas, les données capteur (ici la vision) servent à positionner la cible (à savoir, le véhicule tracteur) dans l'espace du travail. Cette position relative et les informations qui l'accompagnent (inter – distance, orientation et vitesse) sont utilisées pour calculer la commande à appliquer au véhicule tracté. Cette approche est encore parfois appelée *Asservissement visuel 3D*.
- *Asservissement visuel 2D* : La configuration relative à obtenir entre véhicule tracteur et véhicule tracté est exprimée directement sous la forme de mesures à observer. On précise en fait la façon dont le véhicule tracteur doit être vu. L'originalité de cette approche est qu'elle ne nécessite pas de repasser dans l'espace de travail, les commandes à appliquer au système se déduisant de l'écart entre informations visuelles « de consigne » et informations visuelles observées au travers de la matrice d'interaction associée au système.

Dans la mesure où les informations 3D sont par ailleurs nécessaires, c'est au travers d'une approche de type « Asservissement Visuel 3D » que l'équipe a entrepris de traiter le problème. Cependant, quelle que soit l'approche choisie, la première étape consiste à extraire au sein des informations visuelles des primitives permettant de mener à bien la tâche d'asservissement. C'est à ce problème, ainsi qu'à celui de la localisation 3D à partir des informations visuelles que s'attachent les sections qui vont suivre. Comme nous l'avons indiqué, le « point d'accroche » visuel est ici la plaque d'immatriculation puisque celle-ci est

une entité inhérente aux véhicules. La première partie a pour objectif de préciser les différentes contraintes pesant sur le choix des éléments constitutifs de notre capteur stéréoscopique. Ces contraintes (tant matérielles qu'applicatives) ayant été posées et les composants du système dimensionnés en conséquence, nous présentons ensuite notre méthode de détection de la plaque d'immatriculation. Des résultats expérimentaux montrant l'efficacité de notre algorithme de détection sont présentés. Dans la seconde partie, nous décrivons un filtre de Kalman destiné à effectuer un tracking de la plaque d'immatriculation dans une séquence d'images, ceci afin d'augmenter les performances du système en matière de temps de réponse. La troisième partie de ce chapitre traite de la localisation de la plaque dans l'espace Euclidien tridimensionnel. L'approche monoculaire (une seule caméra) est particulièrement détaillée, compte-tenu de son importance lors du fonctionnement en mode dégradé. Des résultats expérimentaux permettent enfin de dresser un bilan des capacités opérationnelles du système mis en œuvre. Ce chapitre se termine par une brève description des travaux à venir.

III.2. LES CONTRAINTES ET LES EXIGENCES DE L'APPLICATION

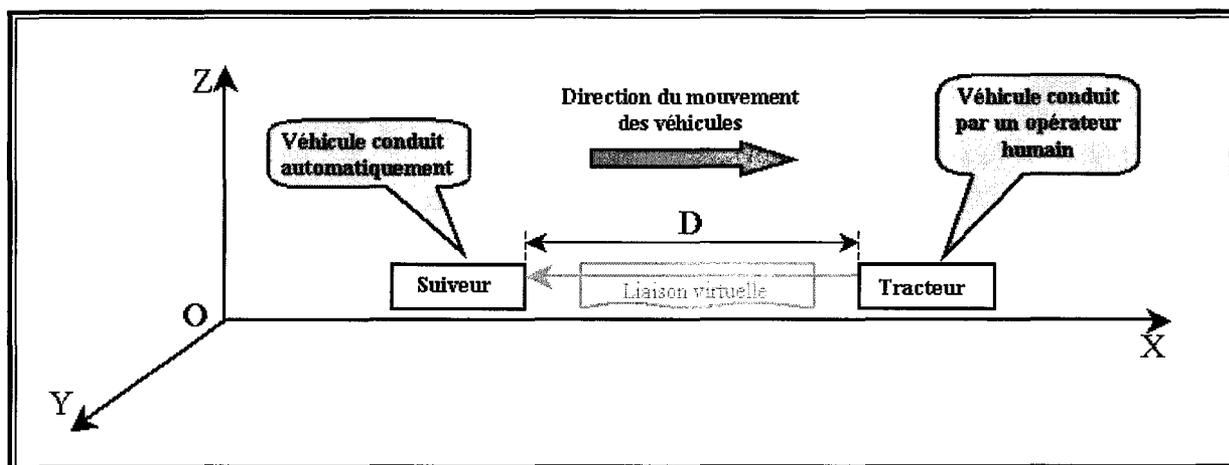
III.2.1. AU NIVEAU DE L'APPLICATION

L'énumération des contraintes pesant sur notre système nécessite que soient rappelés les principes de l'attelage virtuel « basé vision » dans le cas simple où l'attelage ne comporte que deux véhicules. Le terme « basé vision » signifie simplement ici que nous nous basons uniquement sur des propriétés visuelles intrinsèques et sur les positions et vitesses relatives des deux véhicules pour mener à bien la tâche de conduite.

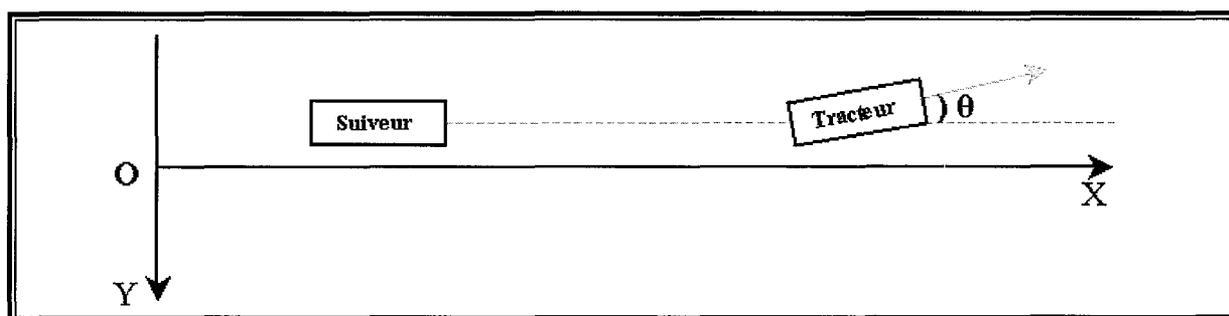
Ainsi, si nous considérons deux véhicules nommés respectivement « Tracteur » et « Suiveur », le véhicule *Tracteur* est conduit par un opérateur humain, tandis que le véhicule *Suiveur* suit automatiquement le véhicule *Tracteur*, comme sous l'emprise d'une liaison mécanique, alors que cette dernière est en fait réalisée au travers de mesures et d'une loi de commande (justifiant l'appellation « d'*Attelage virtuel* »).

Pour que le véhicule suiveur puisse suivre effectivement le véhicule tracteur, celui-ci doit disposer en temps réel des informations suivantes :

- *L'interdistance* entre le véhicule *Tracteur* et le véhicule *Suiveur*.
- *La vitesse relative* entre le véhicule *Tracteur* et le véhicule *Suiveur*.
- *La différence de cap* entre le véhicule *Tracteur* et le véhicule *Suiveur*.



(a)



(b)

Figure III.1 : Les informations nécessaires pour effectuer l'attelage virtuel (a) Schéma descriptif de la tâche de l'attelage virtuel (b) La projection des véhicules, tracteur et suiveur sur le plan OXY

Comme nous l'avons évoqué déjà dans le chapitre 1, afin de pouvoir estimer les informations nécessaires à la tâche d'attelage, les données visuelles utilisées à cet effet doivent présenter certaines propriétés :

- Etre nécessairement présentes dans le cadre de la mission "habituelle" du système, autrement dit, présentes dans la scène observée.
- Etre liées aux véhicule tracteur mais indépendantes des disparités s'appliquant aux véhicules. On peut considérer à ce niveau, tous les éléments visibles de l'arrière du véhicule, à savoir, phares, plaque, calandre, feux arrières, etc.
- Pouvoir être extraites en un temps compatible à la nature temps réel de la tâche, ce qui présuppose une forme géométrique simple ou encore, une couleur particulièrement discriminante par rapport à l'environnement (phares allumés, feux stop ou plaque).

Certes, les propriétés citées ci-dessus sont nécessaires pour toute entité qui doit être utilisée en tant qu'objet décrivant la trajectoire du véhicule tracteur, mais s'avèrent insuffisantes pour des raisons de sécurité. En effet, ces propriétés ne permettent pas de faire la différence entre le véhicule suivi et les autres véhicules. Pour cette raison, la primitive à suivre doit présenter aussi les particularités suivantes :

- Présenter une forme et des caractéristiques suffisamment « robustes » pour être détectable même dans des conditions de fonctionnement perturbées (brouillard « raisonnable » , pluie, etc.).
- Présenter des aspects spécifiques au véhicule à suivre. De ce point de vue, les inscriptions portées par la plaque d'immatriculation sont toutes indiquées puisqu'elles permettent de distinguer le véhicule suivi parmi les autres, même au sein d'un même modèle de même coloris.

L'entité qui correspond le mieux aux exigences précédemment formulées est de loin la plaque d'immatriculation du véhicule tracteur.

Le capteur que nous avons développé a donc pour objectif de localiser en temps réel l'image de la plaque d'immatriculation du véhicule tracteur, de « certifier » ces données à partir des inscriptions portées par la plaque et d'exploiter ces informations pour estimer les grandeurs requises par la loi de commande.

III.2.2. AU NIVEAU MATERIEL

Outre le fait que les choix technologiques doivent permettre l'implémentation des traitements répondant aux besoins de l'application, les matériels choisis doivent par ailleurs respecter un certain nombre de contraintes, à savoir :

- Avoir le moins d'impact possible sur le véhicule suivi (d'où le choix de la vision, comme évoqué précédemment).
- Avoir un coût global ne dépassant pas 10% du coût du véhicule. Notre approche a consisté à exploiter du matériel « grand-public », autant que faire se pouvait (Webcams modifiées, Caméscopes, PC standard).

Ces contraintes ayant été posées, nous passons maintenant à l'analyse fonctionnelle de notre système.

III.3. DECOMPOSITION FONCTIONNELLE DU CAPTEUR STEREOSCOPIQUE DEVELOPPE

Le capteur que nous avons mis en place est constitué d'une unité d'acquisition d'images (deux caméras couleurs facilement interchangeables) et d'une unité de traitement et de communication (un PC standard). La conception modulaire de l'application permet de s'affranchir des disparités présentes entre les différents types de capteurs possibles pour l'information visuelle. Ainsi, à chaque type de caméra est associé une interface logicielle respectant des normes bien définies et fournissant une vision abstraite du matériel aux autres entités impliquées dans le traitement global. A titre d'exemple, il est possible d'utiliser indifféremment des Webcams basées sur les interfaces USB 1.1 ou 2.0 ou des caméras exploitant le protocole IEEE1394 ou tout autre matériel pour lequel un « filtre logiciel » est disponible.

Cette configuration est facilement extensible à l'utilisation d'une carte d'acquisition d'images spécifique, puisque celle-ci s'effectuera juste par l'ajout d'un composant logiciel (un filtre) qui va jouer le rôle d'interface, entre les outils logiciels déjà développés et le matériel en question.

Le déclenchement ou l'arrêt du système se fait sous la responsabilité d'un opérateur humain. Il acquiert et traite les images en temps-réel (temps de réponse du système inférieur à 100 ms) afin d'extraire les paramètres nécessaires à la loi de commande, à savoir : l'interdistance ainsi que la vitesse et les orientations relatives. Après une phase de calibration au démarrage du système, la série de traitements typiquement appliqués à chaque image est :

- Acquisition
- Extraction de zones d'intérêt (zones de l'image où la présence de la plaque est probable, compte-tenu des caractéristiques de celle-ci, des mouvements relatifs estimés du véhicule tracteur et des mouvements propres du véhicule suiveur).
- Extraction et classification des caractères de la plaque d'immatriculation.
- Mise en correspondance et calcul des coordonnées 3D des points caractéristiques de la plaque, permettant de calculer la position relative du véhicule tracteur par rapport au véhicule tracté. La figure III.2 présente une décomposition fonctionnelle du capteur stéréoscopique développé. D'après ce qui précède, son fonctionnement peut être scindé en trois opérations distinctes, qui sont :

- **L'Acquisition** des images gauches et droites.
- **La Calibration** du capteur.
- **Les Traitements** spécifiques à l'application.

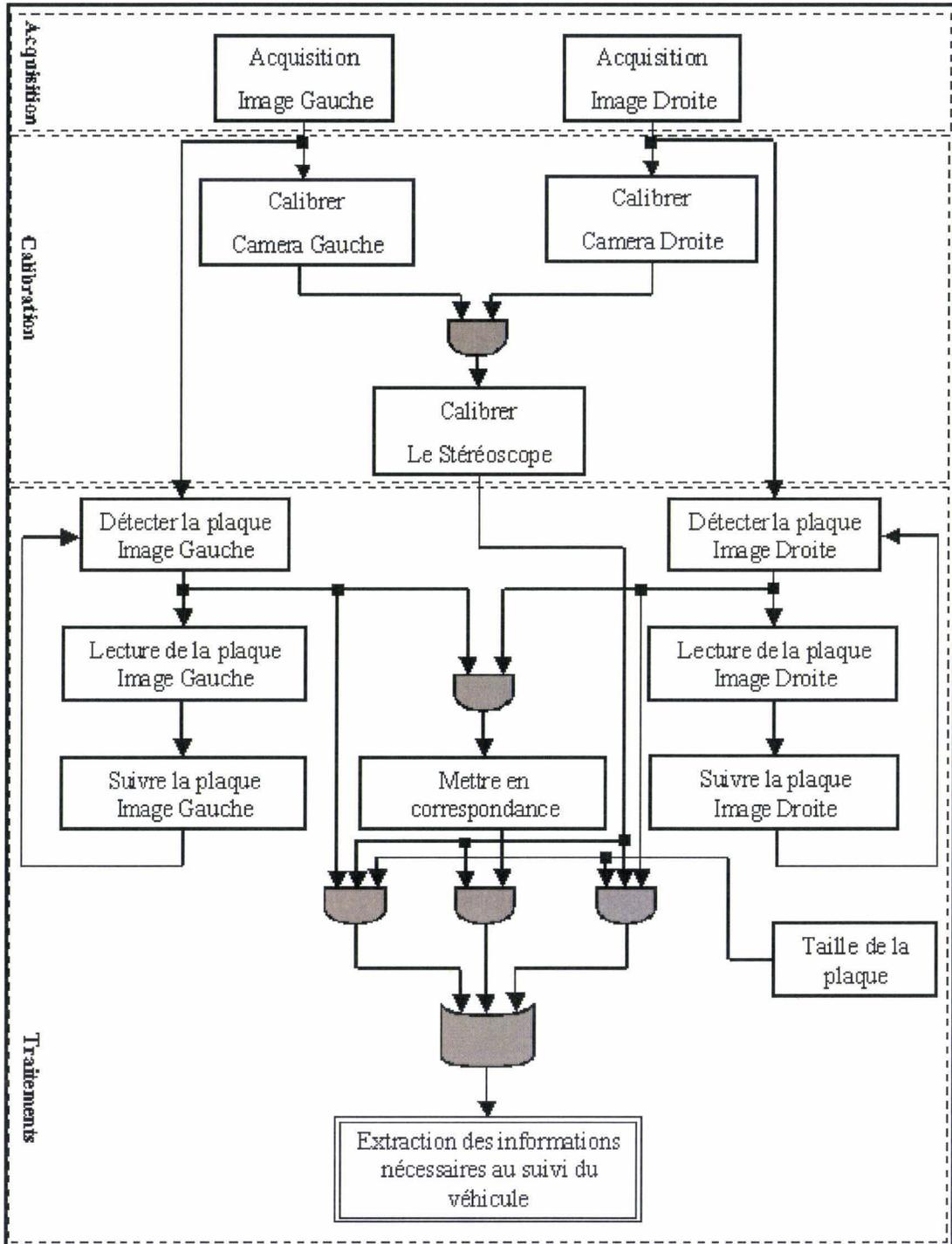


Figure III.2 : Décomposition fonctionnelle du capteur stéréoscopique mis en place.

Les sections qui suivent détaillent les opérations ainsi réalisées et justifient les choix effectués en matière d’implantation.

III.3.1. ACQUISITION

L’acquisition des images est l’opération consistant à transformer l’énergie photonique, issue de la partie du monde réel observée par le dispositif, en une énergie électrique qui (après quantification) sera stockée sous forme d’images et accessible *via* le programme

d'application. Le module qui permet d'effectuer l'opération d'acquisition d'images peut être décomposé en plusieurs sections : l'optique, le capteur, le transfert des données, les entrées associées à l'ordinateur (Figure III.3).

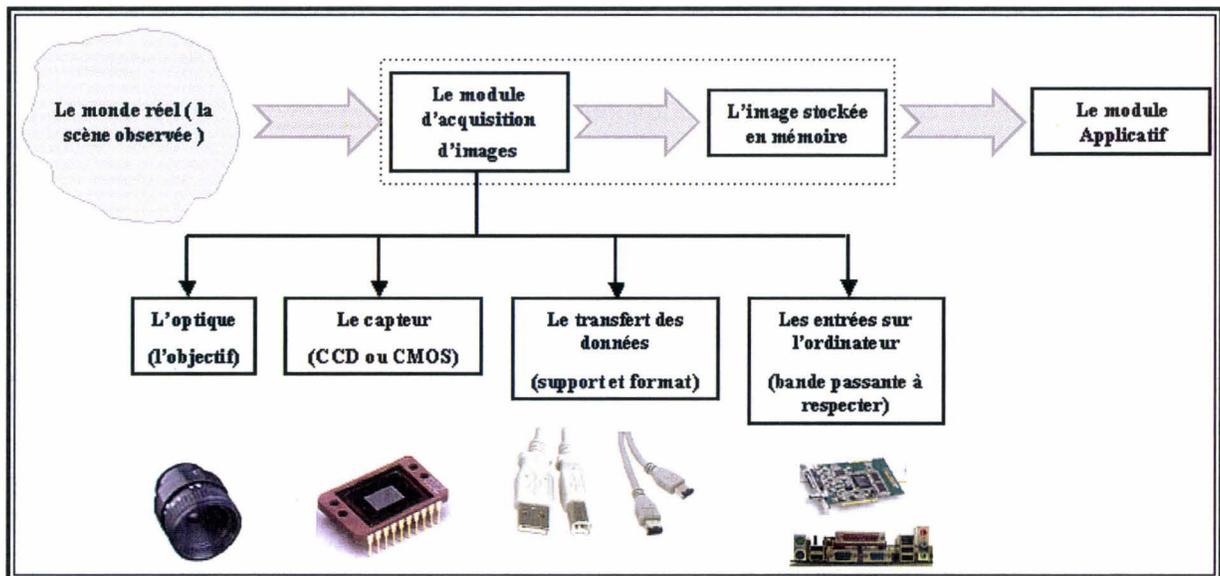


Figure III.3 : Décomposition du module d'acquisition d'images.

Dans le cadre de cette thèse, nous nous sommes intéressés exclusivement aux capteurs de type CCD et CMOS.

III.3.1.1. LES DISPOSITIFS D'ACQUISITION D'IMAGES

L'utilisation des caméras CCD dans le domaine de la vision par ordinateur (pour les mesures tridimensionnelles) est prédominante. Cependant, en raison de certaines particularités (par exemple la possibilité d'un accès individuel aux pixels de la matrice, la courbe caractéristique du capteur en lui-même) des capteurs CMOS de la même résolution ont été employés dans d'autres domaines (Interférométrie, holographie digitale, ...). Les capteurs CMOS actuels sont-ils applicables à des systèmes stéréoscopiques à vocation généraliste ?

Afin d'apporter des éléments de réponse, la section suivante dresse un comparatif entre les caractéristiques principales de l'une et l'autre de ces deux technologies et décrit leur influence sur les mesures issues de la vision [Zay.04a].

A) TECHNOLOGIE CMOS CONTRE TECHNOLOGIE CCD

Qu'ils s'agisse des capteurs CCD ou CMOS, tous deux sont fabriqués dans des fonderies de silicium et sont constitués à la base des mêmes matériaux (le silicium, l'oxyde de silicone, le polysilicium, etc.). Cependant les capteurs CCD profitent de trois décennies d'expérience qui ont permis de développer des processus optimisés pour les dispositifs d'acquisition d'images. En conséquence, ces capteurs présentent aujourd'hui d'excellentes caractéristiques grâce à des taux de bruits extrêmement bas, des « courants d'obscurité » très

faibles, un facteur de rendement et un facteur de remplissage très élevés [Bla.01]. La suite fournit quelques explications sur la signification de ces différents paramètres qui constituent les aspects techniques guidant le choix d'un capteur spécifique.

1) LES ASPECTS TECHNIQUES

Dans ce qui suit, nous détaillons donc la signification de paramètres pertinents dans la caractérisation d'un capteur destiné à l'acquisition d'images.

a) Le courant d'obscurité et le bruit

Le courant d'obscurité résulte principalement de l'énergie thermique répartie dans le treillis de silicium. Ainsi, indépendamment de la lumière projetée sur les capteurs, des porteurs de charge négatifs apparaissent aléatoirement au fil du temps [Mag.03]. L'occurrence, tant spatiale que temporelle, de ce phénomène connu sous le nom du "bruit du courant d'obscurité" peut être décrite en termes statistiques. Les fluctuations qu'il implique sont par nature présente même en l'absence de lumière. En pratique, ce « courant d'obscurité » ne présente un réel problème que pour les temps d'exposition relativement longs (de l'ordre de quelques secondes).

b) Le bruit du modèle

Au contraire des capteurs CCD, les capteurs CMOS sont des capteurs à pixels actifs où chaque pixel possède son propre nœud de détection et son propre amplificateur [Jan.02a, Jan.02b, Eas.00]. À cause de la dispersion paramétrique affectant la fabrication de chacun de ces composants pendant le processus industriel, les valeurs des niveaux du gris varient du pixel en pixel, même sous un éclairage homogène. Ce "bruit de modèle" correspondant au capteur peut être mesuré une fois pour toutes pour les différentes valeurs possibles des niveaux du gris. Ainsi, ces données pourront être stockées par exemple dans une matrice et pourront ensuite être employées pour la correction des intensités des pixels [Hel.03]. En outre, la réduction du bruit du modèle est possible par la méthode dite de « corrélation du double échantillon » dont on trouve les détails dans [Jan.02b, Kav.00].

c) L'efficacité quantique

L'efficacité quantique (EQ) est la mesure de l'efficacité d'un capteur à produire une charge électronique en fonction de l'énergie lumineuse incidente.

L'efficacité quantique est définie physiquement par un pourcentage représentant le nombre d'électrons emmagasinés sur le nombre des photons incidents sur une zone du capteur totalement sensible à la lumière. Pour les capteurs CCD et CMOS, l'efficacité quantique est excellente sur l'ensemble du domaine visible (400 - 700 nm)[Bla.01]. Cependant, les capteurs CMOS sont généralement moins sensibles dans le domaine du bleu (300 - 400 nm) que leurs homologues CCD.

d) Le facteur de remplissage

Le facteur de remplissage est défini comme étant la proportion de la partie sensible à la lumière par rapport à la surface totale du pixel, et détermine aussi la sensibilité maximale. En raison de leur architecture et de leur mode d'exploitation, le facteur de remplissage des capteurs CCD à transfert de trame complète est proche de 100 %. En revanche, le facteur de remplissage des capteurs CCD à transfert de trames entrelacées et des capteurs CMOS est de l'ordre de 30%. Il convient d'attirer l'attention du lecteur à ce niveau qu'il est possible d'augmenter le facteur de remplissage des capteurs CMOS en diminuant la surface occupée par le transistor sur chaque pixel.

e) La consommation d'énergie

Les caméras CCD nécessitent de nombreux circuits additionnels (par exemple des conditionneurs de signaux, des amplificateurs, etc.) « autour » du capteur proprement dit. Par contre, la technologie standard des capteurs CMOS permet l'intégration de différentes fonctions (horloge, contrôle des temps d'exposition, etc.) sur la même puce de silicium. Par conséquent, les capteurs CMOS consomment beaucoup moins d'énergie (au moins 10 fois moins) que leurs homologues utilisant des matrices CCD.

2) LES ASPECTS ECONOMIQUES

En raison de la large utilisation de la technologie CMOS pour les nombreux types de circuits intégrés, ces puces profitent de l'industrialisation à grande échelle qui diminue considérablement la densité des défauts et les dépenses de production. Des rendements plus importants et des taux de rebut très faibles font de la technologie CMOS une technologie moins coûteuse que celle associée aux capteurs CCD pour l'acquisition des images. Des composants moins nombreux, une taille plus réduite et une fiabilité plus importante du produit final constituent des sources d'économies non négligeables pour les fabricants et les acquéreurs de ce type de dispositifs.

Ainsi, en dépit des indéniables performances que présentent les matrices CCD dans le domaine de l'acquisition d'image, il est clair que les capteurs CMOS exhibent des avantages intrinsèques d'une grande importance : une basse consommation électrique, capacité d'intégration, accès aléatoire aux pixels, etc.

Par ailleurs, la technologie CMOS profite, elle aussi, des travaux qui ont été menés initialement dans le domaine des capteurs CCD. Ainsi, plusieurs techniques sont maintenant appliquées à la technologie CMOS, comme par exemple les micro-lentilles, permettant de pallier le faible facteur de remplissage [Lab.03].

Le tableau ci-dessous récapitule les points forts et les points faibles de chacune de ces deux technologies de capteurs dans le domaine de l'acquisition d'images :

Technologies Caractéristiques	La technologie CCD	La technologie CMOS
Le courant d'obscurité et le bruit	présent	présent
Le bruit du modèle	Moins sensible car impliquant moins de composants	Un problème intrinsèque à la technologie CMOS.
L'efficacité quantique	Excellente	Excellente, comparativement moindre dans le bleu
Le facteur de remplissage	Proche de 100 % pour transfert de trame complète	Relativement faible (30%) compte tenu des éléments associés au pixel lui même (ampli).
Consommation d'énergie	Consommation relativement importante des indispensables circuits périphériques.	Consommation très faible du à l'intégration de toutes les fonctions sur le même silicium.
Le coût	Coût relativement élevé.	Une technologie « low cost » profitant d'économies d'échelle.

Tableau III.1 : Récapitulation des caractéristiques des technologies CCD et CMOS.

Comme indiqué, le but de l'acquisition d'images est donc de convertir une vue du monde réel en une image numérique. Une image numérique est représentée par un ensemble de points de coordonnées (x, y) et d'intensité lumineuse v . Le passage d'un signal analogique à un signal numérique évoque incontestablement les deux opérations clefs dans le traitement du signal que sont l'échantillonnage et la quantification [Mat.00].

La bande passante du système est limitée par la résolution des images utilisées d'une part et la période d'acquisition des images d'autre part. Il est par conséquent nécessaire de restreindre le domaine fréquentiel des phénomènes observés de sorte à pouvoir conserver un compromis acceptable entre résolution « spatiale » et résolution « temporelle ». L'opération de quantification consiste à diviser la dynamique du signal analogique en un certain nombre d'intervalles et à affecter la même valeur à toute intensité contenue dans un intervalle donné. Généralement, la grille de quantification est « linéaire », c'est à dire que la largeur des intervalles est constante et le nombre d'intervalles est codé sur k bits (généralement $k = 8$ ou 12). A noter qu'une image au sens « informatique » du terme a subi deux opérations de quantification, à savoir une quantification au niveau de chaque pixel, mais aussi une quantification « spatiale », induite par l'utilisation d'une matrice CCD ou CMOS.

Dans ce qui suit, nous allons aborder quelques concepts clefs de l'acquisition d'images :

III.3.1.2. LA TAILLE DU CAPTEUR

La taille des capteurs du type CCD ou CMOS est habituellement spécifiée en « pouce ». On parle de « quart de pouce », de « demi-pouce », etc. A vrai dire, ces dimensions ne correspondent à rien du point de vue physique. Il s'agit en fait d'équivalence avec la diagonale des capteurs des caméras à tube [Mat.00]. Le tableau suivant donne les dimensions du capteur en fonction de sa dénomination.

Taille du capteur	Horizontale (mm)	Verticale (mm)	Diagonale (mm)
1/6"	2.4	1.8	3.0
1/4"	3.2	2.4	4.0
1/3"	4.8	3.6	6.0
1/2"	6.4	4.8	8.0
2/3"	8.8	6.6	11.0
1"	12.6	9.5	15.8

Tableau III.2 : Tailles des capteurs CCD ou CMOS

La taille du capteur (ou plus exactement ses dimensions) permettent de calculer le champ de vision d'une caméra en fonction de la focale de l'objectif utilisé.

III.3.1.3 LE MODELE GEOMETRIQUE DE LA CAMERA

Dans notre contexte, un modèle est une formulation mathématique qui rend compte du comportement d'un dispositif physique donné en utilisant un ensemble d'équations. La modélisation d'une caméra consiste ici à exprimer le lien entre les points de l'image formée par la caméra avec les points de la scène observée, ceci en fonction de paramètres inhérents à la caméra elle-même (paramètres intrinsèques) et de la configuration relative caméra / scène (paramètres extrinsèques). Plusieurs modèles de caméras peuvent être adoptés selon la précision requise [Ito.91]. Les plus simples sont basés sur des transformations linéaires et ne tiennent pas compte des distorsions dues aux imperfections des objectifs. A l'inverse, certains modèles non-linéaires modélisent les lentilles des caméras avec précision [Wen.92].

Le modèle de caméra le plus simple est celui proposé par Hall [Hal.82]. L'idée principale de ce modèle est de trouver une relation entre tout point 3D de la scène et sa projection 2D sur le plan rétine de la caméra. Cette relation peut être exprimée sous forme matricielle de la façon suivante:

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix} \begin{pmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{pmatrix} \quad (3.1)$$

avec : $(X_m \ Y_m \ Z_m \ 1)^T$ les coordonnées homogènes d'un point de la scène **dans le repère 3D du monde** et $(su \ sv \ s)^T$ les coordonnées homogènes de la projection de ce point sur le plan image, exprimées **par rapport au repère image**.

Cette relation permet donc d'associer tout point P_m de la scène, défini grâce à ses coordonnées dans le repère 3D du monde, à sa projection P_i exprimée en coordonnées pixel dans le repère image. La matrice 3×4 exprimant cette relation est fonction de paramètres intrinsèques, liés à la géométrie propre de la caméra, et de paramètres extrinsèques exprimant la relation entre la position de la caméra et le repère de la scène (appelé aussi repère du monde).

De nombreux chercheurs se sont penchés sur la modélisation des caméras. Nous pouvons citer Aloimonos, qui a défini une hiérarchie des modèles de caméras, incluant même les para-perspectives [Alo.90]. Mundy et Zisserman [Mun.92] ont généralisé cette approche, au travers de la caméra affine. Faugeras a mis en évidence les propriétés de la caméra projective [Fau.93]. Plus de détails concernant le modèle « linear pushbroom camera » peuvent être retrouvés dans [Gup.97], et la caméra 2D est expliquée dans [Qua.97b].

Le modèle géométrique de la caméra le plus utilisé est le *modèle à Sténopée*. Celui-ci modélise une caméra finie, c'est à dire une caméra dont le centre de projection est bien déterminé dans l'espace Euclidien, contrairement à la caméra affine (son centre de projection est à l'infini). Pour plus d'information voir [Har. 03]. C'est ce modèle que nous allons détailler dans ce qui suit.

A) LE MODELE A STENOPEE

La caméra est un outil de projection de l'espace 3D (l'espace de l'objet) vers l'espace image 2D. Autrement dit, cet outil nous permet de « résumer » l'information contenue dans l'espace 3D dans un espace 2D représenté par l'image, au prix bien entendu de certaines pertes.

Les entités géométriques constitutives du modèle à Sténopée le plus simple sont :

- Un centre de projection C des points de l'espace vers un plan. Ce centre de projection est aussi considéré comme l'origine du repère euclidien.
- Un plan à une distance $Z = f$, qui est appelé aussi le *plan image* ou le *plan focal*.

- Un système d'axes défini par rapport au centre de projection et au plan image. Les deux premiers vecteurs de base sont en effet tel qu'avec le point C , ils définissent un plan parallèle au plan image. Le trièdre associé à ce repère est direct.

Dans le cadre du modèle à Sténopée, la projection p d'un point $P(X\ Y\ Z)$ de l'espace sur le plan image I est définie par $\{p\} = (CP) \cap I$ (voir Figure III.4). Par triangulation, nous pouvons déduire facilement les coordonnées de p en fonction des coordonnées 3D de P :

$$(X\ Y\ Z) \mapsto (fX/Z\ fY/Z) \quad (3.2)$$

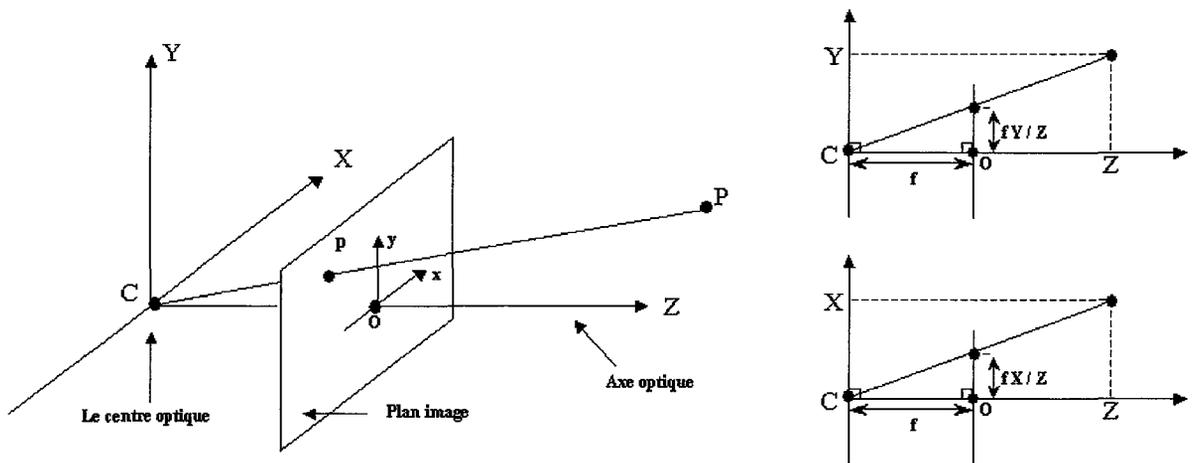


Figure III.4 : Géométrie du modèle à Sténopée de la caméra. C est le centre optique et O le point principal.

Cette transformation est une projection de l'espace Euclidien tridimensionnel \mathcal{R}^3 vers l'espace bidimensionnel \mathcal{R}^2 . Dans ce cas, le centre de projection est appelé *centre de la caméra* ou encore le *centre optique*. La ligne passant par le centre optique et perpendiculaire au plan image est appelée l'*axe optique* de la caméra, et le point où l'axe optique coupe le plan image est appelé, le *point principal*. Par ailleurs, le plan parallèle au plan image, et contenant le centre de projection est communément appelé le *plan principal*.

La relation établie précédemment peut être exprimée linéairement en exploitant les coordonnées homogènes du point de l'espace et de sa projection. C'est ainsi que nous aboutissons facilement à :

$$\begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (3.3)$$

Avec $(X \ Y \ Z \ 1)^T$ les coordonnées homogènes du point P de l'espace Euclidien tridimensionnel lié à la caméra et $(fX \ fY \ Z)^T$ les coordonnées homogènes de sa projection p sur le plan rétiné de la caméra, autrement dit, l'espace image 2D. En notant Pr la matrice de projection de la caméra, nous pouvons écrire le système d'équations (3.3) sous forme condensée :

$$p = Pr P \quad (3.4)$$

Les équations proposées jusqu'à présent sont encore un peu trop simples pour être réalistes. En effet, l'équation (3.4) suppose que l'origine des coordonnées de l'image se trouve au point principal (la projection du centre optique sur le plan de l'image). En pratique, l'origine des coordonnées de l'image se situe dans le coin supérieur gauche de l'image (Figure III.5):

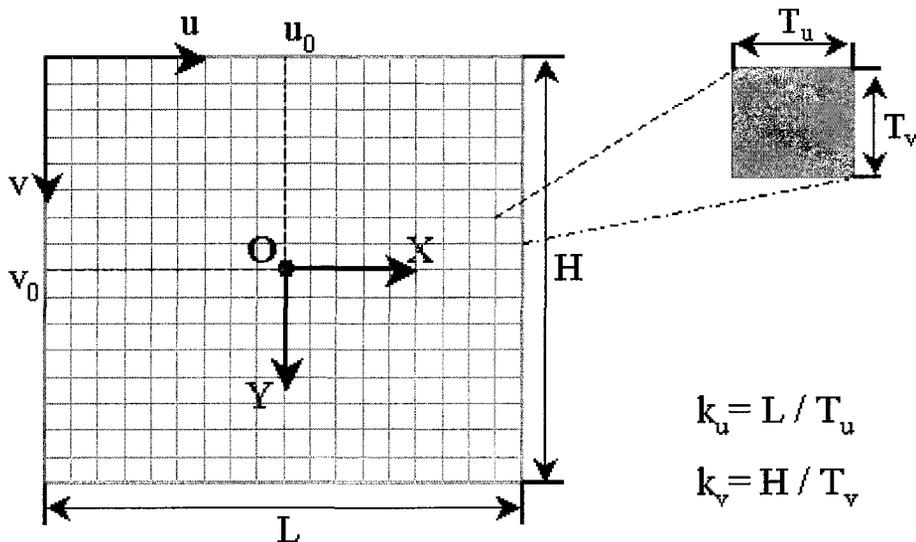


Figure III.5 : Les coordonnées image (u, v) et les coordonnées caméra (X, Y)

Pour que nous puissions effectuer le changement de repère approprié, nous devons introduire les coordonnées $(u_0 \ v_0 \ w_0)^T$ du point principal par rapport au repère de l'image, ainsi les coordonnées du point p dans le repère image s'exprimeront de la manière suivante :

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} f & 0 & u_0 & 0 \\ 0 & f & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (3.5)$$

Cependant, le système d'équations (3.5) suppose que les coordonnées image sont des coordonnées euclidiennes qui ont le même facteur d'échelle sur les deux directions. Or, dans le cas des capteurs CCD ou CMOS, les pixels de la caméra sont rarement carrés. Si les coordonnées image sont mesurées en pixels, nous serons obligés d'introduire des facteurs d'échelle différents pour chaque direction. Ainsi, le système d'équations (3.5) s'écrira :

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} f k_u & 0 & u_0 & 0 \\ 0 & f k_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (3.6)$$

Où, k_u et k_v représentent respectivement les facteurs d'échelle sur l'axe des abscisses et des ordonnées. Ils représentent physiquement la « densité en pixel » dans une direction donnée (verticale ou horizontale).

Le système d'équations (3.6) exprime les coordonnées homogènes 2D d'un pixel sur l'image en fonction des coordonnées homogènes du point 3D appartenant à la scène observée, par rapport au repère de la caméra.

1) LES PARAMETRES INTRINSEQUES

En développant le système d'équations (3.6), nous pouvons écrire les équations du modèle géométrique de la caméra, soit la relation entre les coordonnées caméra $(X \ Y \ Z)^T$ du point P et les coordonnées image $(u \ v)^T$ du point p (sa projection):

$$\begin{cases} u = u_0 + k_u f \frac{X}{Z} \\ v = v_0 + k_v f \frac{Y}{Z} \end{cases} \quad (3.7)$$

En effet la matrice :

$$I_c = \begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Représente une application linéaire de l'espace projectif vers le plan projectif exprimant la transformation perspective :

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = I_c \begin{pmatrix} X \\ Y \\ Z \\ s \end{pmatrix}$$

Ce modèle comporte quatre paramètres : $\alpha_u = k_u f$, $\alpha_v = k_v f$, u_0 et v_0 . Ce sont ces paramètres qui vont être déterminés *via* la phase de *calibration*. Notons que la distance focale ne peut être calculée explicitement.

En introduisant maintenant des coordonnées caméra sans dimension telles que :

$$X_c = X/Z$$

$$Y_c = Y/Z$$

$$Z_c = 1$$

Nous pouvons écrire explicitement, la relation entre les coordonnées image et les coordonnées caméra :

$$\begin{cases} u = \alpha_u X_c + u_0 \\ v = \alpha_v Y_c + v_0 \end{cases} \quad (3.8)$$

ou sous forme matricielle :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = C \begin{pmatrix} X_c \\ Y_c \\ 1 \end{pmatrix} \text{ avec } C = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

La matrice C est appelée matrice des paramètres intrinsèques de la caméra et représente une transformation projective qui permet de calculer les coordonnées image d'une projection en sachant les coordonnées du point 3D de la scène par rapport au *repère Euclidien de la caméra* (rappelons en effet que, jusqu'à présent, nous avons supposé que celui-ci était aussi le repère du monde).

2) LES PARAMETRES EXTRINSEQUES

L'hypothèse de confusion entre le repères du monde et le repère de la caméra ne peut plus être faite dès lors que sont envisagées des applications où cette dernière est mobile. Un modèle réaliste exige donc la prise en compte de paramètres extrinsèques, permettant d'exprimer la transformation euclidienne entre le repère du monde (repère de la scène

observée) et le repère de la caméra (voir Figure III.6). Cette transformation peut être décomposée en une matrice de rotation R et un vecteur de translation T :

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \underbrace{\begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}}_R \begin{pmatrix} X_m \\ Y_m \\ Z_m \end{pmatrix} + \underbrace{\begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}}_T \quad (3.9)$$

La transformation rigide correspondante (rotation et translation) peut s'écrire sous la forme d'une translation homogène [Hor.96] :

$$A = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \quad (3.10)$$

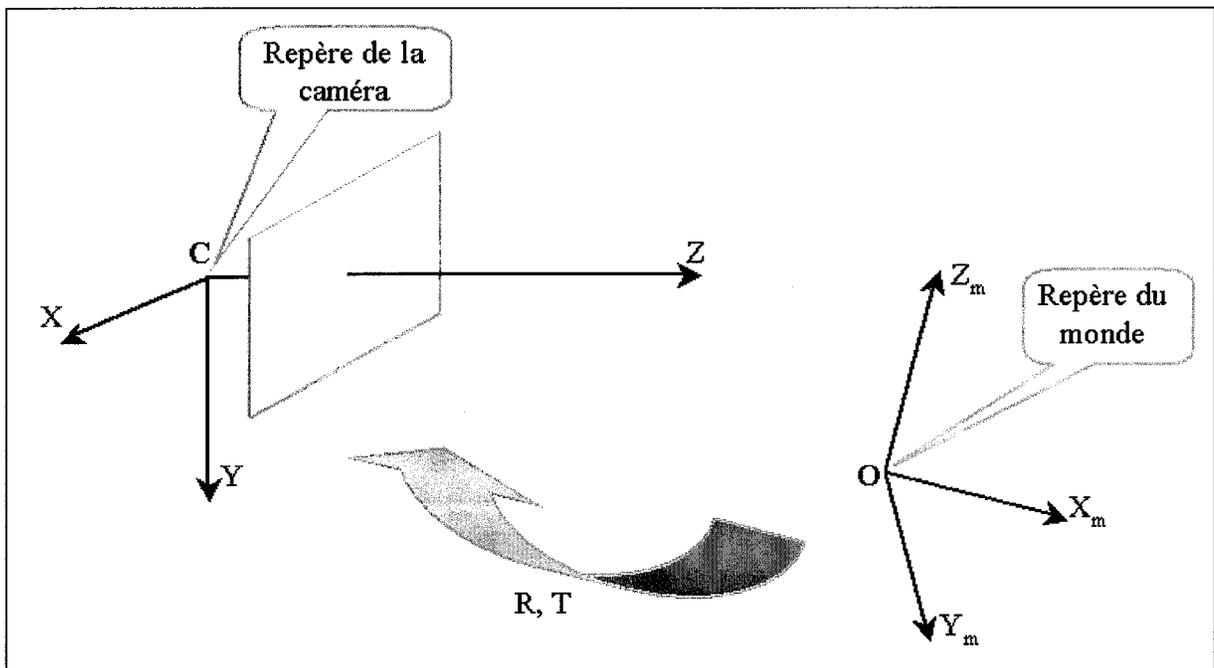


Figure III.6 : La transformation euclidienne entre le repère de la scène et le repère de la caméra.

La matrice représentant cette transformation rigide (composée de 3 rotations et de 3 translations) correspond à l'influence des paramètres extrinsèques. La matrice inverse exprimant les mouvements de la caméra par rapport à la scène est donnée par l'expression suivante :

$$A^{-1} = \begin{pmatrix} R^T & -R^T T \\ 0_3^T & 1 \end{pmatrix} \quad (3.11)$$

3) LA MATRICE DE PROJECTION

En combinant les équations (3.6) et (3.10) nous obtenons les équations de projection pour une caméra avec des paramètres intrinsèques spécifiques et une position et une orientation données :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R & T \\ 0_3^T & 1 \end{pmatrix} \begin{pmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{pmatrix} \quad (3.12)$$

Ceci nous permet donc de définir une matrice de projection 3×4 d'un point de l'espace Euclidien 3D de la scène, exprimé dans le repère du monde, en un point de l'image, exprimé dans le repère 2D attaché à celle-ci :

$$P = \begin{pmatrix} \underbrace{\begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}}_C & \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \end{pmatrix} \begin{pmatrix} R & T \\ 0_3^T & 1 \end{pmatrix} \quad (3.13)$$

Avec : C : matrice 3×3 des paramètres intrinsèques de la caméra.

R : matrice de rotation 3×3 .

T : vecteur de translation 3×1 .

A l'aide des équations (3.1) et (3.13), nous pouvons extraire les paramètres intrinsèques et extrinsèques en fonction des coefficients de la matrice de projection (3.1), et cela en tenant compte de l'orthogonalité de la matrice de rotation :

$$\begin{aligned} u_0 &= P_3 P_1^T & R_2 &= \frac{1}{\alpha_v} (P_2 - v_0 P_3) \\ v_0 &= P_3 P_2^T & R_3 &= P_3 \\ \alpha_u &= \sqrt{P_1 P_1^T - (P_3 P_1^T)^2} & t_x &= \frac{1}{\alpha_u} (p_{14} - u_0 p_{34}) \\ \alpha_v &= \sqrt{P_2 P_2^T - (P_3 P_2^T)^2} & t_y &= \frac{1}{\alpha_v} (p_{24} - v_0 p_{34}) \\ R_1 &= \frac{1}{\alpha_u} (P_1 - u_0 P_3) & t_z &= p_{34} \end{aligned} \quad (3.14)$$

avec P_i : les trois premières colonnes de la ligne i de la matrice de projection 3×4 exprimée dans l'équation (3.1).

R_i : la ligne i de la matrice de rotation R .

4) INTERPRETATION GEOMETRIQUE DE LA MATRICE DE PROJECTION

Soient p_1, p_2, p_3 et p_4 les colonnes de la matrice de projection P de la caméra. Ceux-ci sont des vecteurs 3×1 qui représentent des points image bien particuliers. Les trois premiers vecteurs représentent les images par la matrice de projection des vecteurs supports des axes X_m, Y_m et Z_m du repère du monde. Ils correspondent aux coordonnées des « points de fuite » (*vanishing points*). Par exemple, la direction de l'axe Z_m est représentée par le vecteur $(0 \ 0 \ 1 \ 0)^T$ et sa projection en utilisant le relation (3.1) donne le vecteur p_3 . Le vecteur p_4 , quant à lui, représente l'image de l'origine du repère du monde.

Dans le même ordre d'idée, les lignes de la matrice de projection de la caméra sont des vecteurs 1×4 permettant de retrouver des plans bien particuliers. La troisième ligne permet ainsi de retrouver le plan principal (le plan parallèle au plan image et qui passe par le centre de projection de la caméra). Les deux premières lignes permettent de retrouver les deux autres plans définis par le centre optique et les vecteurs de base pris distincts deux à deux. Le lecteur intéressé par des informations complémentaires se rapportera avec profit à [Har.03].

5) MODELISATION DE LA DISTORSION

Le modèle à Sténopée (modèle linéaire) de la caméra décrit relativement bien le processus de formation d'image pour la plupart des caméras. Cependant, quand un modèle de caméra extrêmement précis est exigé ou quand des objectifs « exotiques » (du type « fisheye ») ou de bas de gamme (du type de ceux utilisés sur certaines WebCam) sont utilisés, des effets supplémentaires doivent être pris en considération. En effet, l'alignement supposé (voire « fondateur » pour le modèle à Sténopée !) entre le centre de projection, le point de la scène et sa projection sur l'image n'est pas toujours vérifiée. Ceci est dû principalement aux propriétés du système optique. La déviation la plus perceptible est généralement celle due à la distorsion radiale. Pour remédier à ce type de distorsion, nous devons corriger les positions des pixels de telle manière que ces derniers aient les coordonnées qu'ils auraient dû avoir dans le cadre d'une projection linéaire parfaite (voir Figure III.5). Pour de plus amples explications sur ce sujet voir [Wil.94]. En tenant compte de ces imperfections de projection, les équations (3.8) s'écrivent sous la forme suivante :

$$\begin{cases} u = \alpha_u X_c + u_0 + \Delta u \\ v = \alpha_v Y_c + v_0 + \Delta v \end{cases} \quad (3.15)$$

Dans cette expression, nous voyons apparaître les deux termes Δu et Δv modélisant la distorsion radiale. Ces termes sont généralement non linéaires (voir [Hor.96, Fau.93, Har.03]). Les nombreuses études menées sur ce sujet ont montré que l'estimation du premier coefficient de la distorsion radiale est suffisante dans la plupart des cas [Tsa.87]. Les relations correspondantes sont alors :

$$\begin{cases} \Delta u = k_1 r^2 (u - u_0) \\ \Delta v = k_1 r^2 (v - v_0) \end{cases} \quad (3.16)$$

$$\text{avec, } r^2 = (u - u_0)^2 + (v - v_0)^2$$

Un tel modèle de caméra contient cinq paramètres intrinsèques (α_u , α_v , u_0 , v_0 et k_1) à estimer au lieu des quatre paramètres du modèle parfaitement linéaire.

III.3.1.4. CHAMPS DE VISION – VOLUME DE TRAVAIL

Le champ de vision donne le volume de la scène qui peut être observé par la caméra. Il se calcule en fonction des dimensions du capteur (horizontale et verticale), de la distance focale de l'objectif et des angles de vue. L'exploitation des notions de champ de vision et de volume de travail s'effectue souvent suivant deux cas, qui sont :

1^{er} Cas :

Dans ce premier cas, nous considérons que :

- La caméra est connue, donc les dimensions du capteur sont aussi connues.
- L'objectif est connu, donc on connaît la distance focale.

On cherche alors la taille d'un objet à une certaine distance D ou vu sous un angle θ , en utilisant les formules suivantes (voir figure III.7) :

$$\text{Taille de l'objet : } T_{\text{Objet}} = (T_{\text{capteur}} \times D) / f \quad (3.17)$$

$$\text{L'angle de vue : } \theta = \arctan(T_{\text{capteur}} / f) \quad (3.18)$$

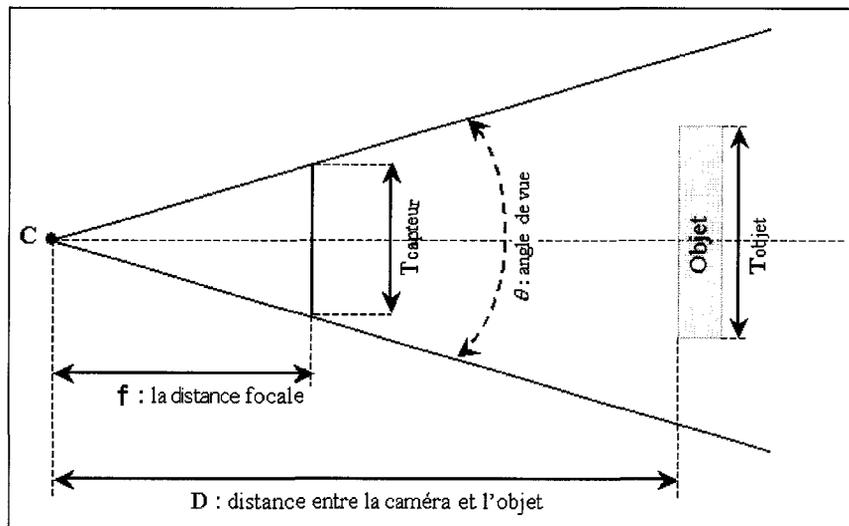


Figure III.7 : Le calcul du champ de vision

2^{ème} Cas :

Dans ce deuxième cas, nous considérons que :

- L'objet est connu, donc la taille de ce dernier est aussi connue.
- La distance D entre la caméra et l'objet est également connue.

On cherche alors à « dimensionner » l'ensemble du système caméra + objectif permettant de visionner cet objet à pleine échelle.

Dans la plupart des cas, on connaît également la taille du capteur, puisque la caméra existe déjà. De la figure III.7, nous pouvons déduire facilement la focale qui permet d'avoir à pleine échelle un objet de taille connue T_{objet} à la distance D . Ainsi, la focale f est donnée par:

$$f = (T_{capteur} \times D) / T_{objet} \quad (3.19)$$

Ce type de raisonnements géométrique très simples va nous permettre de dimensionner le champ visuel de notre système de stéréovision (deux caméras, dans notre cas) et la distance minimale requise pour que la plaque d'immatriculation soit totalement visible par les deux caméras en même temps.

Parmi les systèmes d'acquisition connectables, nous disposons de deux caméscopes identiques (Samsung VP-D80), équipé d'un capteur CCD de taille $1/6''$ et d'un objectif dont la distance focale peut varier de 2.7 à $27mm$. Pour des raisons qui seront détaillées par après, les caméscopes ont été disposés de sorte à ce que leurs axes optiques soient sensiblement parallèles et distants de $300mm$, tandis que les plans rétines des deux caméras sont supposés confondus. La disposition du capteur en fonctionnement normal est telle que le plan défini par les deux axes optiques soit parallèle à la chaussée. Le système conçu doit nous permettre de visualiser une plaque d'immatriculation française, d'une hauteur de $100mm$ et d'une largeur de $510mm$, à des distances raisonnables.

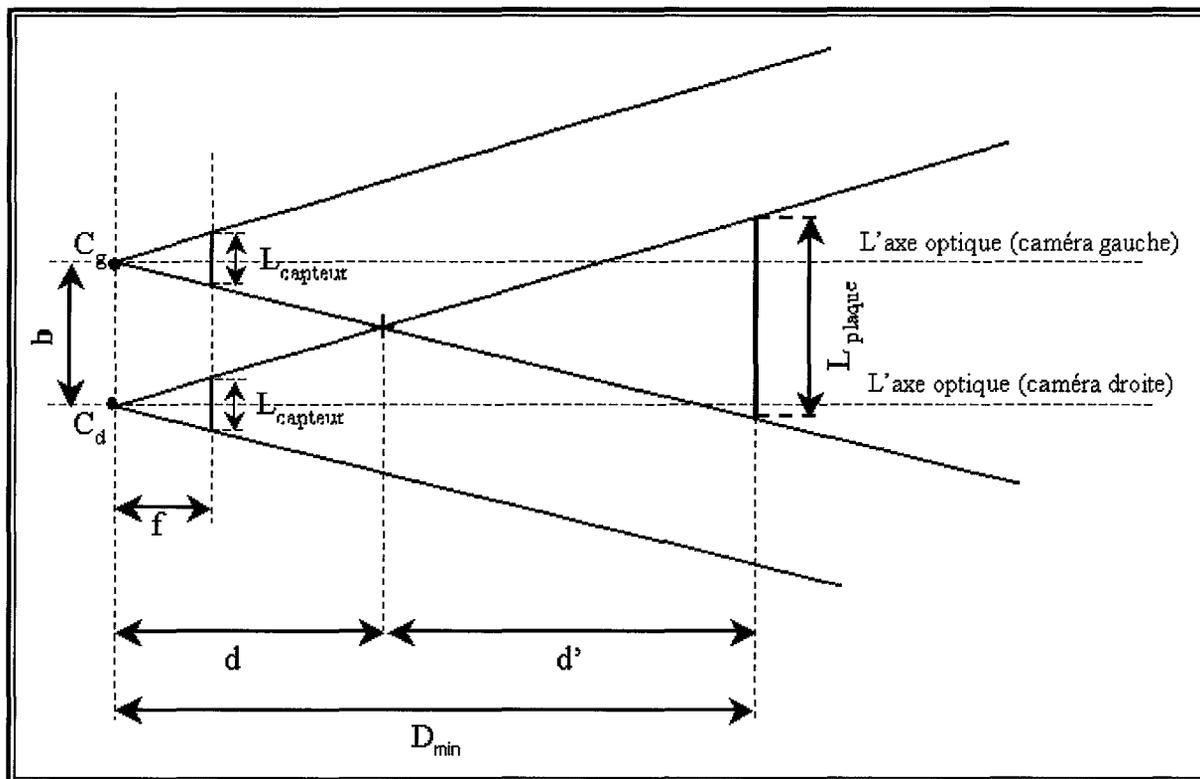


Figure III.8 : Schéma représentant la projection du système de stéréovision et la scène observée (la plaque d'immatriculation) sur le plan de la chaussée.

Comme le montre la Figure III.8, nous pouvons exprimer la distance qui sépare l'objet observé (en l'occurrence la plaque d'immatriculation) du système de stéréovision en fonction de la distance focale f (supposée constamment commune aux deux caméras) :

$$D_{\min} = \frac{(L_{\text{plaque}} + b)}{L_{\text{capteur}}} \cdot f \quad (3.20)$$

Le tableau ci-dessous indique la distance minimale à partir de laquelle la plaque d'immatriculation sera visible pour les deux caméras, pour différentes valeurs de la distance focale :

Distance focale (mm)	Distance minimum requise (mm)
6	2025
8	2700
10	3375

12	4050
16	5400
25	8437.5

Tableau III.3 : Distance minimum à respecter pour que la plaque soit complètement visible par les deux caméras.

Le calcul de la distance minimum D_{\min} , dont les résultats sont repris dans le tableau III.3, suppose que le plan de la plaque d'immatriculation est parallèle au plan contenant les rétines des deux caméras du capteur de stéréovision, comme l'indique la figure III.8.

III.3.2. LE SYSTEME DE STEREOVISION

La localisation d'un point dans l'espace tridimensionnel exige que celui-ci soit au moins observé de deux points de vue différents. Cette exigence est satisfaite en utilisant deux caméras ou en déplaçant une seule caméra effectuant deux prises de vue consécutives, toutes choses égales par ailleurs. Bien sûr, il faudra connaître suivant le cas la position relative des deux caméras ou le déplacement effectué entre les deux images. Compte tenu de l'application, le système de stéréovision est plus approprié pour effectuer des mesures directes et rapides.

III.3.2.1. UNE CONFIGURATION GEOMETRIQUE PARTICULIERE

Dans une configuration quelconque d'un système de stéréovision, l'exploitation de la géométrie épipolaire²⁹ impose de calculer pour chaque point d'une image l'équation de la droite épipolaire correspondante dans l'autre image. Cependant, il existe une configuration particulière intéressante permettant de s'affranchir de ce calcul. En effet, lorsque les deux caméras sont disposées de telle manière que leurs axes optiques soient parallèles et que la droite reliant les centres optiques soit confondue avec leurs axes horizontaux (axe des x), les droites épipolaires conjuguées sont confondues et les deux épipoles sont à l'infini (voir, figure III.9).

²⁹ C'est la géométrie qui relie les points d'une image du système stéréo à son correspondant dans l'autre image.

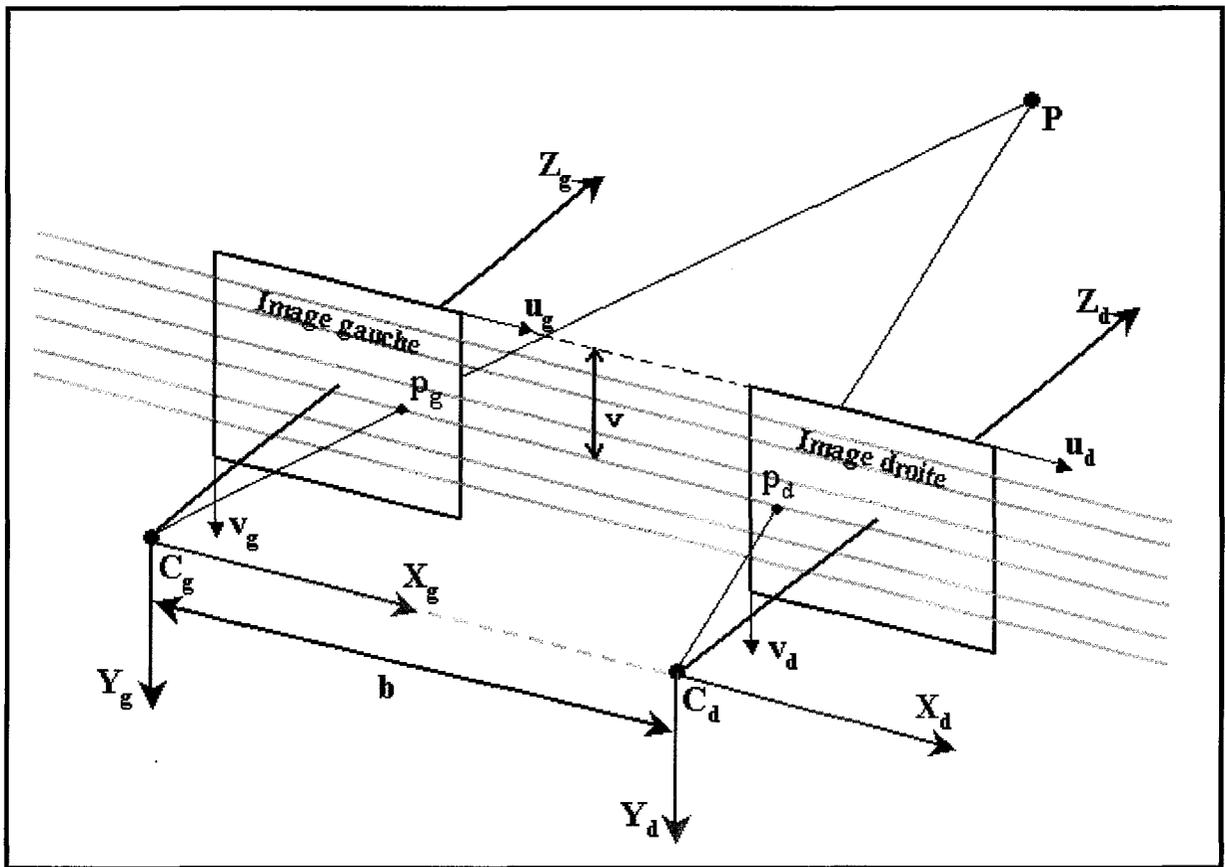


Figure III.9 : Configuration particulière d'un système de stéréovision.

Les droites épipolaires jouent un rôle primordial dans la vision stéréoscopique. Comme nous l'avons déjà évoqué, cette géométrie permet en particulier de restreindre la zone de recherche du correspondant d'un pixel donné à la droite épipolaire correspondante.

La configuration décrite ci-dessus permet ainsi de simplifier grandement les expressions mathématiques de ces droites. La configuration de la figure III.9 peut être aussi obtenue analytiquement à partir d'une configuration quelconque en effectuant une rectification épipolaire [Har.00]. Celle-ci consiste à appliquer aux images une transformation linéaire en coordonnées projectives, de façon à obtenir une paire d'images stéréoscopiques coplanaires, parallèles à la droite passant par les centres de projection des deux caméras et dont les axes des abscisses sont confondus. Notons au passage que cette opération accroît évidemment le temps de calculs !

III.3.2.2. DESCRIPTION DU SYSTEME DE STEREOVISION MIS EN PLACE

La description de notre système fait appel au modèle de caméra à *Sténopée*. Le système de stéréovision que nous avons bâti se compose de deux caméras, placées dans la configuration particulière décrite sur la figure III.9. La relation liant les images des deux caméras est décrite par une matrice d'homographie 3×3 appelée « *matrice fondamentale* ». Cette matrice dispose de plusieurs propriétés importantes [Har.00, For.03, Tru.98], parmi lesquelles la plus utile pour notre application se trouve être :

$$m'^T F m = 0 \quad (3.21)$$

L'équation (3.21) décrit en effet la relation reliant les coordonnées homogènes du point m' appartenant à l'image de droite, aux coordonnées homogènes de son correspondant m dans l'image de gauche. Dans notre cas particulier, la matrice fondamentale aura la forme suivante :

$$F = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \quad (3.22)$$

Cette matrice fondamentale décrit également la géométrie épipolaire de notre système de stéréovision. L'analyse de cette forme homographique spéciale permet de conclure que n'importe quel point appartenant à l'image de gauche aura un correspondant, le cas échéant, sur l'image droite dont l'ordonnée sera identique à celle du point de l'image de gauche. L'expression (3.22) de la matrice fondamentale est obtenue dans le cas où les plans rétine des deux caméras se déduisent l'un de l'autre grâce à une simple translation le long de l'axe des x (voir figure III.9).

La phase nous permettant de calculer la matrice fondamentale exploite la méthode des moindres des médianes des carrés [Zay.03b]. La comparaison entre valeurs réelles et théoriques des coefficients de la matrice fondamentale permet de vérifier *a posteriori* que le positionnement des caméras entre elles est bien conforme à ce que décrit la figure III.9.

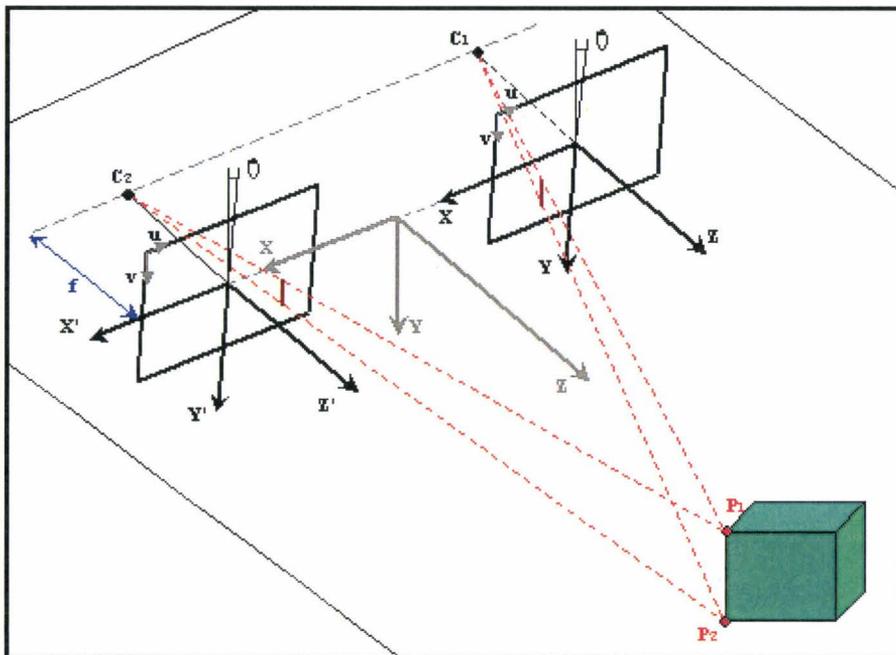


Figure. III.10 : Description de la géométrie du système de vision.

On rappelle qu'étant donné le modèle à *Sténopée* d'une caméra, la relation entre les coordonnées de la projection du point P sur l'image et ses coordonnées dans le repère de la scène (le repère du monde) s'exprimera par :

$$\begin{pmatrix} x \\ y \\ s \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_A \begin{pmatrix} R_g & T_g \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (3.23)$$

$$\begin{pmatrix} x' \\ y' \\ s' \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{A'} \begin{pmatrix} R_d & T_d \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (3.24)$$

Avec:

- $(x \ y \ s)^T$ et $(x' \ y' \ s')^T$ les coordonnées homogènes de la projection du point P sur l'image de gauche et l'image droite respectivement.
- α_u, α_v : définis comme suit; $\alpha_u = \frac{f}{k_u}$ et $\alpha_v = \frac{f}{k_v}$.

où k_u et k_v : la largeur et la hauteur du pixel, respectivement.

- R_g et R_d : les matrices de rotation entre la référence de la scène et les caméras gauche et droite, respectivement.
- $(X \ Y \ Z \ 1)^T$: les coordonnées homogènes du point P par rapport à la référence de la scène.
- T_g, T_d : les vecteurs de translation, exprimant la transformation en terme de translation entre la référence de la scène et les références des caméras gauche et droite, respectivement.
- A, A' : les matrices des projections perspectives. Elles expriment la transformation entre le repère des cameras et celui des images.

Dans notre cas, nous utilisons deux cameras identiques. Il est ainsi légitime de considérer que $A \cong A'$ (cette hypothèse a au demeurant été vérifiée expérimentalement).

D'après la Figure III.7, les transformations respectives entre la référence de la scène et les références des caméras peuvent être exprimées par des rotations autour de l'axe des X et une translation au long de l'axe des X . Ainsi, les matrices de rotation et les vecteurs de translations s'expriment de la manière suivante :

$$R = R_g = R_d = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix} \quad (3.25)$$

$$T_g = \begin{pmatrix} \frac{b}{2} \\ 0 \\ 0 \end{pmatrix} \quad (3.26) \quad T_d = \begin{pmatrix} \frac{-b}{2} \\ 0 \\ 0 \end{pmatrix} \quad (3.27)$$

III.3.3. LA CALIBRATION

Après la description d'un modèle de caméra suffisamment « générique » pour l'application envisagée, il importe de savoir de quelle façon il est possible d'accéder aux valeurs des paramètres du modèle. C'est le rôle de la phase de calibration. Il s'agit d'une étape nécessaire dans le domaine de la vision par ordinateur tridimensionnelle, ceci afin de pouvoir extraire l'information 3D à partir des images 2D. Plusieurs travaux ont été effectués dans ce sens. La phase instigatrice revient à la communauté de la photogrammétrie [Bro.71, Fai.75], et c'est un peu plus tard que le domaine de la vision par ordinateur apporta sa pierre à l'édifice.

Dans la mesure où notre application nécessite la connaissance d'informations métriques tridimensionnelles suffisamment précises, la calibration des caméras composant le système de stéréovision est d'une importance primordiale. La technique de calibration mise en œuvre au niveau de ce capteur se doit de satisfaire à un certain nombre d'exigences, dont les principales sont la flexibilité et la précision.

Rappelons que notre objectif est d'aboutir à un attelage virtuel dont la seule source de données extéroceptives est constituée par notre système de vision. Sachant que la technique de calibration employée conditionnera directement la précision des mesures obtenues, il est clair que celle-ci doit être choisie avec un soin tout particulier. La technique adoptée par la plupart des groupes de recherches dans le domaine des transports intelligents consiste, jusqu'à présent, à utiliser une mire peinte sur la chaussée. Les coordonnées tridimensionnelles de certains points caractéristiques (généralement extraits à la main) sont connues avec précision par rapport à une référence liée au monde. Il s'agit donc d'une technique basée sur la calibration photogramétrique. Cette méthode souffre d'un réel manque de flexibilité, puisque la calibration est basée sur une mire tridimensionnelle non transportable ! Ainsi, la calibration ne peut être effectuée qu'à un endroit bien précis. Des techniques de mise à jour dynamique des paramètres de calibration ont été proposées dans [Bro.99b].

Nous pouvons distinguer trois techniques principales de calibration :

- La calibration photogramétrique (appelée également la calibration classique): La calibration de la caméra est effectuée en observant un objet (« mire ») de calibration dont la géométrie dans l'espace euclidien est connue avec une bonne précision. Cette méthode permet d'aboutir à une calibration très fine. La mire de calibration se compose généralement de deux ou trois plans orthogonaux entre eux, supportant un ensemble de points facilement et précisément repérables. Une alternative (qui se ramène toutefois au procédé décrit ici) consiste à utiliser un même plan qu'on déplace d'une façon connue avec précision entre chaque prise de vue [Tsa.87].
- La calibration hybride: La technique proposée dans cette catégorie nécessite l'utilisation d'une mire plane dont les coordonnées bidimensionnelles de certains points caractéristiques sont connues, basée sur les homographies définissant les transformations entre les plans de la mire (dans ses différentes positions successives) et le plan de la caméra (ou des caméras) à calibrer [Zha.00]. La caméra ou la mire peuvent être déplacées et il n'est pas nécessaire de connaître *a priori* les déplacements relatifs effectués. Cette approche se situe entre la calibration photogramétrique et l'auto calibration puisque la connaissance préalable de certaines informations dimensionnelles relatives à la mire est requise et que cette méthode utilise par ailleurs des propriétés de la géométrie projective elles-aussi exploitées par l'auto calibration.
- L'auto calibration: Les techniques associées à cette catégorie ne nécessitent l'utilisation d'aucune mire de calibration, elle se base essentiellement sur la géométrie épipolaire (*matrice fondamentale* et *matrice essentielle*). L'idée consiste à déplacer la caméra dans une scène statique. La rigidité de la scène fournit généralement deux contraintes [Luo.92, May.92] sur les paramètres intrinsèques de la caméra. Si les images sont prises avec la même caméra (dont on suppose les paramètres intrinsèques fixes), la mise en correspondance entre trois images est suffisante pour extraire les paramètres intrinsèques et extrinsèques [Luo.97, Har.94]. Malgré la flexibilité de cette approche, celle-ci manque encore de maturité. Vu le nombre de paramètres à estimer, on n'obtient pas toujours des résultats fiables [Bou.98].

Les calibrations photogramétrique et hybride présentent de meilleurs résultats comparativement aux techniques d'auto calibration. En effet, les techniques d'auto calibration proposées jusqu'à présent sont adaptées aux problématiques des systèmes de vision actifs dont les paramètres des caméras varient continuellement mais ne conviennent pas aux systèmes qui exigent une précision importante des informations métriques (la majorité de ces techniques supposent l'absence de distorsions optiques ou supposent connaître les coordonnées de la projection du centre optique).

Pour les raisons citées ci-dessus, nous nous sommes plutôt orientés vers la calibration hybride puisque cette technique, moins précise que la calibration photogramétrique, fournit néanmoins un compromis tout à fait valable entre la précision et la flexibilité, répondant ainsi aux exigences que nous nous étions fixées.

Dans ce qui suit, nous allons mettre en avant les fondement théoriques de chacune des méthodes citées ci-avant. Une attention particulière sera accordée à la technique de la calibration hybride.

III.3.3.1. LA CALIBRATION PHOTOGRAMETRIQUE

A partir de l'équation (3.1), nous pouvons écrire les coordonnées image de la projection d'un point de la scène, en fonction de ses coordonnées par rapport au repère euclidien tridimensionnel du monde, de la manière suivante :

$$\begin{cases} u = \frac{p_{11}X_m + p_{12}Y_m + p_{13}Z_m + p_{14}}{p_{31}X_m + p_{32}Y_m + p_{33}Z_m + p_{34}} \\ v = \frac{p_{21}X_m + p_{22}Y_m + p_{23}Z_m + p_{24}}{p_{31}X_m + p_{32}Y_m + p_{33}Z_m + p_{34}} \end{cases} \quad (3.28)$$

Notons que le système d'équations (3.28) décrit la droite passant par le centre de projection de la caméra C et le point image $(u \ v)^T$, dans le repère de la scène. Cette droite correspond donc à la droite (Cp) de la Figure III.4. On l'appelle également *la droite de vue* associée à un point image de coordonnées $(u \ v)^T$.

Les équations de (3.28) peuvent se réécrire sous la forme d'une combinaison linéaire de p_{ij} , reliant les coordonnées d'un point de la scène de coordonnées homogènes $(X_{mi} \ Y_{mi} \ Z_{mi} \ 1)^T$ aux coordonnées homogènes image de sa projection $(u_i \ v_i \ 1)^T$ sous la forme suivante :

$$\begin{cases} X_{mi}p_{11} + Y_{mi}p_{12} + Z_{mi}p_{13} + p_{14} - u_i X_{mi}p_{31} - u_i Y_{mi}p_{32} - u_i Z_{mi}p_{33} = u_i p_{34} \\ X_{mi}p_{21} + Y_{mi}p_{22} + Z_{mi}p_{23} + p_{24} - v_i X_{mi}p_{31} - v_i Y_{mi}p_{32} - v_i Z_{mi}p_{33} = v_i p_{34} \end{cases} \quad (3.29)$$

Nous obtenons ainsi $2n$ équations pour n points. Pour pouvoir déterminer les composantes de la matrice de projection, nous devons disposer d'au moins 11 équations. Par conséquent, il nous faut connaître la position 3D de 6 points de la scène et de leurs projections respectives sur l'image. Etant donnés n points de la scène, nous obtenons un système d'équations que nous pouvons exprimer sous forme matricielle (les indices représentent les dimensions des matrices ou des vecteurs) :

$$K_{2n \times 11} x_{11} = u_{2n} \quad (3.30)$$

La « solution » de ce système d'équations linéaires, éventuellement surabondantes, peut s'exprimer au sens des moindres carrés sous la forme classique suivante :

$$x_{11} = K^+ u_{2n} \quad (3.31)$$

Avec, x_{11} le vecteur de dimension 11×1 contenant les composantes de la matrice de projection à un facteur d'échelle près.

K^+ la matrice pseudo – inverse de dimension $2n \times 11$ de la matrice K telle que :
 $K^+ = (K^T K)^{-1} K^T$.

u_{2n} le vecteur de dimension $2n$ contenant les coordonnées image des projections des points de la scène utilisés pour la construction du système d'équations.

Comme nous l'avons évoqué, la calibration photogramétrique nécessite l'utilisation d'une mire de calibration supportant des points dont les coordonnées sont connues avec une grande précision dans le repère euclidien de la scène. Il est alors possible de calculer directement la matrice de projection en utilisant l'équation (3.31). Les équations (3.14) sont quant à elles exploitées pour l'extraction explicite des paramètres de la caméra.

La technique de calibration décrite ci-dessus est par nature adaptée à un modèle linéaire de la caméra (modèle à Sténopée parfait), ne tenant pas compte des distorsions. Ce modèle « simple » est adopté comme une première approche par plusieurs chercheurs (Hall [Hal.82], Toscani-Faugeras [Fau.86, Tos.87] et Ito [Ito.91]).

Si l'on souhaite tenir compte des distorsions liées à l'optique des caméras (dans le but par exemple d'obtenir un jeu de paramètres plus « robuste » pour le modèle), il est nécessaire d'adopter les équations (3.15), ce qui conduit à :

$$\begin{cases} (u_i - u_0) - \alpha_u \frac{r_{11}X_{mi} + r_{12}Y_{mi} + r_{13}Z_{mi} + t_x}{r_{31}X_{mi} + r_{32}Y_{mi} + r_{33}Z_{mi} + t_z} - k_1 r^2 (u_i - u_0) = 0 \\ (v_i - v_0) - \alpha_v \frac{r_{21}X_{mi} + r_{22}Y_{mi} + r_{23}Z_{mi} + t_y}{r_{31}X_{mi} + r_{32}Y_{mi} + r_{33}Z_{mi} + t_z} - k_1 r^2 (v_i - v_0) = 0 \end{cases} \quad (3.32)$$

Le système d'équations (3.32) est non linéaire. Sa résolution implique donc la mise en œuvre de techniques d'optimisation non-linéaire, telle que la méthode de Newton-Raphson ou celle de Levenberg-Marquardt [Pre.88]. Comme c'est souvent le cas, les résultats produits par ces algorithmes dépendent essentiellement du choix des valeurs initiales des paramètres.

En la matière, une des solutions consiste à évaluer les paramètres (sauf k_1) avec une des techniques linéaires citées ci-dessus et à utiliser les résultats obtenus comme valeurs initiales. La littérature fournit plusieurs méthodes utilisant la calibration photogramétrique non linéaire, telle que celle de Tsai [Tsa.87] (ne tenant compte que de la distorsion radiale) ou encore celle de Weng [Wen.92] qui inclut dans son modèle différents types de distorsions. Une étude comparative entre ces techniques de calibration peut être trouvée dans [Sal.02].

III.3.3.2 LA CALIBRATION HYBRIDE

L'approche proposée ici se situe entre la calibration photogramétrique et l'auto calibration, du fait qu'on utilise des informations métriques 2D au lieu des informations métriques 3D. Par ailleurs, les fondements théoriques même de l'auto calibration apparaissent clairement dans la formulation de la calibration hybride (au travers de la notion d'image du cône absolu). Les propriétés essentielles de cette approche sont sa flexibilité par rapport à l'approche photogramétrique, et sa robustesse relativement à l'auto calibration [Zha.00].

A) L'HOMOGRAPHIE ENTRE LE PLAN DE LA MIRE ET SON IMAGE

Le point de départ de cette approche réside dans la définition de la transformation (homographie) entre le plan de la mire et son image. Supposons que la mire se situe initialement à $Z = 0$ dans le repère de la scène. Soit alors r_i , la $i^{\text{ème}}$ colonne de la matrice de rotation R . Le système d'équations (3.12) peut ainsi s'écrire de la manière suivante:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} r_1 & r_2 & r_3 & T \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_m \\ Y_m \\ 0 \\ 1 \end{pmatrix} \Leftrightarrow \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_H \begin{pmatrix} r_1 & r_2 & T \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_m \\ Y_m \\ 1 \end{pmatrix} \quad (3.33)$$

Il est donc possible de travailler avec le point $P = [X_m \ Y_m \ 1]^T$, compte-tenu de la contrainte $Z = 0$. Par ailleurs, un point P de la mire est relié à son image p par l'homographie H :

$$s p = HP \quad \text{avec} \quad H = \underbrace{\begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}}_c \begin{pmatrix} r_1 & r_2 & T \end{pmatrix} \quad (3.34)$$

La matrice d'homographie H est une matrice (3×3) définie à un facteur d'échelle près.

B) LES CONTRAINTES SUR LES PARAMETRES INTRINSEQUES

Ayant l'image d'une mire plane, plusieurs méthodes permettent d'estimer l'homographie introduite dans ce qui précède. Parmi ces méthodes nous en présentons une basée sur un critère de maximum de ressemblance. En effet, d'un point de vue pratique, un point P_i de la scène et son image p_i sur le plan de la rétine de la caméra ne satisferont pas en toute rigueur l'équation (3.34), à cause du bruit affectant les points d'intérêt extraits de l'image. Il est commode en l'espèce de supposer que les p_i sont affectés par un bruit Gaussien de moyenne nulle et de matrice de covariance Λ_{p_i} . L'estimation au maximum de ressemblance de H est alors obtenue par la minimisation de la fonction suivante :

$$\sum_i (p_i - \hat{p}_i)^T \Lambda_{p_i}^{-1} (p_i - \hat{p}_i), \quad (3.35)$$

Avec $\hat{p}_i = \frac{1}{h_3^T P_i} \begin{bmatrix} \bar{h}_1^T P_i \\ \bar{h}_2^T P_i \end{bmatrix}$ et h_i la $i^{\text{ème}}$ ligne de H .

En pratique, on suppose simplement que $\Lambda_{p_i} = \sigma^2 I$, avec I est la matrice identité. Cette hypothèse est raisonnable si les points sont extraits de façon indépendante et en appliquant toujours la même procédure. Dans ce cas, le problème ci-dessus devient un problème des moindres carrés non-linéaire, c.-à-d., $\min_H \sum_i \|p_i - \hat{p}_i\|^2$. La minimisation non-linéaire peut être conduite en appliquant l'algorithme de Levenberg-Marquardt [Pre.88, Mor.77]. Cette procédure exige au demeurant une estimation initiale des coefficients de H , estimation qui peut être obtenue de la manière suivante :

Estimation initiale des coefficients de H :

Soit h , un vecteur contenant les coefficients de la matrice d'Homographie H . A partir de l'équation (3.34) on obtient :

$$\underbrace{\begin{bmatrix} \tilde{P}^T & 0^T & -u\tilde{P}^T \\ 0^T & \tilde{P}^T & -v\tilde{P}^T \end{bmatrix}}_L h = 0 \quad (3.36)$$

Si on dispose de n paires de points mis en correspondance entre la mire et son image, on obtient un système d'équations du type $L \cdot h = 0$, avec L une matrice $2n \times 9$. Puisque h est définie à un facteur d'échelle près, la solution de ce système d'équations est le vecteur singulier de droite de L , associé à la plus petite valeur singulière ou, d'une façon équivalente, le vecteur propre associé à la plus petite valeur propre de la matrice $L^T L$.

Remarque : La matrice L contient des valeurs constantes égales à 1, alors que d'autres coefficients sont en pixels ou encore en coordonnées associées à la scène, voire en une combinaison de ces deux types de données. Pour cette raison, la matrice L est mal conditionnée numériquement. Des résultats plus stables sont obtenus en appliquant une normalisation des données (voir, [Har.95]) avant d'entamer la procédure décrite ci-dessus.

Nous supposons dans ce qui suit que la matrice de l'homographie $H = [h_1 \ h_2 \ h_3]$ est calculée à partir de (3.34) :

$$(h_1 \ h_2 \ h_3) = \lambda \underbrace{\begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}}_c (r_1 \ r_2 \ r_3) \quad (3.37)$$

où λ représente un scalaire arbitraire.

En prenant en considération le fait que r_1 et r_2 sont orthogonaux, il vient:

$$h_1^T (C^{-1})^T (C^{-1}) h_2 = 0 \quad (3.38)$$

$$h_1^T (C^{-1})^T (C^{-1}) h_1 = h_{21}^T (C^{-1})^T (C^{-1}) h_2 \quad (3.39)$$

Les équations (3.38) et (3.39) sont les contraintes de base sur les paramètres intrinsèques, à partir d'une seule homographie. Notons au passage que $(C^{-1})^T (C^{-1})$ décrit l'image du cône absolu [Luo.97,Pol.99].

C) EXTRACTION DES PARAMETRES INTRINSEQUES ET EXTRINSEQUES

Dans ce qui suit, nous allons décrire la méthode qui permet l'extraction des paramètres de calibration. Posons $B = (C^{-1})^T (C^{-1})$, on aura donc :

$$B = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} = \begin{bmatrix} \frac{1}{\alpha_u^2} & 0 & -\frac{u_0}{\alpha_u^2} \\ 0 & \frac{1}{\alpha_v^2} & -\frac{v_0}{\alpha_v^2} \\ -\frac{u_0}{\alpha_u^2} & -\frac{v_0}{\alpha_v^2} & \frac{u_0}{\alpha_u^2} + \frac{v_0}{\alpha_v^2} + 1 \end{bmatrix} \quad (3.40)$$



Notons que la matrice B est une matrice symétrique définie par un vecteur de dimension (6×1) :

$$b = [B_{11} \quad B_{12} \quad B_{22} \quad B_{13} \quad B_{23} \quad B_{33}]^T \quad (3.41)$$

Soit le $i^{\text{ème}}$ vecteur colonne de H , tel que: $h_i = [h_{i1}, h_{i2}, h_{i3}]^T$. Nous aurons donc:

$$h_i^T B h_j = v_{ij}^T b \quad (3.42)$$

Avec, $v_{ij} = [h_{i1} h_{j1}, h_{i1} h_{j2} + h_{i2} h_{j1}, h_{i2} h_{j2}, h_{i3} h_{j1} + h_{i1} h_{j3}, h_{i3} h_{j2} + h_{i2} h_{j3}, h_{i3} h_{j3}]^T$

En adoptant l'écriture ci-dessus et à partir des équations (3.38) et (3.39), nous aurons :

$$\begin{pmatrix} v_{12}^T \\ (v_{11} - v_{22})^T \end{pmatrix} b = 0 \quad (3.43)$$

Si nous observons n images de la mire (plane par hypothèse), nous obtiendrons n équations du type de (3.68). Nous aurons un système d'équations de la forme $V \cdot b = 0$, avec V une matrice $2 \cdot n \times 6$. Si $n \geq 3$, le système a une solution unique et générale dans le cas des cinq

inconnues (c'est à dire cinq paramètres intrinsèques au lieu de quatre³⁰). Si $n \geq 2$, le système d'équations a une solution unique en imposant une contrainte additionnelle, qui consiste ici à supposer que les axes de l'image sont orthogonaux. Par contre, si on ne possède qu'une seule image ($n = 1$), on ne peut résoudre le système d'équations que pour deux variables seulement, en l'occurrence u_0 et v_0 ³¹.

La solution du système d'équations obtenue à partir des images de la mire et de l'expression (3.43) est le vecteur propre de $V^T V = 0$, associé à la plus petite valeur propre ou, d'une façon équivalente, le vecteur singulier de droite associé à la plus petite valeur singulière de V .

En utilisant, les équations (3.40) et (3.43) nous pouvons maintenant estimer les paramètres intrinsèques de la caméra. La matrice B est estimée à un facteur d'échelle près puisque, $B = \lambda(C^{-1})^T(C^{-1})$, avec λ un scalaire arbitraire. Donc, les paramètres intrinsèques sont donnés par :

$$\left\{ \begin{array}{l} v_0 = -\frac{B_{23}}{B_{22}} \\ \lambda = B_{33} + v_0 B_{23} - \frac{B_{13}^2}{B_{11}} \\ \alpha_u = \sqrt{\frac{\lambda}{B_{11}}} \\ \alpha_v = \sqrt{\frac{\lambda}{B_{22}}} \\ u_0 = -\frac{B_{13}}{B_{11}} \end{array} \right. \quad (3.44)$$

Il convient de noter que pour ces expressions, nous avons supposé que les axes de l'image étaient orthogonaux (ce qui est tout à fait légitime d'un point de vue pratique). Cependant, une solution tenant compte du cinquième paramètre intrinsèque est décrite dans l'article de Zhang [Zha.00].

Les paramètres intrinsèques ayant été estimé, les paramètres extrinsèques associés à chaque image se calculent en exploitant (3.34) de la façon suivante :

³⁰ Si on suppose que les axes de l'image ne sont pas tout à fait orthogonaux.

³¹ Dans ce cas, on suppose que u_0 et v_0 se trouvent dans le centre de l'image, et que les axes de l'image sont orthogonaux.

$$\begin{cases} r_1 = \lambda C^{-1} h_1 \\ r_2 = \lambda C^{-1} h_2 \\ r_3 = r_1 \times r_2 \\ t = \lambda C^{-1} h_3 \\ \lambda = \frac{1}{\|C^{-1} h_1\|} = \frac{1}{\|C^{-1} h_2\|} \end{cases} \quad (3.45)$$

Du fait de l'influence de différents bruits, la matrice de rotation calculée $R = (r_1 \ r_2 \ r_3)$ ne satisfera pas en toute rigueur les propriétés des matrices de rotation. Toutefois, il est possible d'estimer une meilleure matrice de rotation depuis une matrice 3×3 (pour plus d'information à ce sujet voir [Gol.96]).

La calibration hybride permet de prendre en compte plus facilement les différentes distorsions, du fait de l'utilisation d'une mire.

Pour terminer, ZhangYou Zhang [Zha.00] propose l'estimation d'un critère de maximum de ressemblance en tenant compte de tous les paramètres, ceci afin d'affiner les résultats obtenus. Ceci nous conduirait à un problème d'optimisation non linéaire, qu'on pourrait résoudre en adoptant l'algorithme de Levenberg-Marquardt [Mor.77].

Remarque: Cette technique de calibration dégénère dans le cas où la mire effectuerait une translation pure. Cette dégénérescence est due essentiellement au fait que les équations (3.38) et (3.39) sont issues des propriétés de la matrice de rotation. Si $R_1 = R_2$, la deuxième image ne génère aucune contrainte supplémentaire !

III.3.3.3. L'AUTOCALIBRATION (SELF-CALIBRATION)

Le but de l'auto calibration est d'élaborer une méthode qui permette de calibrer un système de vision en exploitant les images acquises par ce système pendant la réalisation de sa tâche habituelle. Cette technique exige que la caméra (ou le système stéréo) subisse une série de déplacements dans une scène rigide et que nous soyons en mesure d'établir les correspondances entre les points dans les images acquises par la caméra aux différentes positions. Le problème est donc d'identifier les points dans chacune des images qui représentent les différentes projections d'un même point de la scène sur la rétine de la caméra. Comme cette technique n'exploite aucune information métrique *a priori*, les paramètres extrinsèques sont extraits à un facteur d'échelle près, et ne permettent ultérieurement qu'une reconstruction projective. Dans le cas d'un système stéréoscopique, cela ne présente pas un problème, dans la mesure où cette information est suffisante pour déterminer la géométrie épipolaire et effectuer la reconstruction 3D à un facteur d'échelle près. Malgré tout, cette information métrique peut être extraite simplement en exploitant un objet de la scène dont la taille est connue. En ce qui concerne sa mise en œuvre effective, l'auto calibration peut être décomposée en six étapes distinctes [Luo.93] :

A) EXTRACTION DES POINTS D'INTERET

Puisque l'auto calibration implique la faculté de pouvoir isoler et détecter des points particuliers dans chaque image de la séquence (les « points d'intérêt », principalement des « coins ») nous donnons ici les trois grandes catégories d'algorithmes permettant d'accomplir cette tâche :

- a) Les filtres non-linéaires, tels que le détecteur de coins SUSAN proposé par Smith [Smi.92]. L'avantage de ce filtre est sa relative insensibilité aux bruits du fait qu'il n'utilise pas la dérivée de l'image.
- b) Les méthodes basées sur les courbures, telle que la méthode de Kitchen et Rosenfelds [Kit.82]. Cette dernière méthode nécessite l'extraction préalable des contours avant de détecter les « coins », déterminés d'après la courbure des contours.
- c) Les méthodes basées sur les variations d'intensité pixel. La mise en application typique de cette classe d'algorithme est fournie par la méthode de Harris et Stephens [Har.88] qui détecte les coins en analysant les valeurs propres d'une matrice construite localement autour de chaque pixel à partir de l'intensité de ses voisins.

Les deux premières catégories présentent l'inconvénient d'une certaine complexité des calculs qui influence significativement leur temps d'exécution (relativement long). A l'inverse, le détecteur de Harris³² et Stephens présente l'avantage d'un temps d'exécution assez bref vu la technique adoptée par ce filtre (voir annexe). Schmid et al. [Sch.98] ont comparé plusieurs détecteurs de primitives d'intérêt et ils concluent que le détecteur de Harris semble présenter la meilleure « réputation » (c'est à dire qu'il fournit des résultats assez peu sensibles à la rotation, à la résolution de l'image et à l'illumination) tout en exploitant une quantité d'information plus élevée (constitué par le voisinage des points d'intérêt extraits).

B) LA MISE EN CORRESPONDANCE DES POINTS D'INTERET

Après, la phase d'extraction des points d'intérêt dans une image abordée ci-dessus. Le problème qui se pose maintenant est de mettre en correspondance chacun de ces points d'intérêt dans une image donnée avec leurs homologues dans toutes les autres images. Cette « mise en correspondance » des pixels d'intérêt se base essentiellement sur le calcul d'un critère de corrélation entre deux voisinages afin de mesurer la similitude entre ces deux régions. Plusieurs formulations de ces coefficients de corrélations, entre une partie d'une image « source » et une partie d'une image « destination », peuvent être adoptées pour mesurer la similitude entre les pixels d'intérêt [Alh.98] :

³² Ce détecteur est connu aussi sous le nom de détecteur de Plessey.

$$C_1(u, v, d_u, d_v) = \frac{\sum_{(i,j) \in F} (I_s(u+i, v+j) - I_d(u+d_u+i, v+d_v+j))^2}{\sqrt{\sum_{(i,j) \in F} (I_s^2(u+i, v+j))} \sqrt{\sum_{(i,j) \in F} (I_d^2(u+d_u+i, v+d_v+j))}}$$

$$C_2(u, v, d_u, d_v) = \frac{\sum_{(i,j) \in F} (I_s(u+i, v+j) I_d(u+d_u+i, v+d_v+j))}{\sqrt{\sum_{(i,j) \in F} (I_s^2(u+i, v+j))} \sqrt{\sum_{(i,j) \in F} (I_d^2(u+d_u+i, v+d_v+j))}}$$

$$C_3(u, v, d_u, d_v) = \frac{\sum_{(i,j) \in F} \left[(I_s(u+i, v+j) - \overline{I_s(u, v)}) - (I_d(u+d_u+i, v+d_v+j) - \overline{I_d(u+d_u, v+d_v)}) \right]^2}{\sqrt{\sum_{(i,j) \in F} [I_s(u+i, v+j) - \overline{I_s(u, v)}]^2} \sqrt{\sum_{(i,j) \in F} [I_d(u+d_u+i, v+d_v+j) - \overline{I_d(u+d_u, v+d_v)}]^2}}$$

$$C_4(u, v, d_u, d_v) = \frac{\sum_{(i,j) \in F} (I_g(u+i, v+j) - I_g(x, y))(I_d(u+d_u+i, v+d_v+j) - \overline{I_d(u+d_u, v+d_v)})}{\sqrt{\sum_{(i,j) \in F} [I_g(u+i, v+j) - \overline{I_g(u, v)}]^2} \sqrt{\sum_{(i,j) \in F} [I_d(u+d_u+i, v+d_v+j) - \overline{I_d(u+d_u, v+d_v)}]^2}}$$

avec :

- $I_s(u, v)$ et $I_d(u, v)$, les intensités des images source et de destination, à la position du pixel $(u \ v)^T$.
- F , la fenêtre de la corrélation de dimension $(2n+1) \times (2m+1)$ centrée autour du pixel $(u \ v)^T$. On a $-n \leq i \leq n$ et $-m \leq j \leq m$.
- d_u et d_v respectivement les décalages horizontaux et verticaux, par rapport à l'image de destination.
- $\overline{I_s(u, v)}$ et $\overline{I_d(u, v)}$, les moyennes des intensités des images source et destination sur la fenêtre F .
- $C(x, y, d_u, d_v)$, la valeur du critère de corrélation au pixel de coordonnées $(u \ v)^T$ pour une disparité $d = (d_u \ d_v)$.

Les critères C_1 et C_3 utilisent la différence entre les niveaux de gris des images, et doivent être minimisés par rapport à la disparité; par ailleurs, les critères C_2 et C_4 multiplient les niveaux de gris des deux images et doivent être maximisés par rapport à la disparité.

Le critère $C_4(x, y, d_u, d_v)$ produit généralement des résultats nettement supérieurs par rapport aux autres (moins de faux appariements). Il s'agit en effet d'un critère « normalisé » qui se montre de ce fait plus robuste vis-à-vis des changements d'éclairage.

C) DETERMINATION DE LA MATRICE FONDAMENTALE

Nous avons décrit des outils nous permettant d'isoler des points d'intérêt dans une image et d'associer dans une certaine mesure ces points particuliers d'une image à l'autre. La question est maintenant de savoir quel type d'information il est possible d'en tirer dans le

contexte de l'auto calibration. C'est à ce titre que la matrice fondamentale joue un rôle très important.

En effet, la plupart des algorithmes basés sur les points des images non-calibrées [Pol.99, Stu.97], nécessitent la matrice fondamentale en tant que donnée primordiale. Cette matrice représente la seule alternative aux matrices de projection, afin de relier entre elles les différentes vues de la scène. Dans ce qui suit nous allons présenter brièvement le concept de la matrice fondamentale et les différents algorithmes qui permettent de l'extraire.

1) LA GEOMETRIE EPIPOLAIRE

La géométrie épipolaire est la géométrie projective intrinsèque entre deux vues [Har.99]. Elle est indépendante de la structure de la scène et n'est influencée que par les paramètres intrinsèques des caméras et leurs positions relatives. Dans le cas d'un système stéréoscopique, la géométrie épipolaire décrit essentiellement la géométrie d'intersection des plans formés par les centres optiques des deux caméras et les points observés de la scène avec les plans des images. Dans le cas d'une caméra unique se déplaçant, la géométrie épipolaire se rapporte à l'intersection des plans formés par les points observés et les positions consécutives du centre optique d'une part, et les plans images consécutifs d'autre part.

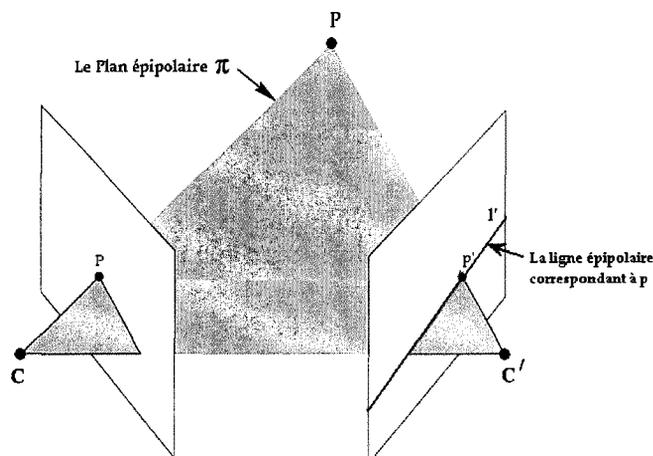


Figure III.11 : La géométrie épipolaire décrivant la correspondance entre deux images d'un point de la scène.

Soit un point P de la scène observé par un capteur stéréoscopique (ou un point P d'une scène rigide observée par une caméra de deux points de vue consécutifs différents), p est sa projection sur l'image de la caméra de gauche (ou sur la première image) et p' sur l'image de la caméra droite (ou sur la seconde image). Il est clair que P , C et C' forment un plan π (appelé le plan épipolaire, voir Figure III.6). Cette construction simple facilite énormément la mise en correspondance entre les images des points observés. En effet, le correspondant d'un point p appartenant à la première image appartiendra forcément à la ligne épipolaire définie par l'intersection entre le plan épipolaire π et le plan image de la deuxième caméra. Par

conséquent, lorsqu'on cherche pour un point p de l'image de gauche (ou de la première image) un correspondant dans l'image de droite (ou dans la seconde image), on peut limiter cette recherche le long de la ligne épipolaire correspondante l' au lieu de chercher dans toute l'image.

La relation entre la représentation projective du point p et sa ligne épipolaire l'_p est projectivement linéaire, du fait que les relations entre p et $\langle C, P \rangle$ d'une part, et $\langle C, P \rangle$ et sa projection l'_p d'autre part sont toutes les deux projectives linéaires. La matrice F de dimensions 3×3 qui décrit cette correspondance est la *matrice fondamentale*.

Essayant maintenant d'exprimer la contrainte épipolaire en utilisant cette matrice fondamentale, dans le cas des caméras non-calibrées. Pour un point donné p dans la première image, la représentation projective l'_p de sa ligne épipolaire dans la deuxième image peut être écrite sous la forme suivante :

$$l'_p = Fp \quad (3.46)$$

Puisque le point p' , correspondant de p , appartient par définition à la ligne l'_p , nous obtenons :

$$p'^T Fp = 0 \quad (3.47)$$

La matrice fondamentale est un concept clé, qui encapsule l'information géométrique qui relie deux images ou plus. L'estimation de la matrice fondamentale peut être faite en ayant un certain nombre de correspondances entre les deux images. On peut constater, à partir de l'équation (3.47), que chacune des paires mise en correspondance fournit une contrainte linéaire sur F . Ainsi, la matrice fondamentale peut être estimée à partir de 8 paires indépendantes. Cependant, telle qu'elle est définie, F a seulement sept degrés de liberté. Algébriquement, F a neuf coefficients à un facteur d'échelle près, mais la condition du rang 2 impose la contrainte additionnelle $\text{Det}(F) = 0$. Ainsi, F peut être calculé à partir de sept paires mises en correspondance et de la contrainte du rang. Par ailleurs, comme la dernière contrainte est cubique, il y aura généralement trois solutions possibles. Pour une estimation correcte, quelques précautions doivent être prises [Har.00]. Pour les détails du calcul de la matrice fondamentale (voir annexe).

D) DETERMINATION DES PARAMETRES INTRINSEQUES DE LA CAMERA

Rappelons que dans le cas d'une caméra en déplacement dans une scène rigide, la matrice fondamentale résulte de l'influence des paramètres intrinsèques et des déplacements effectués entre les vues. A chaque paire de vues on peut donc associer une matrice fondamentale. Si toutes diffèrent du fait qu'elles résultent de déplacements différents, les paramètres intrinsèques n'ont pas changé, par hypothèse. Ainsi, une fois que la caméra a

effectué au moins trois déplacements, il est possible de déterminer ses paramètres intrinsèques à partir des matrices fondamentales. Si on suppose de plus qu'on a seulement quatre paramètres intrinsèques (ce qui est réaliste du fait de l'orthogonalité de la matrice du capteur de vision) deux déplacements seront suffisants. Le lecteur intéressé aux détails d'extraction des paramètres intrinsèques peut se référer à l'annexe X.

E) DETERMINATION DES PARAMETRES EXTRINSEQUES LIES AUX MOUVEMENTS DES CAMERAS

Si l'on s'en réfère à la calibration photogramétrique, rappelons qu'on travaille dans un repère lié au monde, en utilisant une mire 3D présente dans l'environnement de la caméra. On suppose d'ores et déjà connaître les coordonnées de certains points de référence (appartenant à cette mire) relativement à un repère lié à cette mire de calibration. Dans ce contexte, les paramètres extrinsèques représentent le déplacement permettant de passer de ce repère à celui de la caméra. A l'inverse, on n'utilise pas de mire 3D lors d'une auto calibration. Par conséquent, tous les calculs se font par rapport au repère de la caméra (par exemple, celui de la caméra gauche pour un système de stéréovision ou le repère lié à la première position de la caméra pour un dispositif de prise de vue unique se déplaçant dans une scène rigide). Dans le cadre de cette approche, les paramètres extrinsèques décrivent le déplacement entre la première et la deuxième caméra ou entre la première et la deuxième position de prise de vue, suivant le cas. Il est intéressant de rappeler, qu'à ce niveau, on ne possède pas d'informations métriques sur la scène observée. Par conséquent, le déplacement est calculé à un facteur d'échelle près. Néanmoins, il a été démontré [Pol.99, Stu.97] que la reconstruction métrique (entendu par ces auteurs comme le rapport entre les longueurs et les angles entre les droites) est faisable à partir de la connaissance des dimensions d'un objet de la scène.

Toujours dans le contexte de l'auto calibration, nous supposons que nous disposons des paramètres intrinsèques C de la caméra. A partir des équations (3.47), on aura :

$$E = C^T \cdot F \cdot C \quad (3.48)$$

Par ailleurs, nous avons (voir annexe) :

$$E = T \times R \quad (3.49)$$

où R et T sont respectivement la matrice de rotation et le vecteur de translation qui définissent le déplacement entre les deux caméras.

En utilisant la décomposition en valeurs singulières [Pre.88], nous aurons :

$$E = UDV^T \text{ avec, } D = \begin{bmatrix} r & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & q \end{bmatrix} \quad (3.50)$$

Huang et Faugeras [Hua.89] ont démontré que la matrice essentielle E doit avoir deux valeurs singulières non-nulles et égales, et une troisième valeur singulière égale à 0. Hartley [Har.99] a démontré cette propriété autrement. Les valeurs obtenues en pratique peuvent s'éloigner de ces valeurs théoriques du fait des imperfections des mesures, des calculs numériques et des erreurs de modélisation. Une façon de pallier ces problèmes consiste à mettre la valeur de q à zéro et remplacer les valeurs de r et s par leur moyenne. Ceci donne pour D :

$$D = \begin{bmatrix} \frac{r+s}{2} & 0 & 0 \\ 0 & \frac{r+s}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (3.51)$$

La matrice essentielle peut encore s'écrire :

$$E = U \cdot D \cdot G \cdot Z \cdot V^T \quad (3.52)$$

avec :

$$G = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{et} \quad Z = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

De l'équation (3.52) et du fait que $E = T \times R$, la matrice de rotation R et le vecteur de translation T peuvent être calculés grâce à :

$$\begin{cases} R = UGV^T \\ [T]_x = VZV^T \end{cases} \quad (3.53)$$

Avec $[T]_x$, la matrice antisymétrique construite à partir du vecteur de translation T .

En adoptant la méthode décrite ci-dessus, on peut calculer R_1 et T_1 pour la première caméra à partir de la matrice fondamentale F_{11} . De la même façon, on peut calculer R_2 et T_2 associés à la deuxième caméra à partir de la matrice fondamentale F_{22} . Cependant, il reste à calculer la matrice de rotation R_{21} et le vecteur de translation T_{21} du déplacement entre la première et la seconde caméra. Etant donné qu'on dispose des paramètres intrinsèques des deux caméras C et C' , on peut facilement calculer la matrice essentielle $E_{21} = C'^T F_{21} C$. Ainsi, R_{21} et T_{21} se calculent à partir de E_{21} de la même manière que (R_1, T_1) et (R_2, T_2) .

III.3.3.4 LA CALIBRATION D'UN CAPTEUR STEREOSCOPIQUE

La calibration d'un capteur stéréoscopique consiste principalement à déterminer une matrice de rotation R_s et un vecteur de translation T_s exprimant la transformation entre le repère de la caméra gauche et celui de la caméra droite, comme représenté sur la Figure III.8.

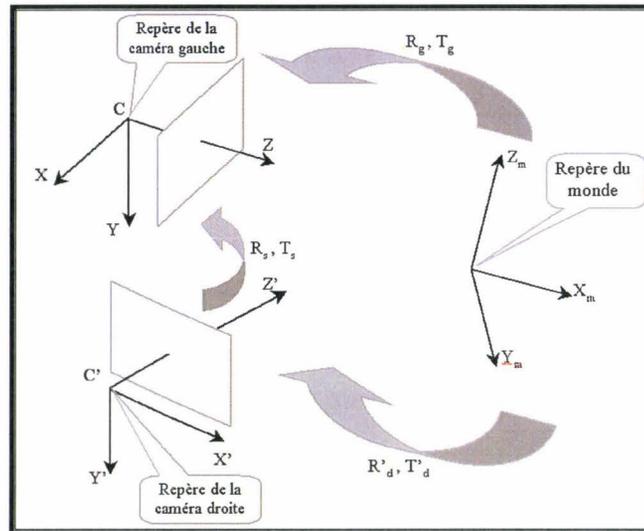


Figure III.12 : Les transformations rigides dans le cas d'un capteur stéréoscopique

Comme le suggère la figure précédente, une démarche valable consiste à calibrer dans un premier temps chacune des deux caméras séparément. Ainsi, nous disposons des paramètres intrinsèques et extrinsèques des deux caméras. Dans une seconde, il suffit de déterminer la transformation qui relie les repères des deux caméras, c'est à dire de déterminer R_s et T_s .

La méthode appliquée pour y parvenir est la suivante : soit P un point de la scène observé par les deux caméras. En supposant que nous disposons des transformations reliant le repère de la scène à celui associé à chacune des deux caméras, le point P sera positionné dans le repère de la caméra gauche (P_g) et dans celui de la caméra droite (P_d) suivant:

$$\begin{aligned} P_g &= R_g P + T_g \\ P_d &= R_d P + T_d \end{aligned} \quad (3.54)$$

Avec :

- R_g et R_d : Les matrices de rotation entre le repère de la scène et les repères des caméras gauche et droite, respectivement.
- T_g et T_d : Les vecteurs de translation entre le repère de la scène et les repères des caméras gauche et droite, respectivement.

En substituant dans (3.54), nous pouvons exprimer les coordonnées du point P de la scène dans la référence de la caméra gauche grâce à :

$$P_g = R_s^T P_d + T_s \quad (3.55)$$

$$\text{ayant posé } \begin{cases} R_s = R_d R_g^T \\ T_s = T_g - R_s^T T_d \end{cases}$$

L'équation (3.55) nous permettra d'exprimer les coordonnées d'un point connu dans le repère d'une caméra dans le repère de l'autre, ce qui est primordial pour la localisation tridimensionnelle.

Dans le cas particulier de la configuration géométrique de notre capteur stéréoscopique (plans rétine des deux caméras confondus, perpendiculaires au plan défini par les deux axes optiques et perpendiculaire au plan de la route) la matrice de rotation R_s et le vecteur de translation T_s correspondant s'exprimeront théoriquement de la façon suivante :

$$R_s = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.56)$$

$$T_s = (b \ 0 \ 0)^T \quad (3.57)$$

Les expressions de (3.56) et (3.57) reflètent le fait que les origines des repères des deux caméras du stéréoscope appartiennent au même plan, se trouvent à des distances égales de la chaussée et ont la même orientation. Ainsi, le déplacement permettant d'aller de l'une à l'autre se réduit à un déplacement le long de l'axe des X (voir la Figure III.7).

La flexibilité de la technique de la calibration hybride est due essentiellement à la mire utilisée pour la calibration. En effet, l'utilisation d'une mire 2D (mire plane) permet une grande mobilité et le choix des motifs constituant cette dernière procure un degré de liberté supplémentaire. En ce qui nous concerne, l'utilisation d'un *damier* dont les dimensions sont connues nous permet d'effectuer la calibration en ligne avec une grande souplesse d'utilisation, sans intervention humaine autre que celle consistant à lancer la procédure et à déplacer la mire devant le système de vision. En effet, comme nous aurons l'occasion de le préciser, l'extraction des points caractéristiques (les coins, dans ce cas de figure) et le calcul du modèle correspondant se feront automatiquement.

L'objectif des paragraphes qui suivent consiste à décrire la procédure que nous avons mise en œuvre pour calibrer le capteur stéréoscopique. Une première partie se consacrera à une étape d'initialisation nécessaire à une mise œuvre efficace de la calibration, à savoir ici l'extraction des points caractéristiques de la mire de calibration. La seconde partie traitera de

la phase de calibration stéréoscopique proprement dite. Le processus global de calibration qui découle de notre méthode a été soumis à une batterie de tests, dont les résultats sont présentés et discutés à la fin de cette section.

III.3.4. MISE EN ŒUVRE DE LA CALIBRATION HYBRIDE

Le cœur de la méthode repose sur l'exploitation d'une mire 2D dont les coordonnées de certains points caractéristiques sont connues par rapport à un référentiel bidimensionnel lié à cette mire. Afin de pouvoir calibrer le système, la mire doit être observée de différents points de vue et avec des orientations différentes (des simples translations sont insuffisantes, comme indiqué précédemment). Dans ce cas, les axes du repère lié à la mire sont choisis de telle sorte que $Z = 0$ pour tout point appartenant au plan de cette dernière (voir, la figure III.13).

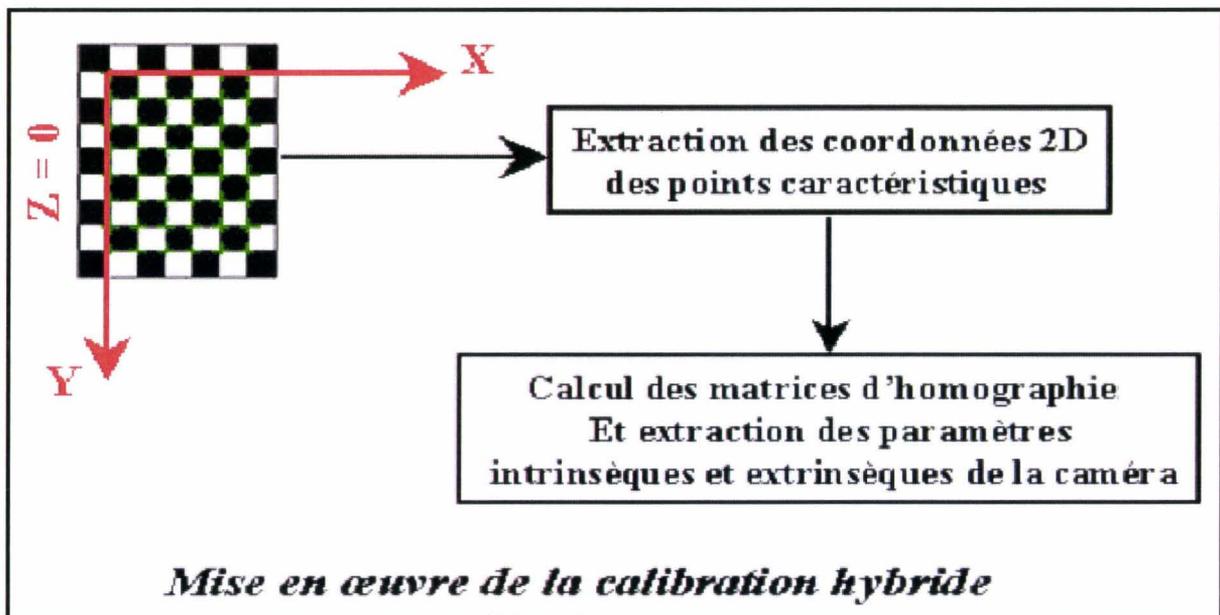


Figure III.13 : Les étapes nécessaires à la mise en œuvre de la calibration hybride.

En ce qui concerne les motifs portés par la mire, plusieurs types sont envisageables dans ce cas de figure. Il apparaît toutefois plus judicieux de privilégier ceux permettant une extraction des points caractéristiques aisée et rapide à réaliser (de sorte à pouvoir s'effectuer en temps réel). Dans cette optique, le damier (noir et blanc) présente des propriétés intrinsèques très intéressantes, puisque la disposition géométrique et colorimétrique des rectangles constituant la mire permettent de la distinguer facilement dans une scène, quelle que soit sa complexité (hormis bien entendu le cas d'un environnement comportant lui aussi d'autres damiers). La section qui suit détaille la façon dont les points caractéristiques sont extraits de la mire de calibration ainsi constituée.

A) DETECTION DE LA MIRE DE CALIBRATION ET L'EXTRACTION DE SES POINTS CARACTERISTIQUES

Le but principal de cette étape est d'extraire les points caractéristiques de l'image de la mire sur le plan rétine de la caméra pour pouvoir les associer à leurs correspondants sur le damier, dont les coordonnées sont exprimées dans le référentiel lié à la mire (nous avons choisi le point haut gauche comme étant l'origine de la mire $(0 \ 0 \ 0)^T$). En matière de points caractéristiques, ceux-ci seront constitués par les coins des carrés constituant le damier de la mire.

L'extraction de ces coins peut être conduite *a priori* de trois manières différentes, à savoir :

- Extraction *manuelle*
- Extraction *semi-automatique*
- Extraction *automatique*

L'extraction manuelle consiste à laisser le soin à un opérateur humain d'identifier les coins de la mire de calibration (par exemple en cliquant sur les coins en question). Du point de vue de la machine, cette technique n'exige aucun traitement mais demande beaucoup d'effort de la part de l'opérateur, tout en étant très lente et fastidieuse à réaliser vu le nombre de coins que contient une mire et le nombre d'images nécessaires pour la calibration ! Cependant, cette technique manuelle permet l'extraction de l'information nécessaire quelle que soit la disposition de la mire (elle n'est pas contrainte par des limites de traitement numérique ou algorithmique, compte tenu des performances naturelles du cerveau humain par rapport à une tâche aussi simple).

L'extraction semi-automatique consiste, quant à elle, à extraire un certain nombre très limité des coins de la mire. Les autres seront extraits d'une manière automatique en se basant sur les points de référence extraits précédemment et en exploitant la connaissance préalable de la géométrie de la mire utilisée. Cette technique n'exige pas de traitements très élaborés et est moins « prenante » que la précédente vis-à-vis de l'opérateur, mais mobilise cependant l'attention de ce dernier pendant un temps non-négligeable et requiert de sa part un certain niveau « d'expertise » (quels points de référence choisir et dans quel ordre ?), ce qui constitue un handicap.

Enfin, l'extraction automatique de coins n'exige pas d'intervention de l'opérateur autre que celle consistant à présenter la mire au système et présente donc une très grande flexibilité ainsi qu'une rapidité inégalable pour ce qui touche à l'extraction des coins. Cependant, cette approche nécessite l'implication inévitable de techniques de traitement d'image et l'exploitation d'un algorithme spécifique à cette tâche, dont les performances doivent être au minimum comparables à celles de l'opérateur humain.

En ce qui nous concerne, nous avons opté pour cette dernière approche afin de pouvoir bénéficier de la flexibilité qu'elle fournit. Les étapes mises en œuvre pour détecter et extraire les coins caractéristiques de la mire sont :

1. La binarisation à partir de l'image en niveaux de gris.
2. La détection des rectangles dans l'image binaire.
3. L'extraction des coins.
4. L'organisation des coins détectés (en prenant comme référence le coin supérieur gauche du damier) afin d'effectuer la mise en correspondance point de l'image / point de la mire.

Dans ce qui suit nous allons détailler ces étapes dans l'ordre où elles sont citées. Chacune d'entre elles sera illustrée par des résultats expérimentaux.

III.3.4.1. BINARISATION DE L'IMAGE

Etant donné que les images acquises par les caméras du capteur stéréoscopique sont des images en couleurs (R, G, B) , la première étape de la binarisation consiste à convertir l'espace de représentation de ces images en niveaux de gris (256 niveaux). La binarisation consiste à réduire davantage l'espace de représentation, tel que chaque pixel soit représenté soit par un 0, soit par un 1. Cette représentation est intéressante, dans la mesure où le damier contient des rectangles blancs et des rectangles noirs et une telle représentation binaire permettra de distinguer clairement les rectangles du damier. A cette fin, nous procédons au calcul du seuil de binarisation tel que :

$$\text{Si } I(i, j) > S_{bin} \text{ alors : } B(i, j) = 1$$

$$\text{Sinon, } B(i, j) = 0$$

Avec : $I(i, j)$: la valeur en niveau de gris du pixel de coordonnées (i, j) .

S_{bin} : le seuil de binarisation compris, entre 0 et 255.

$B(i, j)$: la valeur du pixel de coordonnées (i, j) dans l'image binaire.

Le calcul du seuil de binarisation proprement dit peut se faire de différentes manières. Dans des scènes où la dynamique de la luminance est contrôlée, une valeur fixe telle que 128 permet de différencier clairement le blanc du noir dans une image. A l'inverse, les scènes extérieures posent un certain nombre de problèmes engendrés particulièrement par la dynamique du capteur vis à vis des changements brusques de la lumière. Pour contourner cette difficulté, nous procédons au calcul dynamique du seuil de binarisation en nous basant sur la répartition des niveaux de gris dans l'image. Le but principal de cette opération est de mettre en exergue les propriétés colorimétriques du damier afin de préparer l'image pour des traitements ultérieurs. L'idée est que le seuil de binarisation doit faire apparaître clairement les rectangles constituant le damier et cela quelle que soit la plage des niveaux de gris utilisée pour la présentation du contenu de l'image.

Pour y parvenir, nous procédons au calcul de l'histogramme de l'image en niveaux de gris. L'analyse de cet histogramme fait apparaître particulièrement deux pics qui se situent généralement de part et d'autre de la valeur médiane des niveaux de gris (la valeur 128), nous appelons ces deux pics le *pic blanc* et le *pic noir* respectivement. Le seuil de binarisation choisi est la valeur du niveau de gris correspondant au minimum global entre ces deux pics.

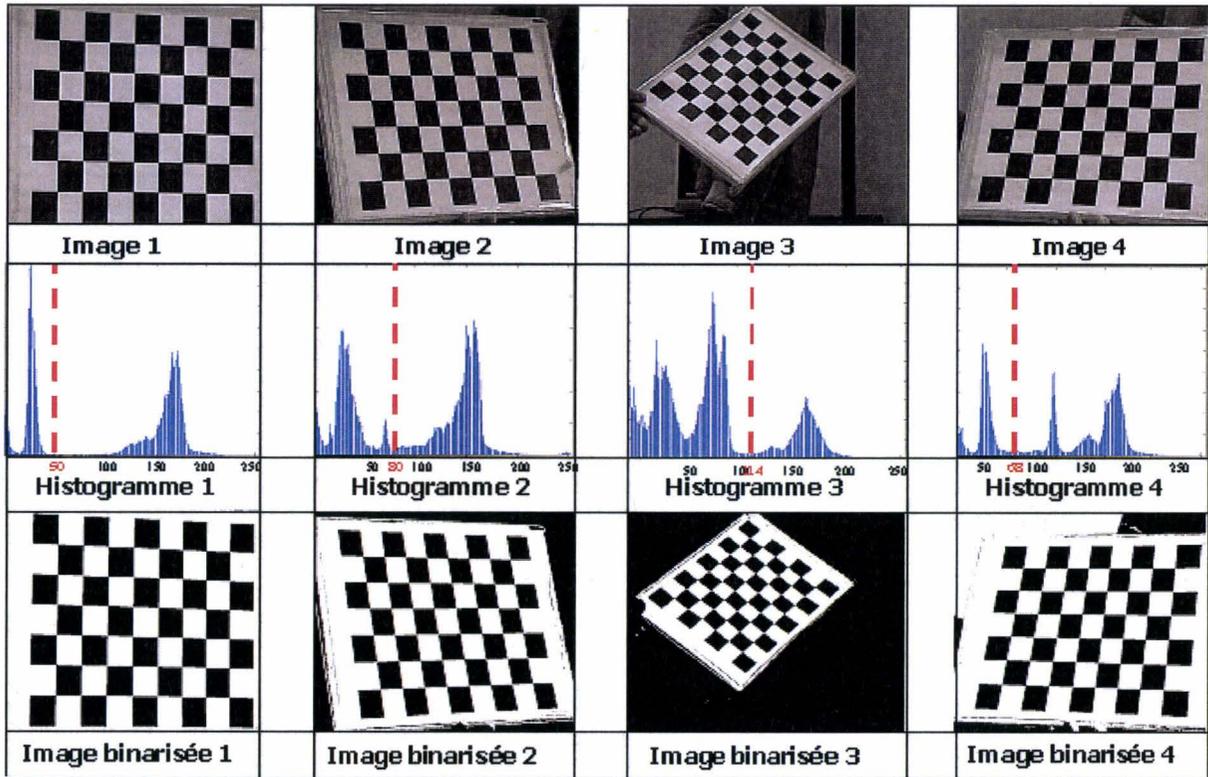


Figure III.14 : Exemples d'images de la mire de calibration et des histogrammes utilisés pour calculer les seuils de binarisation.

La figure III.14, présente quelques images et les histogrammes correspondants, ainsi que les résultats obtenus après seuillage. Le seuillage des images en adoptant la technique décrite précédemment fait apparaître clairement les rectangles constituant le damier.

III.3.4.2. DETECTION DES RECTANGLES ET EXTRACTION DES COINS DE LA MIRE

Suite à l'étape de la binarisation, il s'agit maintenant de détecter les rectangles du damier avant de procéder à la phase d'extraction des coins. En effet, nous aurions pu avoir recours directement à une technique d'extraction des coins, telle que celle de Harris et Stephans [Har.88], mais ce procédé ne nous permettrait pas de différencier les coins qui appartiennent au damier de ceux qui appartiennent aux autres objets de la scène. Même si la prise en considération de certaines contraintes (comme les distances séparant les différents coins) améliore les résultats, cela reste toutefois insuffisant. Nous avons donc plutôt préféré

l'extraction des coins à travers la détection des rectangles (seuls les coins de ces dernières entités nous intéressent).

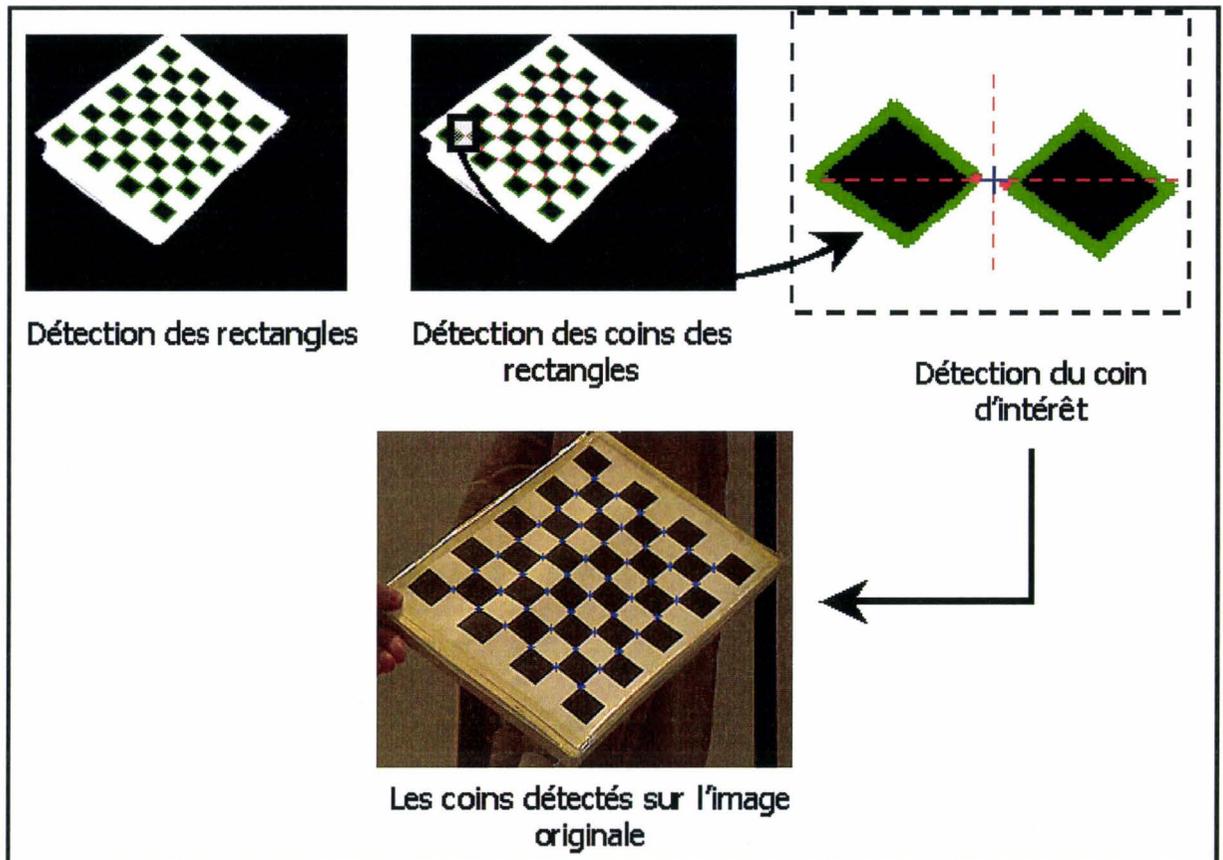


Figure III.15 : Détection des rectangles et extraction des coins d'intérêt de l'image binarisée

Pour y aboutir, nous commençons par l'application d'une dilatation morphologique [Par.96b]. Celle-ci est bénéfique dans la mesure où elle permet la séparation des différentes formes et facilite ainsi la tâche de la détection des rectangles constituant le damier. La deuxième étape consiste à détecter et à extraire les contours de l'image; les contours ouverts ou les petits périmètres seront systématiquement rejetés. Une approximation polynomiale des contours s'avère généralement nécessaire [Dou.73, Her.92] afin de pallier les imperfections de la binarisation et de l'extraction. Ainsi, les contours représentant les rectangles étant classés de haut en bas et de gauche à droite selon leur apparition dans l'image, l'étape suivante consistera à détecter les points des rectangles qui sont les plus proches les uns des autres et qui correspondent aux coins après dilatation. En effet, chaque coin dans l'image originale sera éclaté en deux coins après dilatation. Par conséquent, les coordonnées du coin original est la moyenne des coordonnées des deux points les plus proches entre deux rectangles adjacents (voir la Figure III.15).

III.3.4.3. ORGANISATION DES COINS DETECTES

Après avoir détecté les coins à l'issue de l'étape précédente, la phase suivante consiste à organiser les coins détectés par rapport à l'origine d'un référentiel donné, de telle manière

qu'on puisse mettre en correspondance les coordonnées images des coins et leurs coordonnées par rapport au référentiel 2D du monde d'une part (Figure III.13), et afin de pouvoir mettre en correspondance les coordonnées images issues des images de la mire prises à différentes positions.

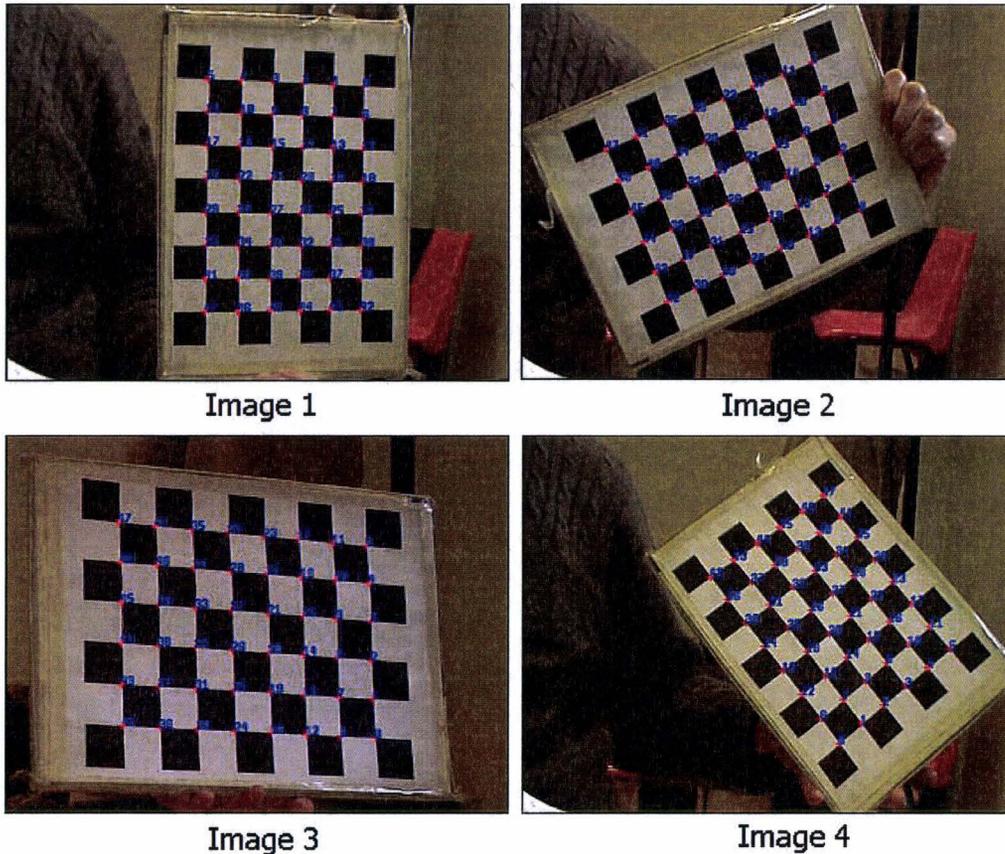


Figure III.16 : Quatre images du damier 7×9 utilisé comme mire de calibration effectuant une rotation autour de l'axe des Z de la mire ; les coins sont ordonnés de la gauche vers la droite et de haut vers le bas dans le repère de la mire.

L'organisation des points caractéristiques se fait en calculant la distance de ceux-ci par rapport à l'origine du référentiel de l'image de la mire (le coin supérieur droit de la mire (Figure III.16 (image 1))), de sorte que les coordonnées image des coins soient triées de droite à gauche et de haut en bas (selon la Figure III.16).

Pour pouvoir ordonner les coins de la mire correctement nous devons donc procéder au calcul des coordonnées des lignes verticales et horizontales matérialisées par les différents coins de l'image de la mire. Ceci permettra de distinguer l'axe des x de celui des y . Par ailleurs, le produit vectoriel entre ces deux axes permettra de déterminer lequel des quatre coins extrêmes de l'image de la mire est la projection de l'origine de cette dernière sur

l'image (qui est, rappelons le, le coin supérieur gauche du damier) comme le montre la figure III.13.

III.4.4.4. EVALUATION DES PERFORMANCES DE LA CALIBRATION DU SYSTEME STEREOSCOPIQUE

Afin d'évaluer les performances et d'établir les conditions d'utilisation optimum de la technique de calibration que nous utilisons (à savoir la calibration *hybride*) nous avons procédé à des expériences sur des données simulées. Il est clair que la calibration avec des données synthétiques nous permet d'avoir une idée précise sur les performances potentielles de la technique, en autorisant la comparaison entre les paramètres estimés et ceux utilisés pour générer les images de la mire (constituant les paramètres de la simulation).

Pour évaluer l'influence du bruit sur les points images des points de la mire, nous avons simulé un capteur stéréoscopique dont la configuration géométrique correspond parfaitement à celle du système de stéréovision utilisé en réalité, observant $n=13$ images d'une mire constituée de 80 points de contrôle distribués régulièrement, espacés de 50 unités (l'unité étant égale à un millimètre). Cette disposition est représentée sur la figure III.17.

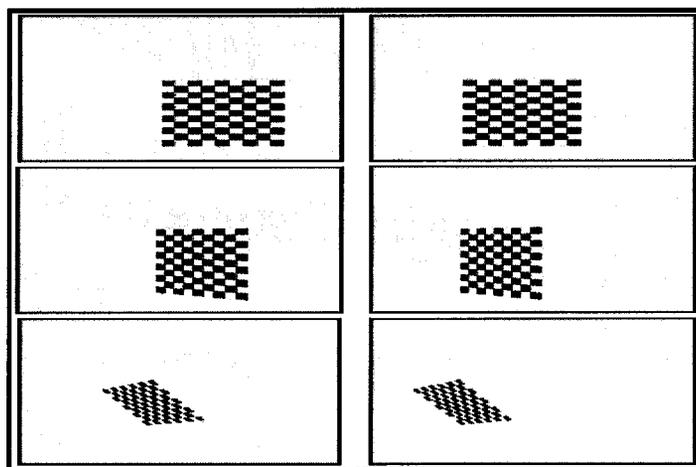


Figure III.17 : Trois paires d'images synthétiques gauche et droite de la mire 8×10 observées par le capteur stéréoscopique simulé.

Les paramètres intrinsèques et les coefficients de distorsion (on ne considère ici que les distorsions radiale et tangentielle) des deux caméras sont donnés dans le Tableau III.4.

	u_0	v_0	α_u	α_v	k_1	k_2	p_1	p_2
Caméra gauche	286	286	1200	1200	0.1	0.01	0.01	0.01
Caméra droite	286	286	1200	1200	0.1	0.01	0.01	0.01

Tableau III.4 : Paramètres intrinsèques et paramètres de distorsion radiale des deux caméras simulées.

Pour chaque point de vue, l'orientation de la mire (le damier) par rapport à la caméra gauche, est représentée par les angles d'Euler en degrés. Sa position est représentée par un vecteur 3D en millimètres (voir le Tableau III.5). Les rotations et les translations reliant les deux caméras du capteur stéréoscopique simulé sont précisées dans le Tableau III.6.

	Rotations en degré			Translations en millimètre		
	θ_x	θ_y	θ_z	T_x	T_y	T_z
Vue - 1	0	0	0	-400	-400	4600
Vue - 2	30	0	0	-400	-600	4700
Vue - 3	0	30	0	230	400	4800
Vue - 4	0	0	30	500	-340	4900
Vue - 5	30	30	30	210	20	5000
Vue - 6	-20	15	10	-60	-370	5100
Vue - 7	-43	11	2	53	42	5200
Vue - 8	12	3	7	90	-120	5300
Vue - 9	-6	23	45	260	-80	5400
Vue - 10	45	45	45	-350	-500	5500

Tableau III.5 : Les paramètres extrinsèques reliant les vues à la caméra gauche.

Rotations en degré			Translations en millimètre		
θ_x	θ_y	θ_z	T_x	T_y	T_z
0	0	0	400	0	0

Tableau III.6 : La position et l'orientation relatives des deux caméras (cette configuration correspond à la configuration particulière décrite dans III.3.2.2)

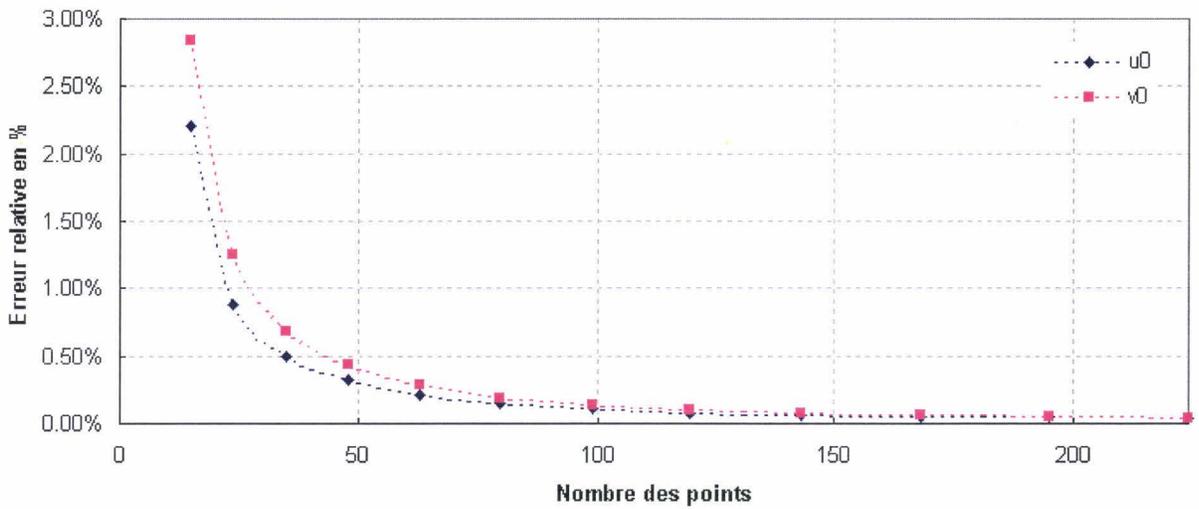
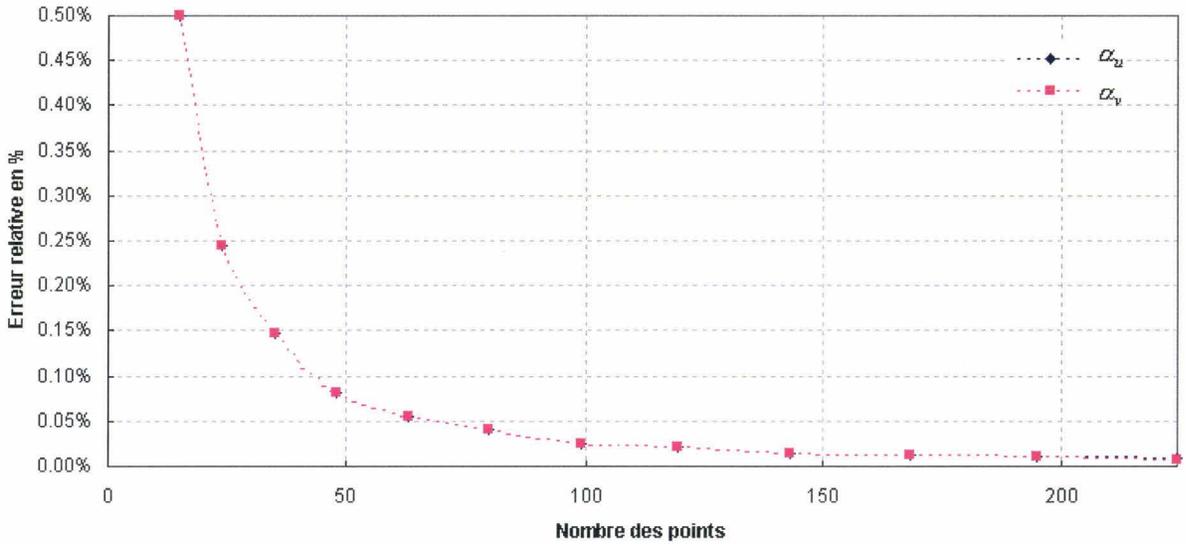
Au départ les paramètres usités pour la génération des images synthétiques sont utilisés afin d'estimer l'erreur entre les paramètres estimés et les paramètres réels d'une part et afin d'estimer l'écart type des paramètres calculés sur plusieurs procédures de calibrations successives d'autre part. L'évaluation de la précision de la technique de calibration mise à l'épreuve se base essentiellement sur l'examen des paramètres intrinsèques et extrinsèques délivrés par la méthode, plus précisément sur les valeurs définissant la matrice de rotation et le vecteur de translation stéréo.

Nous allons commencer par l'évaluation de l'influence du nombre de points de la mire sur la qualité de la calibration. Ensuite, nous allons nous pencher sur l'influence du nombre

d'images utilisées et, enfin, nous évaluerons l'influence du bruit. L'analyse portera aussi bien sur les paramètres intrinsèques que sur les paramètres extrinsèques.

A) INFLUENCE DU NOMBRE DE POINTS DE LA MIRE

Pour évaluer l'influence du nombre de points de contrôle (les coins du damier) sur l'estimation des paramètres de calibration, nous avons effectué une centaine de calibrations en utilisant à chaque fois des mires constituées respectivement de 15, 24, 35, 48, 63, 80, 99, 120, 143, 168, 195 et 224 points de contrôle.



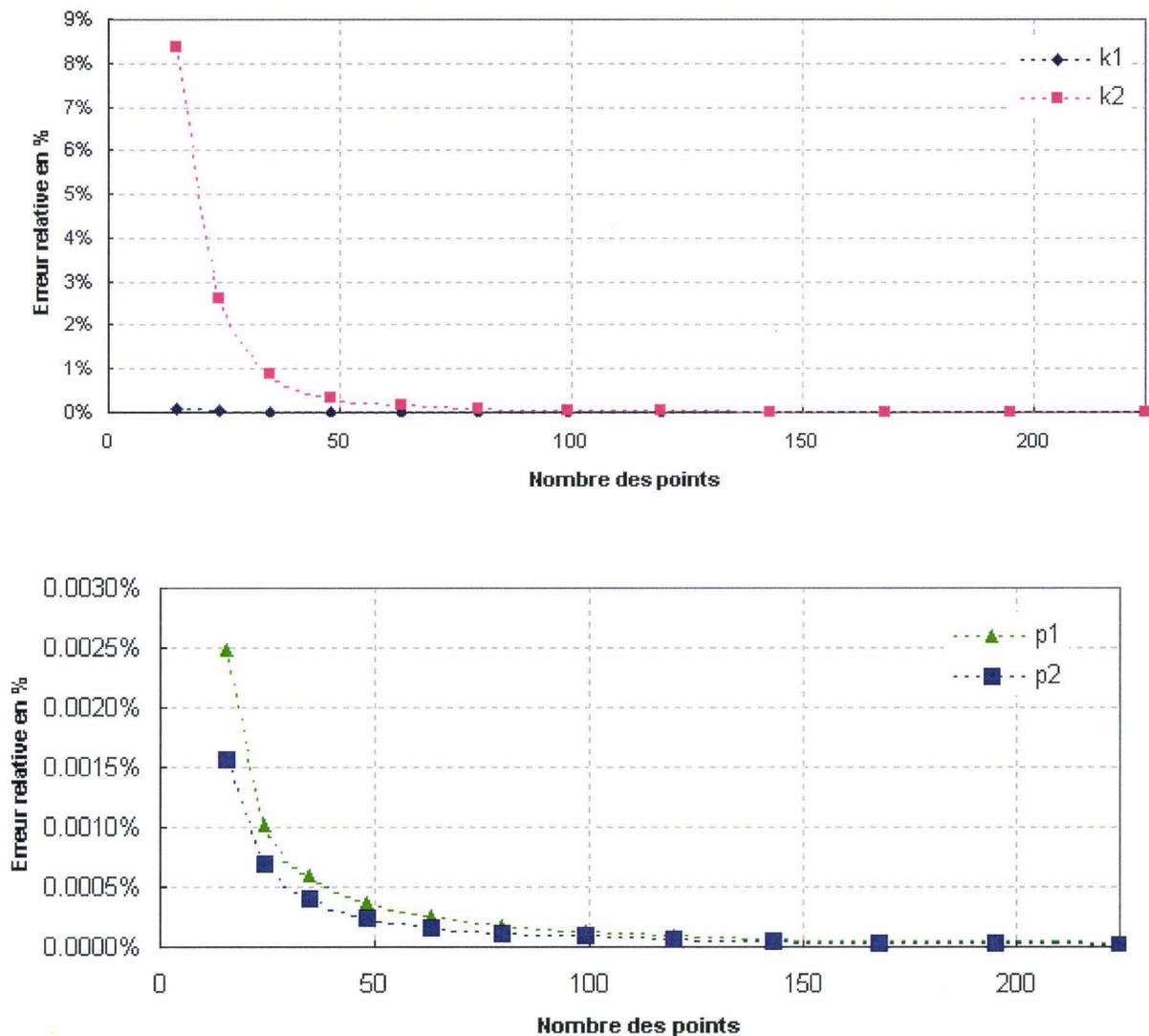


Figure III.18 : Erreurs relatives sur les paramètres estimés après 300 calibrations (avec un bruit ajouté aux coordonnées des points de contrôle de moyenne nulle et d'écart type $\sigma = 0.01$). La simulation porte sur la caméra gauche. La grandeur en abscisse est le nombre de points de contrôle du damier utilisé pour la calibration.

Nous pouvons constater que l'estimation des paramètres intrinsèques de la caméra (α_u, α_v) est la moins sensible au nombre de points utilisés pour la calibration, tandis que les coefficients de distorsion et les coordonnées image du centre de projection le sont davantage. Nous pouvons constater également que l'influence du nombre de points utilisés pour la calibration est plus importante sur l'estimation du coefficient de distorsion radiale k_2 . D'une façon générale, les erreurs relatives d'estimation des paramètres intrinsèques se stabilisent à partir du moment où on calibre avec une mire qui contient plus de 80 points de contrôle. Il semble par ailleurs inutile d'aller nettement plus loin dans le nombre de points de référence, compte tenu du faible accroissement de la précision obtenue en contrepartie.

Dans ce qui suit, nous allons évaluer l'influence du nombre de points de la mire de calibration sur l'estimation des paramètres extrinsèques de la tête stéréo, à savoir les angles d'orientation du repère de la caméra droite par rapport à la caméra gauche et la position du repère de la caméra droite relativement à la caméra gauche.

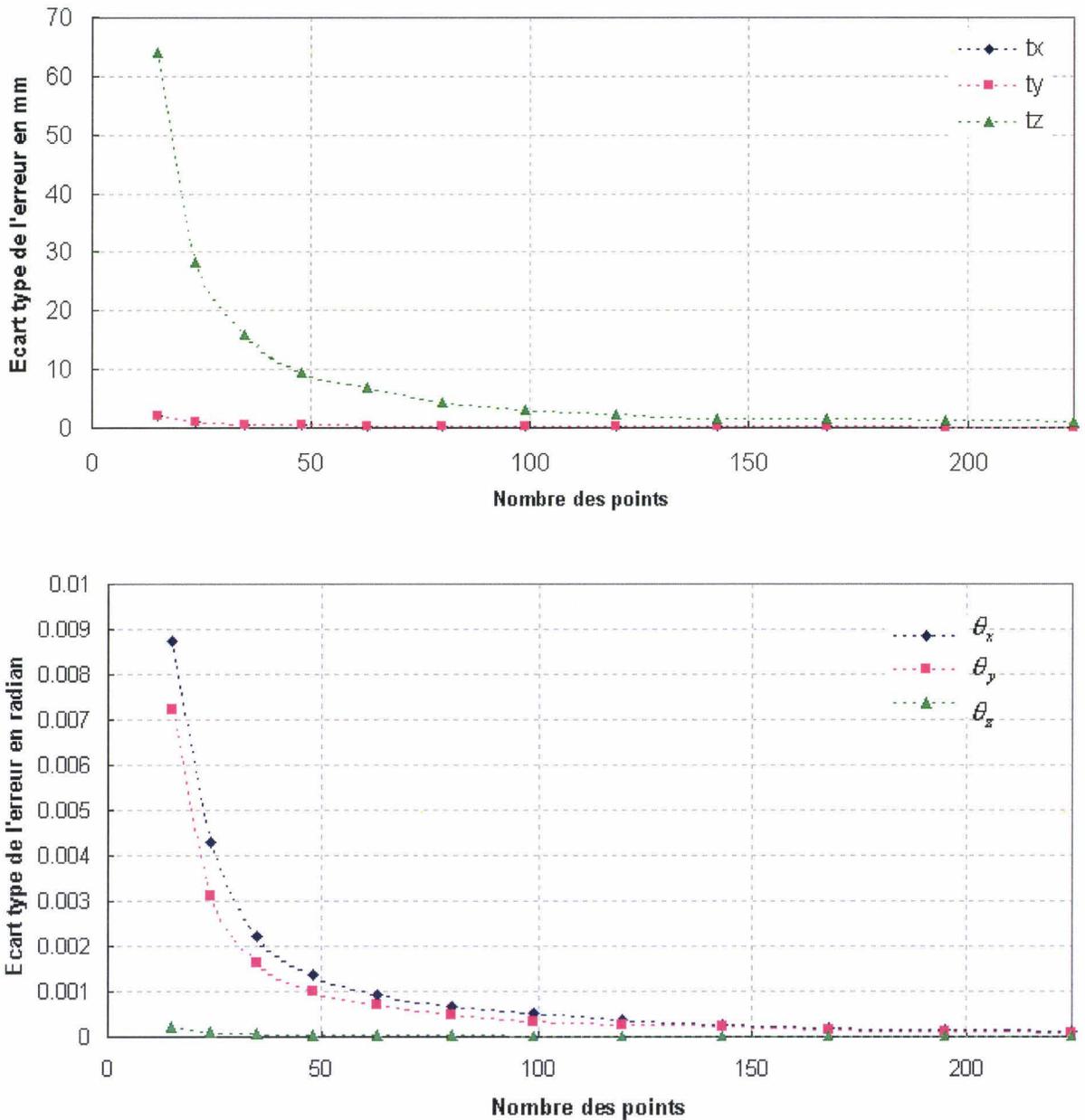


Figure III.19 : Ecarts-types sur l'erreur d'estimation des paramètres extrinsèques du capteur stéréoscopique (angles d'orientations et positions relatives entre les deux repères caméra) en fonction du nombre de points de contrôle de la mire utilisée pour la calibration.

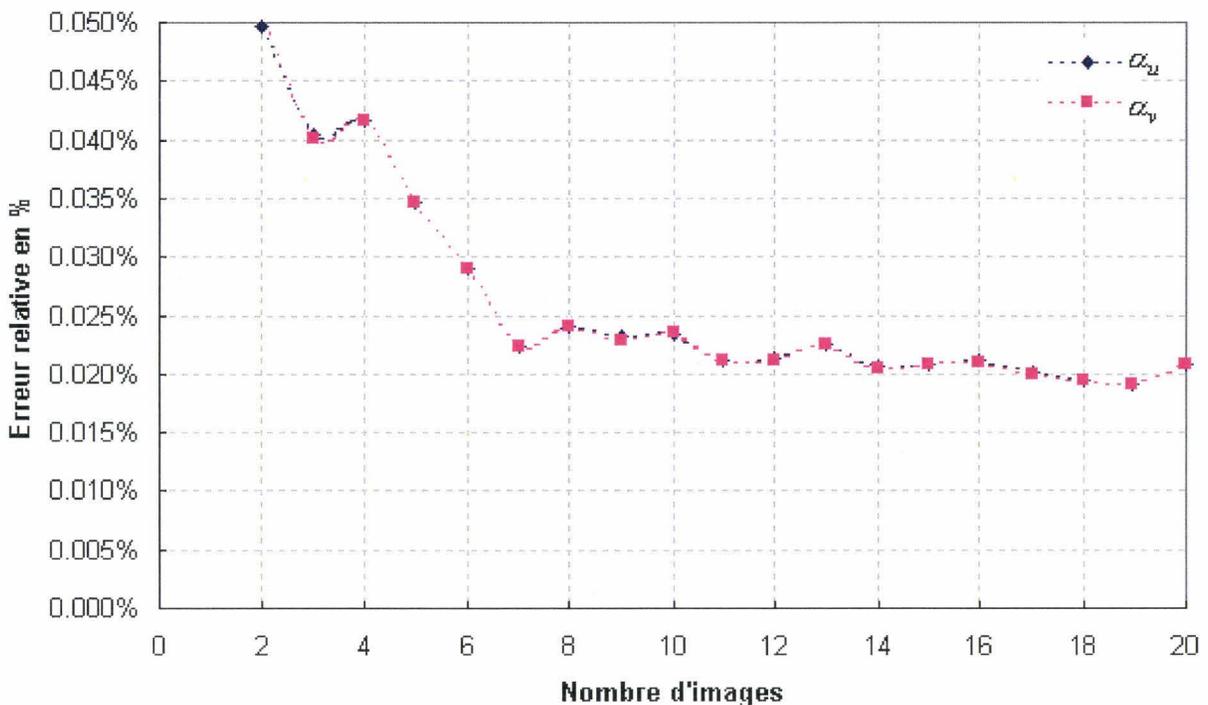
La Figure III.19 représente les écarts-types sur les erreurs d'estimation associées aux paramètres extrinsèques du capteur stéréoscopique en fonction du nombre de coins de la mire de calibration. On s'aperçoit que les écarts-types relatifs aux erreurs d'orientation relatives sont faibles d'une manière générale, tandis que les écarts-types associés aux erreurs de positionnement relatif sont plus importants. Quoiqu'il en soit, les écarts-types de toutes les

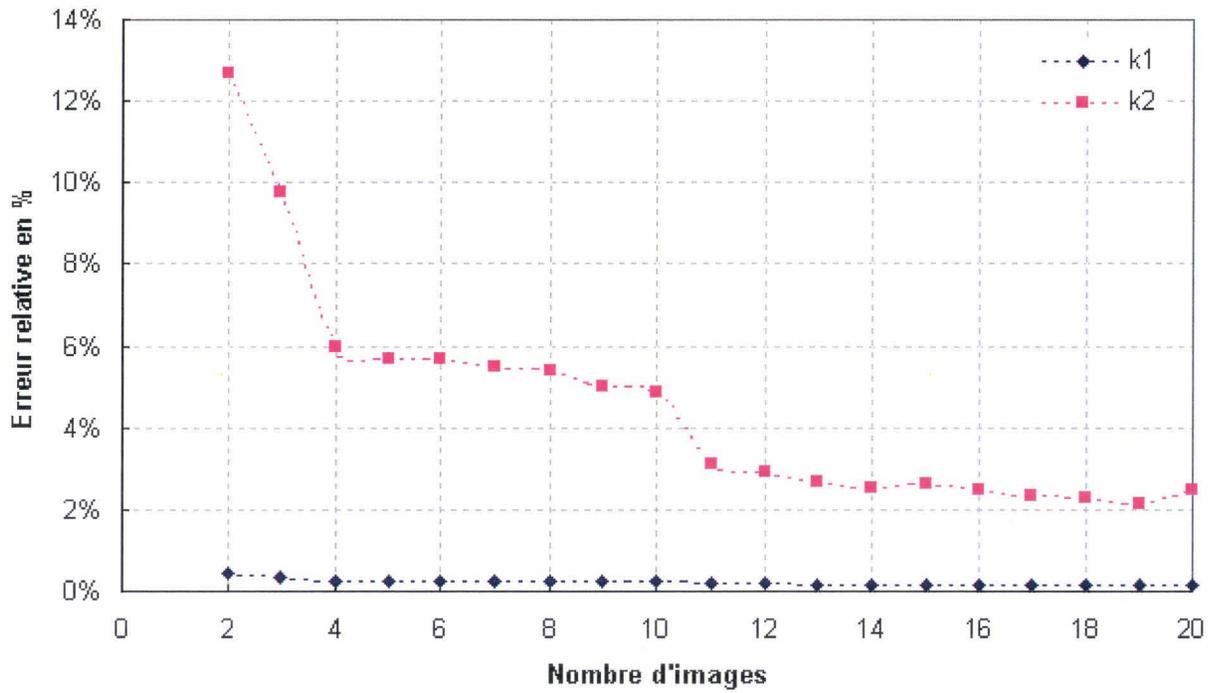
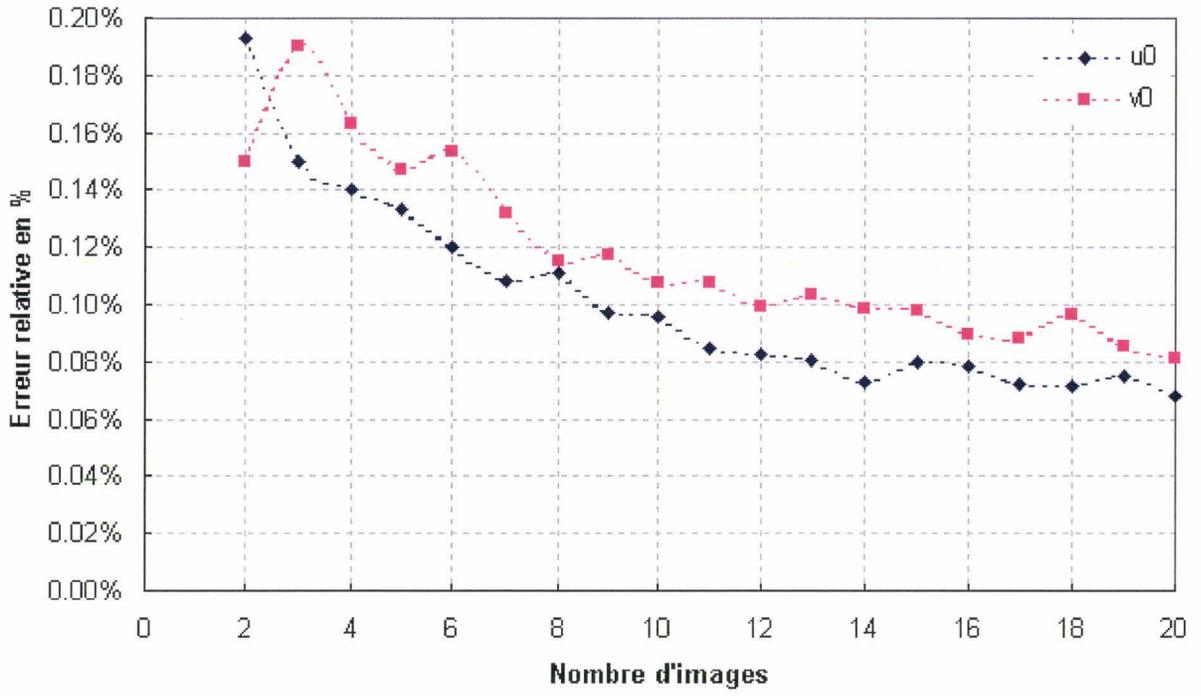
erreurs décroissent avec l'augmentation du nombre de points de contrôle de la mire de calibration. Cette tendance décroissante est plus perceptible sur la coordonnée relative en Z et les orientations relatives autour de X et Y , respectivement. Là encore, la diminution de l'erreur n'est plus réellement significative dès lors qu'on dépasse les 80 points de contrôle.

Les graphiques ci-dessus montrent d'une façon indiscutable l'influence du nombre de points de contrôle sur la qualité des estimations des paramètres intrinsèques et extrinsèques. Il apparaît clairement que le choix d'utiliser une mire comportant approximativement 80 coins paraît judicieux puisque l'influence bénéfique du nombre de points est nettement moins forte au-delà de cette valeur.

B) INFLUENCE DU NOMBRE D'IMAGES UTILISEES POUR LA CALIBRATION

Pour évaluer l'influence du nombre d'image de la mire utilisé pour la calibration, nous avons calibré le système simulé en utilisant de 2 à 20 images en incrémentant le nombre d'images utilisées à chaque itération de la méthode (deux images étant le nombre d'images minimum à utiliser pour effectuer la calibration hybride). La Figure III.15 représente les résultats issus de ces expériences. Les deux premiers graphiques représentent les erreurs relatives commises sur (α_u, α_v) et (u_0, v_0) respectivement. Cette erreur relative a été estimée en utilisant la moyenne des erreurs issues de 300 calculs de calibration, pour un écart type de bruit fixe.





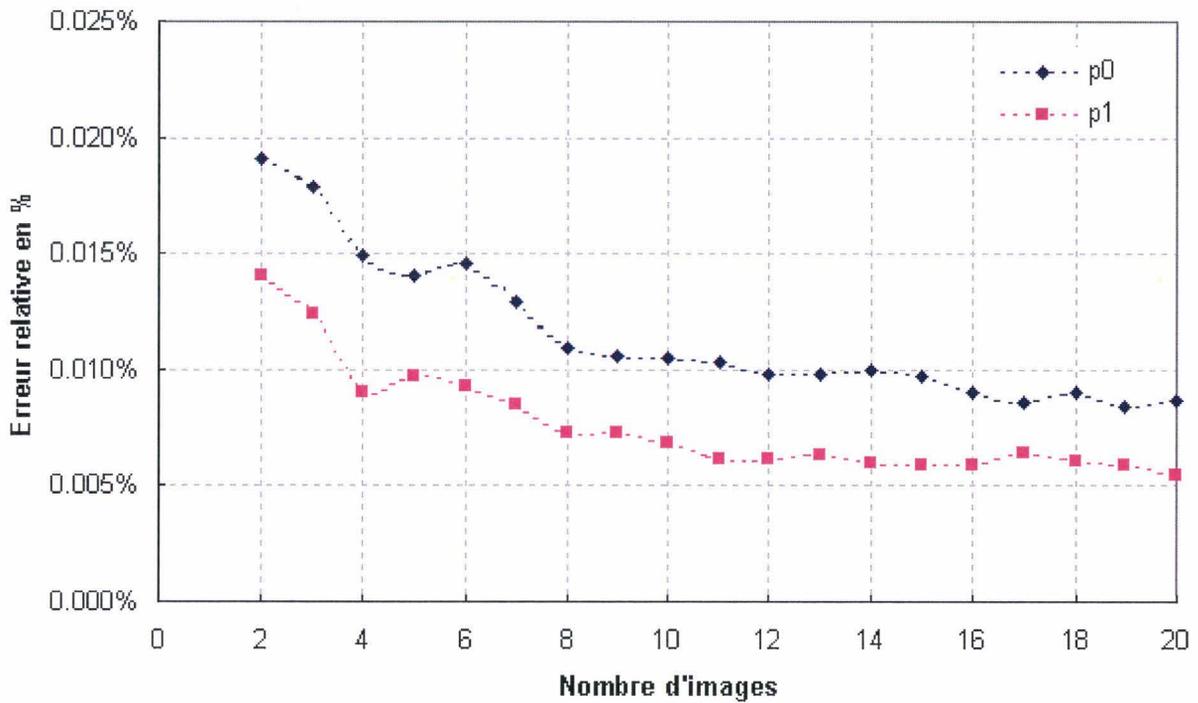


Figure III.20 : Evolution des erreurs relatives sur l'estimation des paramètres intrinsèques de la caméra gauche en fonction du nombre d'images utilisées pour la calibration.

L'erreur relative commise sur l'estimation des paramètres intrinsèques varie en raison inverse du nombre d'images utilisées pour la réalisation de la calibration. Cependant, cette décroissance n'est plus réellement significative dès lors qu'on atteint 13 images. Les erreurs relatives obtenues lors de ces expériences sont très faibles, du fait qu'on a fixé l'écart type de bruit appliqué au point de contrôle à 0.01 (le but de ces expériences était avant tout de montrer l'influence du nombre d'images utilisées pour la calibration sur la qualité de cette dernière).

Il est à noter que l'influence du nombre d'images est plus visible sur l'estimation du coefficient de distorsion radiale k_2 . Cette erreur peut en effet atteindre 13% de la valeur réelle du coefficient lorsqu'il n'est fait usage que de deux images.

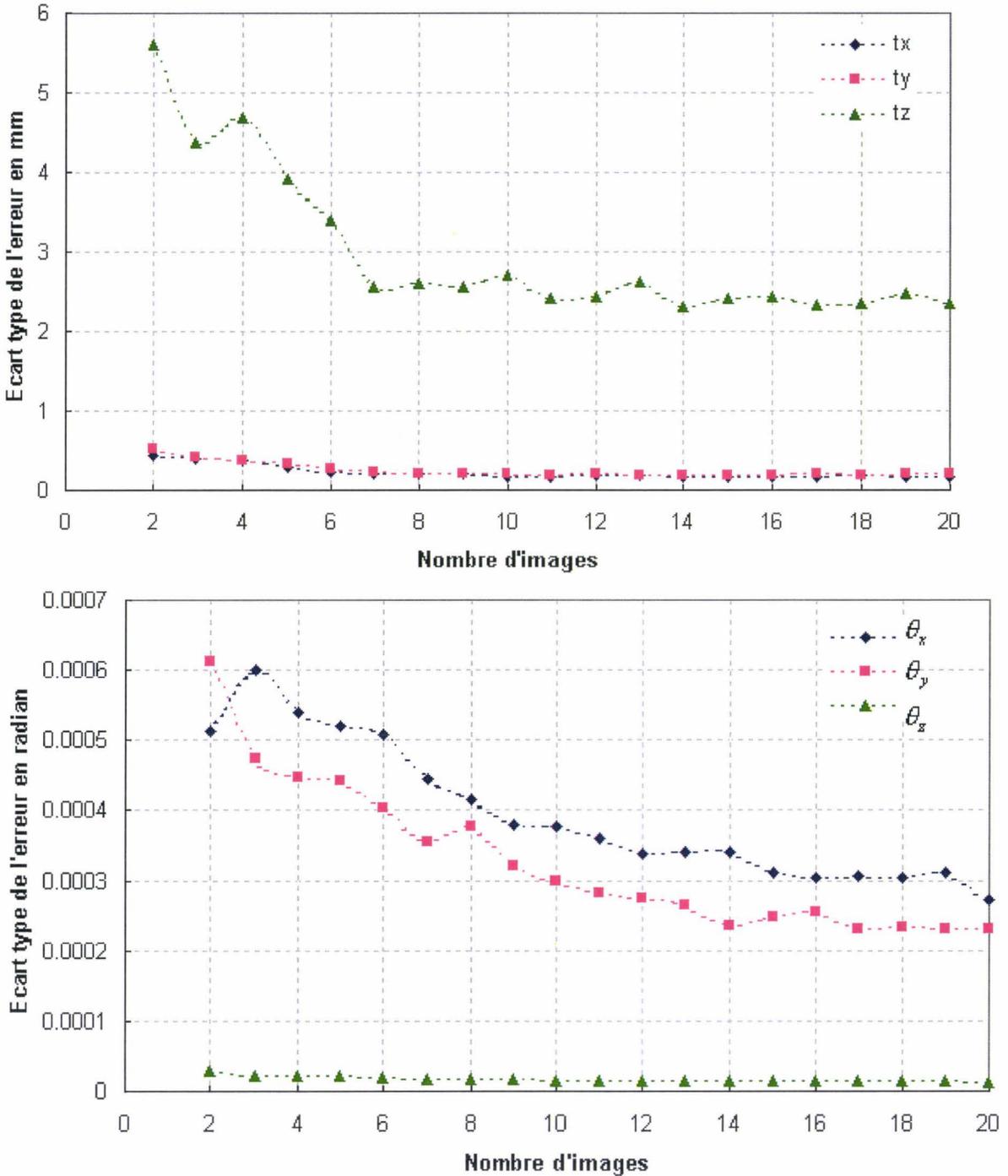


Figure III.21 : Ecarts-types des erreurs d'estimation des paramètres extrinsèques du capteur stéréoscopique (angles d'orientations et positions relatives) en fonction du nombre d'images utilisées pour effectuer la calibration.

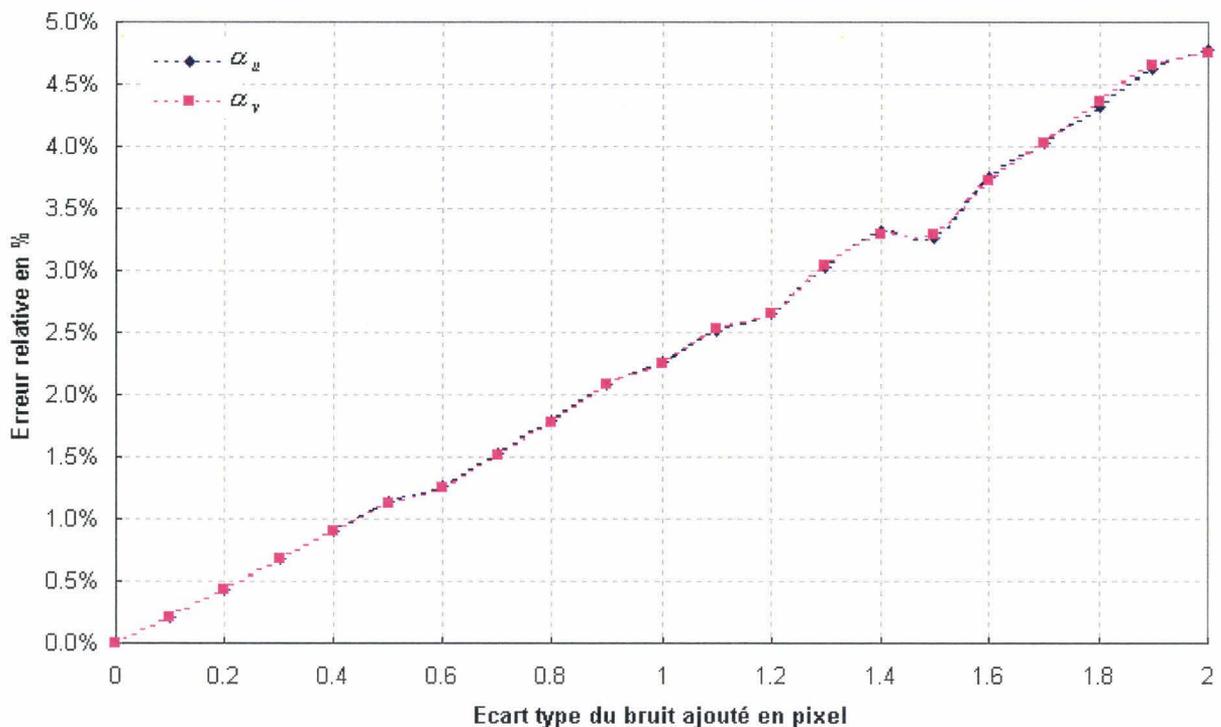
La Figure III.21 représente l'évolution des écarts-types des erreurs relatives à l'estimation des paramètres extrinsèques du capteur stéréoscopique en fonction du nombre d'images utilisées lors de la calibration. Là aussi, on s'aperçoit que les écarts-types des erreurs d'orientation et de positionnement décroissent en fonction de l'augmentation du nombre d'images. Cette tendance est plus marquée sur la coordonnée relative en Z et les orientations relatives autour de X et Y, respectivement. La décroissance de l'écart-type de l'erreur n'est plus significative dès lors qu'on calibre avec plus de 13 images. Il est

intéressant de souligner ici que l'influence du nombre d'images est plus significative que celle du nombre de points de contrôle de la mire. En effet, une augmentation du nombre d'images permet une estimation plus précise de la matrice de rotation et du vecteur de translation stéréo. La raison en est que l'augmentation du nombre d'images (à condition qu'elles correspondent bien à des positions et orientation différentes du damier) génère une augmentation des contraintes sur les paramètres extrinsèques et permet ainsi une meilleure estimation.

Les graphiques ci-dessus illustrent clairement l'influence du nombre d'images de la mire sur la qualité de l'estimation des paramètres intrinsèques et extrinsèques. D'une façon similaire à ce qui se passe pour le nombre de points de contrôle, le nombre d'images cesse d'avoir une influence bénéfique réellement tangible dès lors que celui-ci dépasse un certain seuil (treize images dans notre cas).

C) INFLUENCE DU BRUIT SUR LES POINTS DE LA MIRE SUR L'ESTIMATION DES PARAMETRES

Afin d'estimer l'influence du bruit sur les points de contrôle de la mire (les coins) et, par voie de conséquence, sur l'estimation des paramètres issus de la calibration, nous avons contaminé les points de contrôle avec un bruit gaussien, centré et d'écart type σ variable. Pour chaque valeur de l'écart-type nous avons effectué 300 calibrations consécutives. Les résultats obtenus (synthésisés par la figure III.22) indiquent que l'erreur relative d'estimation sur les paramètres intrinsèques ($\alpha_u, \alpha_v, u_0, v_0$) croît linéairement avec l'augmentation de l'écart type du bruit, pour le domaine de bruit exploré.



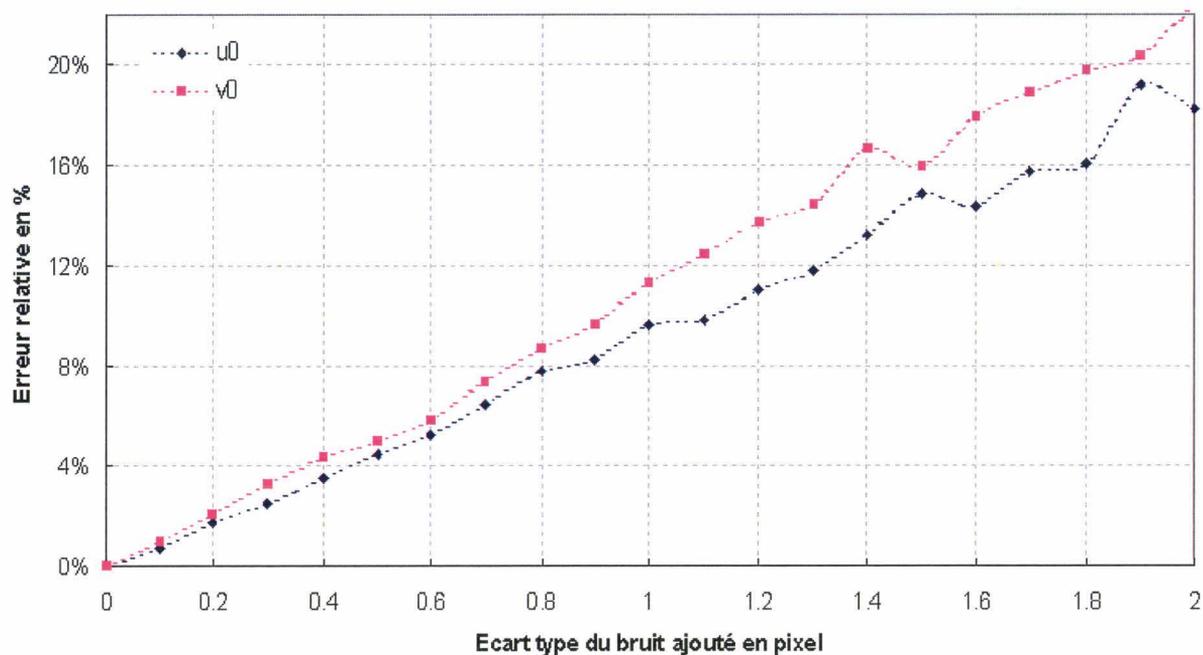


Figure III.22 : Evolution de l'erreur relative sur l'estimation des paramètres de calibration (données moyennes issues de 300 calibrations) de la caméra gauche, en fonction du niveau de bruit appliqué aux positions des projections des points de contrôle de la mire.

La Figure III.22 montre les résultats obtenus pour l'estimation des huit paramètres intrinsèques de la caméra gauche (la caméra droite donne des résultats similaires). Les graphiques ci-dessus permettent de constater que les erreurs relatives sur (α_u, α_v) sont faibles puisqu'elles ne dépassent pas les 5% des valeurs réelles pour un écart type de bruit allant jusqu'à 2 pixels. Ce même bruit a plus d'influence sur l'estimation des coordonnées images $(u_0 \ v_0)^T$ du centre de projection, mais reste toutefois acceptable pour un bruit de 1 pixel (erreur commise inférieure à 10% des valeurs réelles). Le graphique relatif aux coefficients de distorsion n'est pas repris ici puisque celui-ci exhibe les mêmes tendances que celles observées pour les autres paramètres intrinsèques (les mêmes constats sont, bien entendu, valables pour la caméra droite).

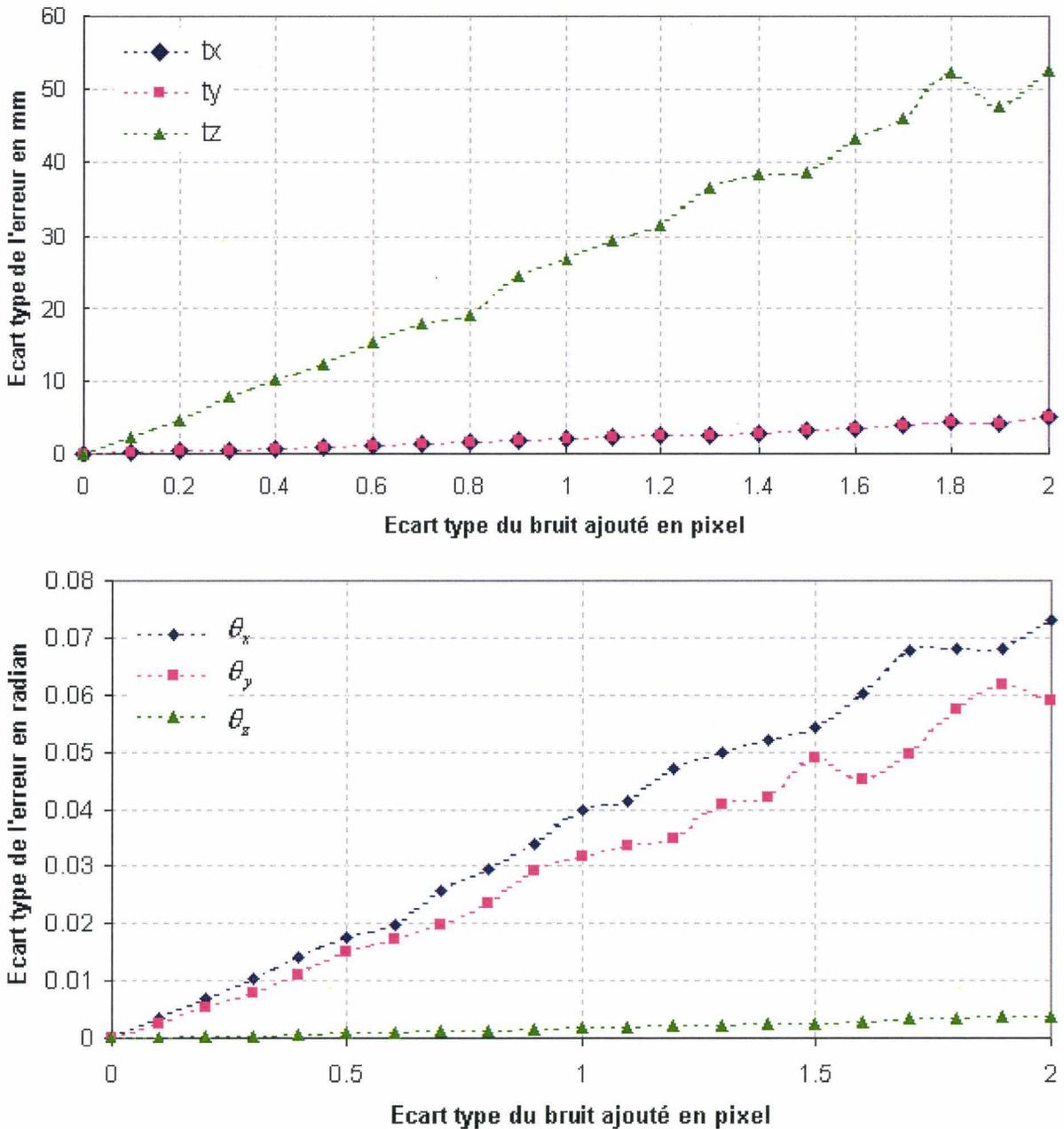


Figure III.23 : Ecarts-types des erreurs relatives à l'estimation des paramètres extrinsèques du capteur stéréoscopique (angles d'orientations et positions relatives) en fonction de l'écart type du bruit ajouté à la position de la projection des points de contrôle sur l'image.

La figure III.23 permet de constater que les écarts-types des erreurs d'estimation de l'orientation et de la position relatives sont linéairement dépendants (au moins quand l'écart type du bruit est entre 0 et 2 pixels) de l'écart type du bruit ajouté aux points utilisés pour la calibration.

D'une manière générale, l'évaluation du comportement de la calibration hybride vis à vis des différents paramètres influençant son estimation des paramètres intrinsèques et extrinsèques, nous a permis d'établir les conditions optimum d'utilisation de cette technique

de calibration. En effet, la mire de calibration doit contenir au moins 80 points de contrôle (pas beaucoup plus, car le faible apport de précision qui en découlerait serait coûteux en terme de temps de calcul) et la calibration doit s'effectuer avec treize images ou un peu plus (comme dans le cas des points de contrôle, prendre trop d'images serait cependant un luxe inutile dans notre cas). Par ailleurs, comme toute autre technique de calibration, le bruit sur l'extraction des points de contrôle utilisés diminue la précision de l'estimation. Pour cette raison, nous optons pour l'utilisation d'une mire dont les cases arborent une taille dans l'image suffisante, ce qui pour notre application conduit à utiliser des cases d'au moins 50 mm de coté.

III.3.5. DETECTION DE LA PLAQUE D'IMMATRICULATION

De façon traditionnelle, un système de reconnaissance de plaques d'immatriculations se compose des éléments suivants :

- Un module de détection de la plaque,
- Un module de segmentation des caractères,
- Un module de reconnaissance de caractères.

Parmi ces modules, le plus important est celui dédié à la détection de la plaque (la fiabilité du processus mis en œuvre par ce dernier affectera bien entendu celle du système tout entier). Pour que cette procédure d'extraction soit rapide, efficace et précise, nous devons faire face à plusieurs difficultés découlant de la nature dynamique des scènes observées et des variations des conditions de prise de vue (éclairage, transparence du milieu, etc.).

Plusieurs entités de recherche ont abordé la problématique de la détection des plaques d'immatriculation en utilisant des techniques diverses et variées [Abe.96, Kim.02a, Gra.99]. Parmi celles-ci, on citera les opérations morphologiques [Hsi.02], l'extraction des contours [Par.96a], la combinaison et l'analyse des gradients [Kim.02b], la classification par réseaux de neurones [Lee.94], la quantification vectorielle [Rov.99], etc. L'inconvénient principal de la majeure partie de ces techniques est le temps de calcul qu'elles requièrent. Qui plus est, si la problématique de la détection de la plaque minéralogique des véhicules à partir des caméras fixes a été abondamment traitée dans la littérature [Kwa.02, Hsi.02, Jun.01], celle-ci est plus difficile à aborder à partir des caméras embarquées.

Dans notre cas, le temps de réponse présente une contrainte critique, puisque la détection de la plaque doit être effectuée en « temps réel » sous peine de collision.

C'est l'une des raisons qui nous a fait choisir la plaque d'immatriculation comme « point d'attache visuel », puisque les calculs peuvent être accélérés grâce aux attributs particuliers cette dernière qui la distinguent des autres objets (hormis, abstraction faite de son contenu, des autres plaques ...) dans l'environnement du véhicule.

La plaque dans notre cas est de couleur « Jaune », de forme « rectangulaire » et son plan est plus au moins « perpendiculaire » à celui de la route. Ainsi, ces particularités simplifient énormément le processus de détection de la plaque, qui se « résumera » à la recherche d'un rectangle jaune de proportion connue dans une image. La Figure III.24 détaille les opérations que nous avons mises en œuvre aboutir à une telle détection :

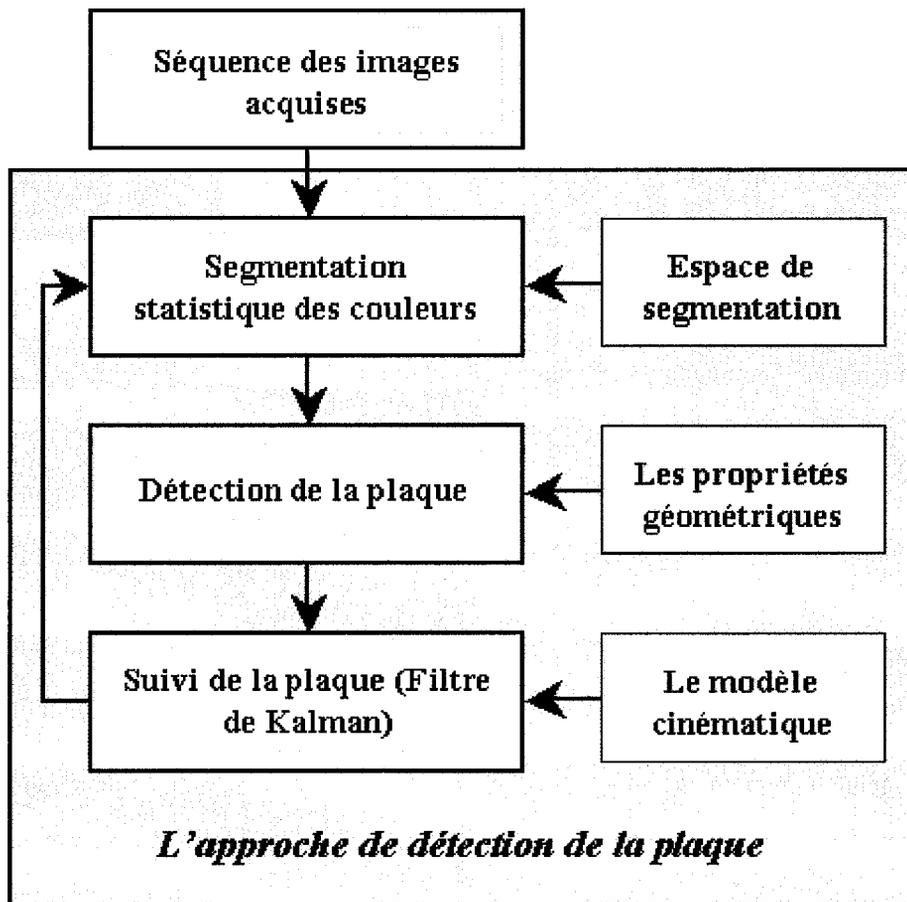


Figure III.24 : La procédure de la détection de la plaque d'immatriculation du véhicule suivi.

Le but des sections qui suivent est donc d'effectuer la description des différentes étapes nécessaires à la localisation de la plaque d'immatriculation dans l'image. En premier lieu, nous introduisons une technique de «segmentation statistique» renforcée par des heuristiques adaptées aux difficultés intrinsèques au domaine du traitement d'images. La problématique du suivi de la plaque d'immatriculation « *tracking* » fera l'objet d'une partie ultérieure.

III.3.5.1. LA SEGMENTATION STATISTIQUE

A) POSITION DU PROBLEME

Différentes techniques de segmentation d'images couleur sont présentées dans la littérature. Beaucoup d'entre elles s'appuient uniquement sur l'hypothèse de cohérence colorimétrique des régions constituant l'image. C'est le cas des méthodes de comparaison d'histogrammes couleurs, de clustering et de comparaison de moments statistiques (voir

[Gev.01, Fun.95, Kan.96, Str.95]). Certaines méthodes plus avancées prennent également en compte la relation spatiale entre les pixels, ce qui revient à exploiter la cohérence spatiale «forme de l'objet» en plus de l'attribut couleur afin de mener à bien la segmentation. Il faut ici souligner que le but principal de ces techniques est de séparer les différentes classes de pixels constituant une image sans connaissance *a priori* de ces classes (on cherche davantage à distinguer qu'à classifier).

Le problème considéré ici est légèrement différent, dans la mesure où on dispose d'une connaissance préalable de la couleur de la plaque minéralogique d'un véhicule, couleur qui constituera « la classe ciblée ». On se basera sur cette couleur pour trouver, un à un, les pixels de la plaque. Ces résultats seront renforcés par l'exploitation de contraintes géométriques associées à la forme supposée de la plaque d'immatriculation.

B) LA DECISION STATISTIQUE

1. ESPACE DE PRESENTATION

L'un des principaux problèmes qui se pose dans ce cas de figure consiste à séparer la «couleur plaque» des autres couleurs constituant l'image. Malheureusement, cette «couleur plaque» n'est pas bien définie, au sens où elle n'est pas unique (dispersion suivant les modèles) et que sa perception dépend bien évidemment des conditions de prise de vue. Autrement dit, elle correspond à un sous espace de l'espace tridimensionnel des couleurs RVB.

On se place dans un espace statistique correspondant aux paramètres sur lesquels s'effectuera la décision, à savoir, « l'espace RVB ». La problématique du choix de l'espace de représentation des couleurs a fait l'objet de nombreux travaux de recherche (voir [Van.00]).

Le choix de l'espace de représentation RVB dans notre cas se justifie par les exigences temporelles strictes de l'application. Le coût, en terme de temps de calcul, de passage à un autre espace hypothèquerait en effet le caractère «temps réel» de l'application.

Notre problème est d'associer un point de l'espace statistique à une classe. Dans notre cas l'espace des couleurs se compose de deux sous – ensembles : la « couleur plaque » et les autres couleurs. Nous faisons ici l'hypothèse que la répartition de la « couleur plaque » est gaussienne, ce qui procure des outils intéressants d'un point de vue pratique et qui met à disposition un certain nombre d'outils mathématiques éprouvés [Pos.87].

En premier lieu, nous commençons par représenter les individus dans l'espace RVB, puis, nous décomposons l'univers statistique en régions caractérisées chacune par un centre et une répartition. Ce partitionnement étant effectué, chaque individu appartient à un seul sous-ensemble, ce dernier définissant la classe à laquelle l'individu a été attribué.

2. LE PRINCIPE DE BASE

Soit un individu x , appartenant à l'espace statistique des couleurs. Nous posons : ω_i , la classe d'indice i ; $P'(\omega_i)$, la probabilité *a priori* de la classe ω_i ; $P'(\omega_i|x)$, la probabilité de x sachant ω_i ; $f(x)$ la densité de probabilité *a priori* de x ; $f'(x|\omega_i)$ la densité *a posteriori* de x .

La loi Bayessienne s'exprime de la façon suivante (voir [Pos.87]) :

$$f'(x) = \sum P'(\omega_i) f(x|\omega_i) \quad (3.58)$$

$$P'(\omega_i|x) = \frac{P'(\omega_i) f(x|\omega_i)}{f(x)} \quad (3.59)$$

La règle d'association d'un individu x à l'une des classes est alors donnée par :

$$\begin{cases} P'(\omega_i|x) = \max_j (P'(\omega_j|x)) \\ P'(\omega_i|x) \geq 1-a \end{cases}, \quad \text{alors } x \in \omega_i$$

$$\begin{cases} P'(\omega_i|x) = \max_j (P'(\omega_j|x)) \\ P'(\omega_i|x) < 1-a \end{cases}, \quad \text{alors rejet d'ambiguïté}$$

Le rejet d'ambiguïté correspond au cas où les classes seraient trop proches les unes des autres pour que la décision soit faite convenablement. La limite de proximité des M classes est ajustée par le coût de rejet en ambiguïté a , dont la valeur doit être estimée en prenant en compte les considérations suivantes :

$$\begin{cases} (M+1) \max_i (P'(\omega_i|x)) \geq 1 \\ \max_i (P'(\omega_i|x)) < 1-a \end{cases} \quad (3.60)$$

Ce qui se traduit par l'inégalité suivante :

$$a < \frac{M}{M+1} \quad (3.61)$$

Cette technique permet de décider entre un nombre connu de classes. Cependant, l'univers statistique peut décrire beaucoup plus de classes que nécessaire³³. De ce fait, il apparaît bien peu rentable de décrire exhaustivement toutes les classes dès lors que seules quelques-unes d'entre elles nous intéressent en pratique. Dans ce contexte, il est intéressant d'introduire une classe dite « de rejet de distance » ω_0 , qui nous permet de décider qu'une

³³ Celles qu'on veut identifier.

classe n'appartient à aucune des classes décrites précédemment, reflétant ainsi la position de l'individu sur lequel porte la décision dans l'univers statistique. Pour introduire la classe de rejet de distance, il va falloir fixer la probabilité *a priori* de la classe ω_0 . Nous pourrions alors écrire :

$$P(\omega_i) = (1 - P(\omega_0))P'(\omega_i) \quad (3.62)$$

Les calculs de la densité de probabilité *a priori* $f(x)$ et de la probabilité $P(\omega_i|x)$ doivent bien sûr se faire après l'introduction de la classe de rejet en distance ω_0 , puisqu'ils dépendent de $P(\omega_i)$.

3. DETERMINATION DES LOIS

Nous déterminons les probabilités *a priori* $P(\omega_i)$ et les densités de probabilité *a posteriori* $f(x|\omega_i)$ (y compris celle de rejet en distance) sous hypothèse gaussienne. Il nous faut donc estimer les paramètres des lois normales $N(\Theta, \Sigma)$, qui seront caractérisées par leurs deux premiers moments (la moyenne Θ et matrice de covariance Σ). Posons donc :

$$f(x|\omega_i) \rightarrow N(\Theta_i, \Sigma_i) \quad (3.63)$$

Où, Θ_i et Σ_i sont respectivement la moyenne et la matrice de covariance associées à la classe ω_i .

L'approche adoptée nécessite un ensemble d'apprentissage. Le nombre d'exemples par classe doit être représentatif de la probabilité *a priori* de la classe, puisque c'est à partir de son cardinal que nous allons calculer $\hat{P}(\omega_i)$, représentant l'estimation de $P(\omega_i)$. En effet, nous aurons :

$$\hat{P}(\omega_i) = \frac{\text{card}(\omega_i)}{\text{card}(\Omega)} \quad (3.64)$$

Ω : Représente l'univers probabilisé.

L'hypothèse gaussienne permet l'estimation directe par calcul de la moyenne et de la matrice de covariance. En l'absence d'information *a priori*, la classe de rejet de distance est supposée suivre une loi uniforme.

4. MISE EN ŒUVRE DE LA DECISION STATISTIQUE

Nous disposons de tous les éléments nécessaires pour mettre en oeuvre un test statistique. Rappelons que ce dernier est censé, au final, répondre à la question suivante : le pixel présenté (dans l'espace RVB) peut-il correspondre à la plaque d'immatriculation ?

L'aspect binaire de la réponse à cette question permet de mettre en œuvre un artifice destiné à pallier la lenteur de la décision statistique. En effet, nous procédons au calcul préalable de toutes les possibilités de décisions : étant donné un triplet (R, V, B) , on lui associe un booléen indiquant si oui ou non la combinaison présentée correspond au « jaune plaque » tel que l'ensemble d'apprentissage a permis de le définir. De cette façon, la prise de décision revient à adresser un élément d'un tableau tridimensionnel, ce qui s'avère extrêmement rapide.

Bien qu'elle nécessite *de facto* la séparation des phases d'apprentissage et d'exploitation, cette façon de procéder nous permet de satisfaire les contraintes « temps-réel » de l'application, c'est-à-dire 10 HZ sur des images de taille 640×480 avec un PIII cadencé à 1GHZ.

C) DETECTION DE LA « COULEUR JAUNE » PAR L'INTRODUCTION D'UNE NOUVELLE DIMENSION

Le problème que nous traitons ici consiste à partitionner l'espace RVB de sorte à caractériser la couleur « jaune plaque ». La méthode décrite précédemment opère cette partition à partir des informations issues d'un procédé assimilable à une phase d'apprentissage. Les résultats obtenus sont par conséquent fortement tributaires de celui-ci, même si l'hypothèse gaussienne offre une certaine « robustesse ». Certes, l'utilisation d'un ensemble d'apprentissage plus conséquent et plus représentatif améliore les résultats de détection, mais la collecte et la mise en forme de celui-ci se révèle être une tâche de longue haleine !

Pour cette raison, nous avons été amené à tenter d'effectuer cette partition de l'espace RGB non-plus directement à partir d'un ensemble d'apprentissage, mais au travers d'une approche analytique permettant de définir la « couleur jaune plaque ». L'idée principale dans ce cas de figure était de définir une « dimension », qu'on appellera J (pour la couleur jaune), de telle manière que cette nouvelle dimension soit une fonction permettant de quantifier directement le caractère « Jaune Plaque » d'un pixel pris dans l'espace RGB.

Contraintes imposées à la fonction J :

- ✓ La fonction de la couleur Jaune J doit être une fonction de l'espace de base RGB , autrement dit, $J = f(R, G, B)$.
- ✓ La fonction de la couleur Jaune J doit implicitement avoir un maximum global pour la couleur jaune pure, autrement dit, J présente un maximum lorsque $R = 255$, $G = 255$ et $B = 0$.

Développement :

Les deux contraintes exprimées ci-dessus peuvent être vérifiées par un nombre illimité de fonctions, parmi lesquelles se trouve:

$$J(R, G, B) = \log \left[\frac{(R+1) \cdot (G+1)}{(B^2+1)} \right] \quad (3.65)$$

En plus de vérifier les contraintes qui ont été imposées, la fonction (3.65) présente une propriété très importante à notre sens qui est de s'annuler pour le Blanc (à très peu de chose près) et le Noir (exactement). La figure suivante représente les graphiques de l'évolution de $J(R, G, B)$ en fonction de (R, G) pour des valeurs de B constantes (Cette évolution est similaire pour toutes les autres valeurs de B).

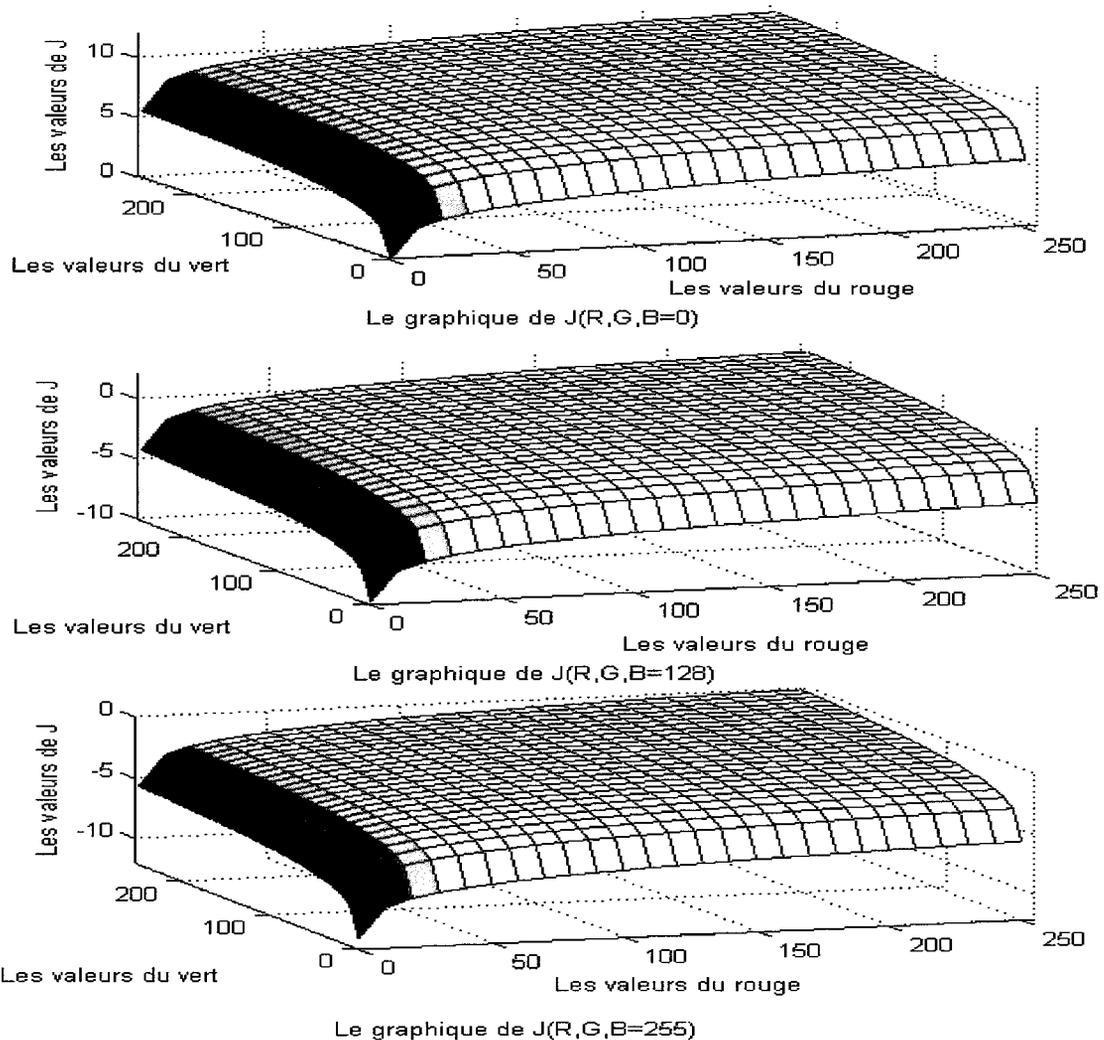


Figure III.25 : Graphique de la fonction J en fonction du vert et du rouge pour une valeur constante du bleu (coupe dans la plan $B = cste$).

De la Figure III.25, on s'aperçoit que la fonction $J(R, G, B)$ est monotone et croissante sur toute coupe dans les plans R et dans les plans G ; ses valeurs sont plus importantes pour les petites valeurs du bleu et des valeurs plus importantes du vert et du rouge. Nous constatons également que le minimum de cette fonction coïncide avec un bleu pur (c'est à dire : $R = 0$, $G = 0$ et $B = 255$). Par ailleurs, le noir et le blanc sont tous les deux égaux à 0 (« presque » en ce qui concerne le blanc mais avec une très bonne approximation). Les propriétés de la « nouvelle dimension » définie par la fonction $J(R, G, B)$ nous permettront de différencier facilement entre le jaune et les autres couleurs en binarisant l'image par rapport à un seuil calculé à partir de la projection de l'image dans le nouvel espace colorimétrique. Ainsi, nous allons pouvoir bénéficier d'une grande flexibilité de détection du jaune quelles que soient les conditions d'éclairage (pas besoin d'un apprentissage *a priori*), puisque cela revient à un seuillage de l'image de J (le jaune correspond aux valeurs élevées de J). Il est vrai qu'en utilisant cette technique de détection nous perdons en précision mais par contre nous gagnons en terme d'adaptabilité dynamique aux conditions d'acquisition, ce qui est appréciable dans une application telle que la notre. Néanmoins, le calcul d'un tel changement d'espace augmente considérablement les temps d'exécution de notre système de vision. Dans ce qui suit nous présentons l'artifice qui nous a permis de faire face à ce désagrément.

1. MISE EN ŒUVRE DU CHANGEMENT D'ESPACE

L'association d'un point (une couleur) de l'espace RGB dans le nouvel espace représentatif de la couleur « jaune » nécessite normalement le calcul de la fonction J . Ce calcul, répété pour chaque triplet associé à tout pixel de l'image, exigerait un temps de calcul considérable, d'autant plus que la fonction en question est une fonction logarithmique. La solution que nous avons retenue pour pallier ce problème consiste à calculer *a priori* les valeurs de J pour chaque triplet (R, G, B) et à les sauvegarder dans un fichier binaire comme nous l'avons fait pour la décision statistique. En effectuant cette opération, nous avons pu diviser le temps nécessaire au changement d'espace par un facteur cent avec un pentium III cadencé à 1 GHz, puisque l'accès aux éléments d'un tableau est de loin plus rapide (de l'ordre de 10^{-9} s) par rapport au calcul d'une fonction logarithmique.

D) OPERATION MORPHOLOGIQUES COMPLEMENTAIRES

Les deux techniques de segmentation de l'image présentées ci-dessus n'étant pas parfaites, l'image issue de cette étape est bruitée (en particulier, les zones détectées ne sont pas nécessairement homogènes) ; il nous a donc semblé intéressant de faire appel à une opération de morphologie mathématique classique, à savoir un « close » (voir [Par. 96b]), afin d'augmenter la cohérence de la détection.

Ainsi, après avoir détecté les pixels qui appartiennent au sous-ensemble de l'espace RGB représentant la couleur distinctive de la plaque d'immatriculation, nous procédons à une extraction des composantes connexes. Cette étape nous permet d'isoler les différentes régions constituant l'image issue du test statistique. Par conséquent, nous passons du niveau le plus

bas de l'information contenue dans une image « pixel », à un niveau « région » plus haut auquel nous pouvons associer des attributs. La prochaine étape consiste à trouver la plaque parmi les primitives détectées.

E) VERIFICATIONS GEOMETRIQUES

Une fois que nous obtenons les différentes régions candidates, nous sélectionnons celles qui ont une taille, une position et une orientation acceptable par rapport à la géométrie « standard » d'une plaque d'immatriculation. Pour exprimer ces différentes valeurs, nous avons établi une analogie avec une distribution Gaussienne 2D. La plaque est alors associée à une ellipse, à laquelle nous faisons correspondre la distribution Gaussienne 2D. En conséquence, les attributs associées à la région détectée, sont :

- *La position* (le centre de gravité de la région).
- *La taille* (calculée à partir de la matrice de covariance).
- *L'inclinaison* (obtenue grâce aux vecteurs propres de la matrice de covariance).

Ces attributs nous permettront de construire les rectangles, qui entourent au mieux ces régions et qui seront les plus représentatifs de la géométrie d'une plaque minéralogique.

La distorsion affectant l'image de la plaque est due essentiellement à la projection projective de la scène sur le plan de la caméra, qui à son tour dépend de l'orientation relative du plan de la plaque par rapport à celui de la caméra. Par conséquent, le rapport de la largeur L de l'image de la plaque sur sa hauteur H reflète sa disposition dans la scène observée. Les conditions d'utilisation de notre algorithme (circulation routière normale), autorisent des hypothèses réalistes quant à l'aspect de la plaque dans l'image. Ainsi, le rectangle contenant celle-ci ne doit pas présenter une forte inclinaison, ni un trop fort « écrasement » dans les deux directions (horizontale et verticale).

Cette relative conservation des proportions de la plaque est exploitée pour distinguer celle-ci des autres objets qui peuvent exhiber les mêmes propriétés colorimétriques, mais des attributs morphologiques assez différents. Les contraintes géométriques s'expriment comme suit :

$$\begin{cases} S_{R_{\min}} < R = \frac{H}{L} < S_{R_{\max}} \\ I_{\text{plaque}} < S_{I_{\max}} \end{cases} \quad (3.66)$$

Avec, R : Le rapport hauteur sur largeur de la plaque.

I_{plaque} : L'inclinaison du rectangle limitant la plaque.

$[S_{R_{\min}}, S_{R_{\max}}]$: L'intervalle tolérable du rapport R identifiant une plaque.

S_{imax} : Le seuil maximal de l'inclinaison permise.

Nous exploitons ici la technique des composantes connexes [Par.96b] pour distinguer les différentes régions détectées par la segmentation couleur (qu'elle soit statistique ou issue de l'application de la fonction J). Nous éliminons ensuite celles qui occupent des surfaces jugées trop petites ou très grandes. La détection des contours des différentes régions et l'approximation polygonale de ceux-ci [Dou.73] nous permettront d'éliminer les régions dont la forme n'est pas rectangulaire.

Il est cependant clair que la projection géométrique ne conserve pas les propriétés angulaires des objets projetés (l'image d'un rectangle telle qu'une plaque d'immatriculation n'aura pas forcément des angles droits), et ceci même dans le cas d'une simple translation. A plus forte raison, le changement d'orientation du véhicule tracteur provoquera quant à lui une déformation angulaire assez nette (voir, Figure III.26).

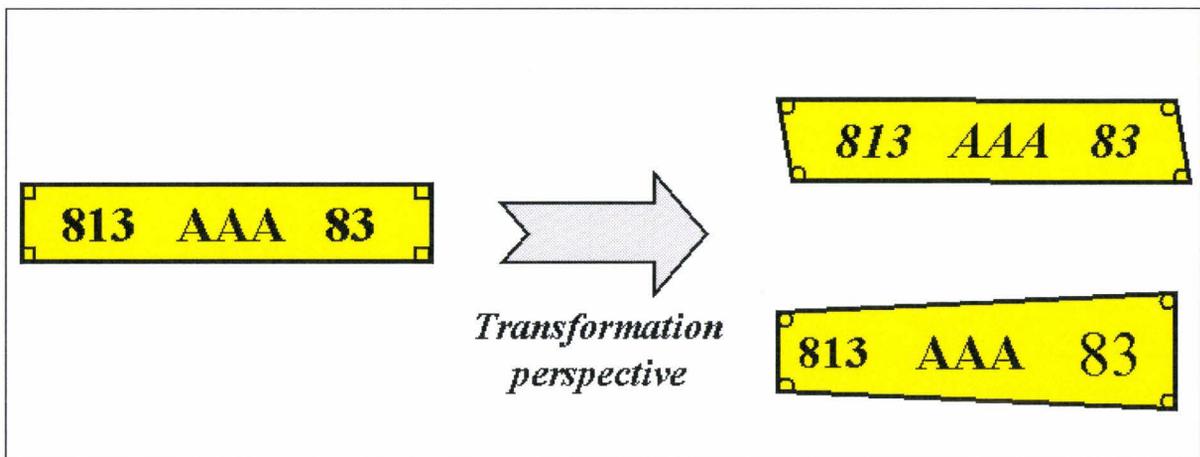


Figure III.26 : Exemples de déformations angulaires de la plaque d'immatriculation dues au processus de formation de l'image

Pour remédier à ce problème, il nous faut faire preuve d'une certaine tolérance vis-à-vis des déformations appliquées aux objets détectés, nous acceptons ainsi la détection de formes géométriques proches des rectangles, dans la mesure où l'angle entre deux lignes adjacentes est compris dans l'intervalle $]\pi/3 \quad 2\pi/3[$.

Il faut noter qu'on peut trouver plusieurs plaques sur une image. L'heuristique permettant de résoudre ce problème consiste à prendre en compte la plaque occupant la surface la plus importante (un périmètre plus important), du fait que celle-ci sera en principe la plus proche du véhicule suiveur. La solution à ce cas de figure sera « affinée » plus loin en prenant en compte « l'immatriculation » du véhicule. Autrement dit, c'est *la lecture de la plaque d'immatriculation* qui nous permettra de distinguer de façon univoque le véhicule suivi parmi les autres.

III.3.4.2. RESULTATS EXPERIMENTAUX

L'efficacité et la robustesse de la phase de la détection de la plaque d'immatriculation ont été évaluées sur des images réelles (voir Figure III.27).

Le taux de bonnes détections est de 80%, pour la segmentation statistique et de 95% pour la technique de seuillage sur la nouvelle dimension J . L'ensemble de test utilisé pour l'évaluation des performances est composé d'une centaine d'images et n'a, comme il se doit, pas été utilisé pour l'apprentissage. Ces expériences montrent les points forts (rapidité) ainsi que les limites (sensibilité à l'ensemble d'apprentissage et au choix des seuils de détection). En ce qui concerne la décision statistique, un ensemble d'apprentissage plus étendu permettra d'améliorer considérablement les résultats obtenus.

	Images originales	Binarisation et détection (décision statistique)		Binarisation et détection (seuillage dans le nouveau espace)	
1					
2					
3					
4					
5					
6					
7					
8					

Figure III.27 : Echantillon de l'ensemble des images de test utilisées pour évaluer les performances de chacune des deux techniques de détection de la plaque d'immatriculation. Les pixels de couleur blanche des images binaires représentent les pixels appartenant à la plaque, celle-ci est encadrée en vert sur l'image en couleurs, les images barrées représentent celles où la plaque n'a pas été détectée.

La Figure III.27 donne un aperçu des résultats obtenus lors de l'application de chacune des deux techniques décrites ci-dessus. Nous constatons que la plaque est détectée convenablement dans la plupart des cas par les deux approches. Néanmoins, la détection avec la deuxième technique apparaît plus adaptée dans certains cas (en présence d'ombres portées notamment, comme sur la figure III.27 (6)). Les plaques dans la septième et la huitième images n'étaient pas détectées et ceci pour deux raisons différentes. En effet, la plaque de l'image 7 présente des caractéristiques géométriques différentes que celles prise en charge par notre algorithme, tandis que la plaque de l'image 8 est partiellement cachée, ce qui entrave l'opération de détection du rectangle contenant la plaque. Dans ces deux derniers cas la défaillance de la détection est due à des causes géométriques et non pas à une erreur de classification des pixels.

En effet, la binarisation de l'image avec la technique de la segmentation statistique donne de bons résultats sur des images dont la couleur faisait effectivement partie de l'ensemble d'apprentissage. Les résultats se dégradent par contre assez facilement dès lors qu'on éloigne de celui-ci. La réponse binaire issue du classificateur représente ici un inconvénient. Néanmoins, cette technique a le mérite d'être très *précise* dans la détection du jaune déjà appris et ne présente que de faibles fausses détections. Un autre point fort de la décision statistique est l'*évolutivité* qui permet d'augmenter ses performances en utilisant de larges ensembles d'apprentissage au cours du cycle de vie du système de détection.

A contrario, la binarisation de l'image obtenue au travers de la fonction J permet d'avoir d'assez bons résultats quelles que soient les conditions de prises de vue. Autrement dit, cette technique présente un caractère de robustesse assez important par rapport aux changements brusques d'éclairage puisque le calcul dynamique d'un seuil de binarisation offre la possibilité de prendre en charge des cas de figures plus variables. Néanmoins, nous avons constaté expérimentalement que les seuils de détection de la plaque ne changent pas ou peu durant une séquence d'images réelles prise en continu. Généralement la valeur du seuil de J est supérieure à 0, sauf dans des cas où l'éclairage change fortement la couleur de la plaque. Dans ce cas, le seuil est négatif mais reste toujours proche de 0. Comme d'habitude avec ce type de techniques, la seule réelle difficulté de mise en oeuvre réside dans *le choix des seuils de détection* qui restent difficiles à définir dans certains cas de figure.

III.3.6. LECTURE DE PLAQUE D'IMMATRICULATION

Les opérations qui vont suivre correspondent à ce qui se passe à l'issue d'une phase de détection réussie de la plaque d'immatriculation. Nous disposons ainsi d'une image de cette dernière et nous allons entreprendre l'extraction de l'information « sémantique » de celle-ci, qui correspond au code d'immatriculation du véhicule suivi. Cette étape renforce les étapes de détection et du suivi de la plaque dans la mesure où elle permet de répondre à la question, somme toute primordiale, suivante :

« *Est-ce que nous suivons toujours le même véhicule ?* »

La réponse à cette question est fondamentale, puisqu'elle permet d'éviter des situations de « faux tracking » qui s'avèrent catastrophique vu la criticité de l'application et les effets irréversibles d'une telle défaillance...

Pour parvenir à « lire le code d'immatriculation », plusieurs techniques de reconnaissance de caractères sont envisageables (par exemple : *reconnaissance statistique*, *template matching*, *réseaux de neurones* tel qu'expliqué dans [Par.96b]). En ce qui nous concerne, nous avons optés pour l'utilisation d'un réseau de neurone basé sur l'algorithme de rétropropagation [Jod.94]. La Figure III.28, situe les étapes correspondantes au sein du processus de reconnaissance de la plaque :

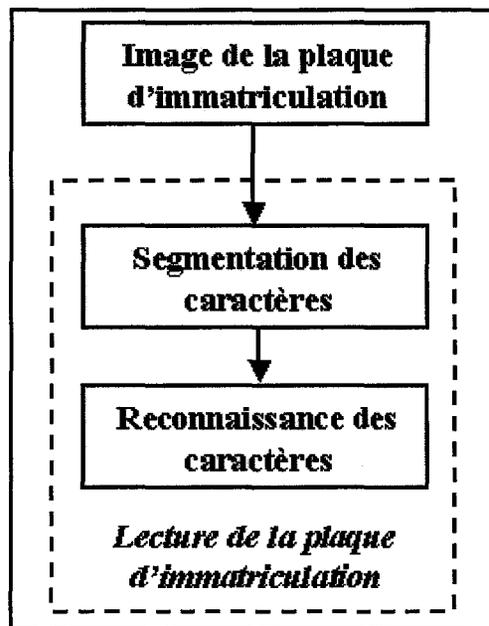


Figure III.28 : Synoptique du système de reconnaissance du code d'immatriculation du véhicule suivi.

La phase de la lecture sera donc décomposée en deux étapes distinctes qui sont :

- la segmentation des caractères.
- la reconnaissance des caractères proprement dite.

Dans ce qui suit, nous allons traiter en détails chacune de ces deux étapes.

III.3.6.1. SEGMENTATION DES CARACTERES

L'opération décrite ici consiste à extraire les caractères contenus dans la forme polygonale (plus au moins rectangulaire) délimitant la plaque d'immatriculation. On notera que la reconnaissance des caractères dans le cadre d'applications liées à la bureautique est une tâche aisée à réaliser, en particulier grâce à l'aisance avec laquelle les formes peuvent être extraites grâce à des caractères faciles à distinguer du fond du document.

A l'inverse, la nature de l'application considérée ici pose de nombreuses difficultés lors de cette phase d'extraction de caractères, en particulier :

- Eclairage non contrôlé (variations lumineuses intempestives et difficilement prévisibles).
- Position de la caméra par rapport à la plaque (distance, angle de vue).
- Variation de la taille des caractères.
- Présence d'informations parasites, du bruit essentiellement (des taches, des rivets, etc.).

La Figure III.29, montre quelques exemples des plaques d'immatriculation sur lesquelles l'extraction des caractères doit s'effectuer. On peut constater une grande variation d'éclairage, de taille et de couleurs entre ces dernières.



Figure III.29: Exemple des plaques d'immatriculation sur lesquelles l'extraction des caractères doit s'effectuer.

La figure ci-dessous donne une idée plus précise des étapes nécessaires à l'extraction des caractères de la plaque d'immatriculation.

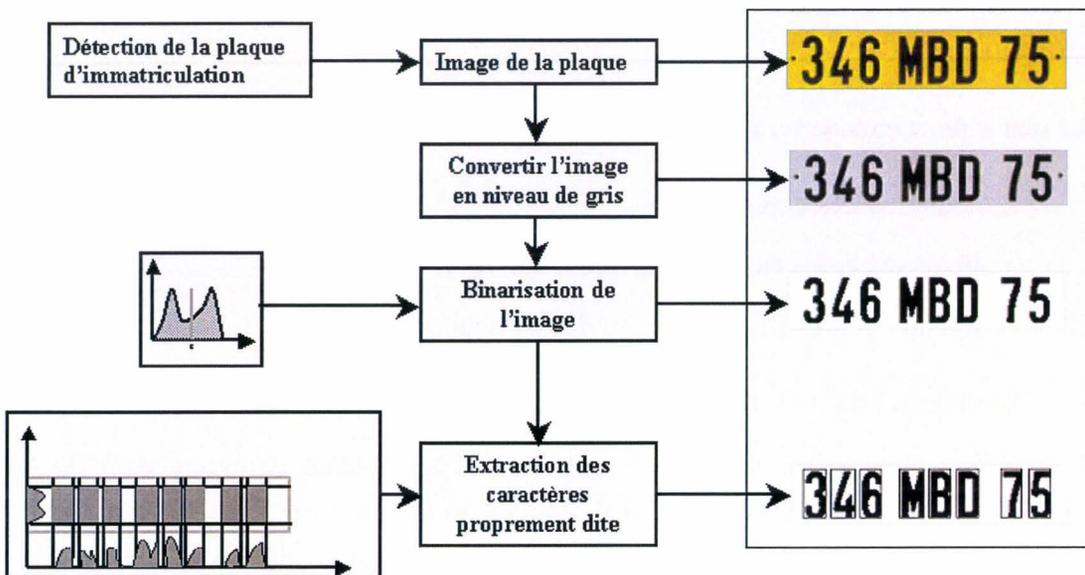


Figure III.30 : synoptique du processus d'extraction des caractères de la plaque d'immatriculation.

La première étape consiste à convertir l'image issue de la phase de détection (image de la plaque en couleur) en une image en niveau de gris en appliquant la formule de changement d'espace de représentation suivante :

$$G(i, j) = 0.212671 \cdot R(i, j) + 0.715160 \cdot V(i, j) + 0.072169 \cdot B(i, j) \quad (3.67)$$

Avec,

- $G(i, j)$: Le niveau de gris du pixel d'indices (i, j) qui représente en réalité la luminance.
- $R(i, j), V(i, j), B(i, j)$: Représentent le triplet rouge, vert et bleu qui définissent la couleur du pixel d'indices (i, j) dans l'image en couleur.

Il est à noter que le temps nécessaire au changement d'espace de présentation (RVB en niveaux de gris) sera évidemment d'autant plus petit que la taille de l'image considérée dans ce cas sera réduite. Comme nous utilisons des rectangles délimitant de façon assez « stricte » les plaques d'immatriculation, dont la surface est relativement faible par rapport à celle de l'image complète, ce temps de conversion est en pratique négligeable.

A) BINARISATION DE L'IMAGE

La binarisation consiste à transformer l'image en niveaux de gris en une image binaire, dans le but de séparer les caractères inscrits sur la plaque du fond de celle-ci. Bien que la plaque minéralogique ait été conçue dans le but d'identifier les véhicules, les caractères ne sont pas nécessairement aussi visibles qu'ils devraient l'être ! Pour cette raison des difficultés d'identification doivent être surmontées.

Pour parvenir à binariser l'image, nous devons procéder à un seuillage. La première difficulté consiste ainsi à trouver un seuil qui nous procure un maximum de séparabilité tout en préservant l'information utile, à savoir le code d'identification du véhicule. La première idée qui vient à l'esprit est de fixer unique un seuil pour toutes les images (par exemple : 128), puisque la plaque est censée être suffisamment contrastée par construction. Cependant, les expériences nous ont montrés que les variations lumineuses rendent inefficaces des approches aussi simplistes.

Pour remédier à cette première difficulté, nous exploitons un algorithme de binarisation à seuillage dynamique. La méthode consiste en premier lieu à construire un histogramme à partir de l'image en niveau de gris de la plaque d'immatriculation(voir, Figure III.31).

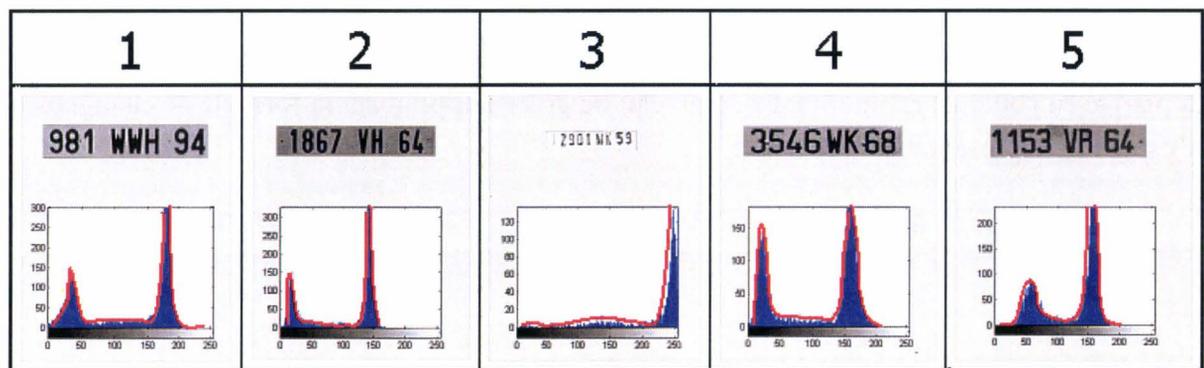


Figure III.31 : Exemples des plaques d'immatriculation en niveaux de gris à binariser, avec leurs histogrammes respectifs.

L'analyse des histogrammes de ces dernières nous dévoile l'existence de deux pics très visibles (sauf pour la Figure III.31- (3)). Cela est du à la structure de la plaque qui présente des caractères sombres sur fond clair. Partant de ces constatations, le seuil optimum de binarisation dans le cas qui nous préoccupe sera le niveau de gris correspondant au minimum compris entre les deux pics globaux de l'histogramme de l'image de la plaque. Par ailleurs, pour pallier les problèmes d'éclairage nous utilisons une propriété intrinsèque de la plaque d'immatriculation, liée au pourcentage de la surface qu'occupe le code d'identification par rapport à la surface globale. D'après nos expérimentations, les pixels appartenant aux caractères (pixels en noir) après binarisation doivent occuper entre 20 et 30 % de la surface totale de la plaque d'immatriculation.

B) ELIMINATION DU BRUIT

La mise en œuvre de la technique décrite ci-dessus nous a permis de séparer avec une grande précision les pixels des caractères du fond de la plaque. Malgré tout, quelques pixels de la plaque ne seront pas affectés à l'entité correcte (caractère ou fond), l'apparition de ces *pixels indésirables* est due essentiellement à :

- L'état de propreté de la plaque (une plaque sale ou maculée risque de générer un bruit important).
- La présence d'un bruit important sur le capteur lui-même (caméra de mauvaise qualité).
- La présence d'une éventuelle bordure autour de la plaque.

La liste ci-dessus n'est pas exhaustive. Elle permet cependant d'illustrer les types de bruit auxquels il est possible d'être confronté. Ainsi, les deux premiers présentent un caractère aléatoire et assez localisé, au sens où un faible pourcentage de pixels est affecté. La technique mise en œuvre pour l'éliminer repose principalement sur l'exploitation d'opérations morphologiques, comme une dilatation. Le dernier cas, bien qu'assez spécifique, est assez fréquent en pratique. Dans la mesure où il est plus au moins structuré et spatialement discernable, nous y remédions par l'intermédiaire d'opérations logiques avec un masque (en

l'occurrence une image) spécialement conçu à cet effet. L'opération logique effectuée peut s'exprimer suivant:

$$I_r = I_p \text{ or } \overline{M} \quad (3.68)$$

Avec,

- I_r : l'image résultat.
- I_p : l'image de la plaque après binarisation.
- M : le masque d'élimination des bordures (voir figure III.32).

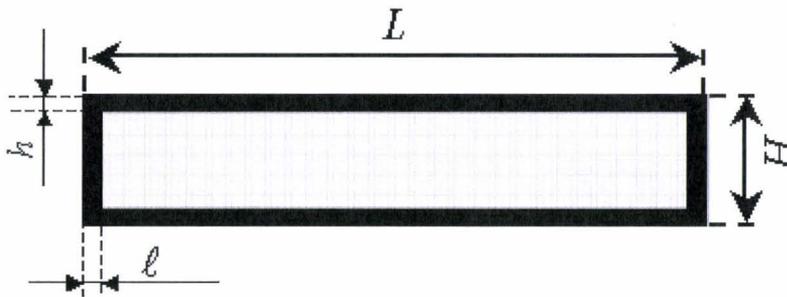


Figure III.32 : Masque d'élimination des pixels de la bordure de la plaque. Les expérimentations ont donné $l = 0.05 \cdot L$ et $h = 0.05 \cdot H$.

La Figure III.33 donne quelques exemples de résultats de binarisation avant et après l'application des techniques dédiées à l'élimination des deux types de bruit évoqués précédemment :

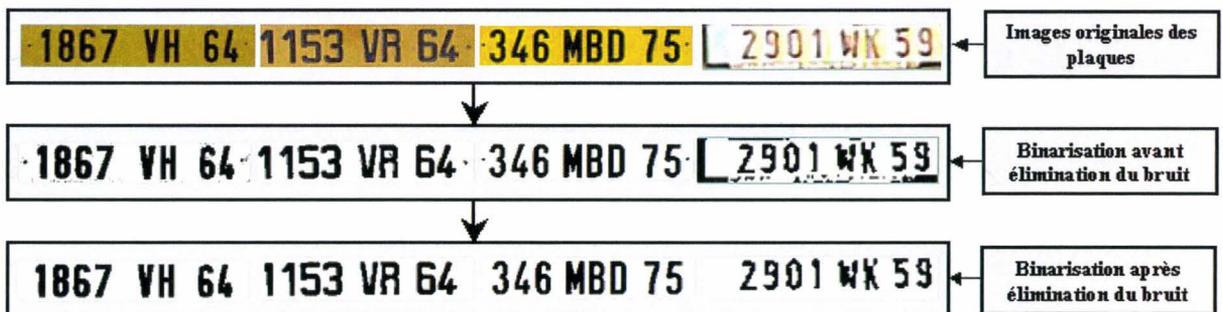


Figure III.33 : Images présentant les plaques d'immatriculation binarisées avant et après l'élimination du bruit.

La dernière étape consiste à « projeter » les pixels qui sont censés représenter les caractères du code d'identification du véhicule à la fois verticalement et horizontalement. En fait de projection, il s'agit plutôt d'effectuer, par ligne et par colonne, le cumul des pixels supposés correspondre aux caractères. Cette « projection » nous permet de trouver les lignes verticales et horizontales qui délimitent au mieux les caractères en détectant les passages entre les zones denses en pixels. L'intersection entre les lignes horizontales d'une part et les lignes

verticales d'autres part, délimitera exactement les régions d'intérêt, à savoir les zones contenant les caractères. Chaque région contenant potentiellement un caractère est ainsi extraite sous forme « d'imagettes » en sélectionnant les régions d'intérêt une à une. Au départ, les imagettes sont de tailles différentes. Nous procédons alors à une normalisation de leurs tailles respectives afin de pouvoir les soumettre à un traitement « générique ». La Figure ci-dessous illustre la procédure mise en œuvre et permet de se faire une idée des résultats issus de cette étape clef.

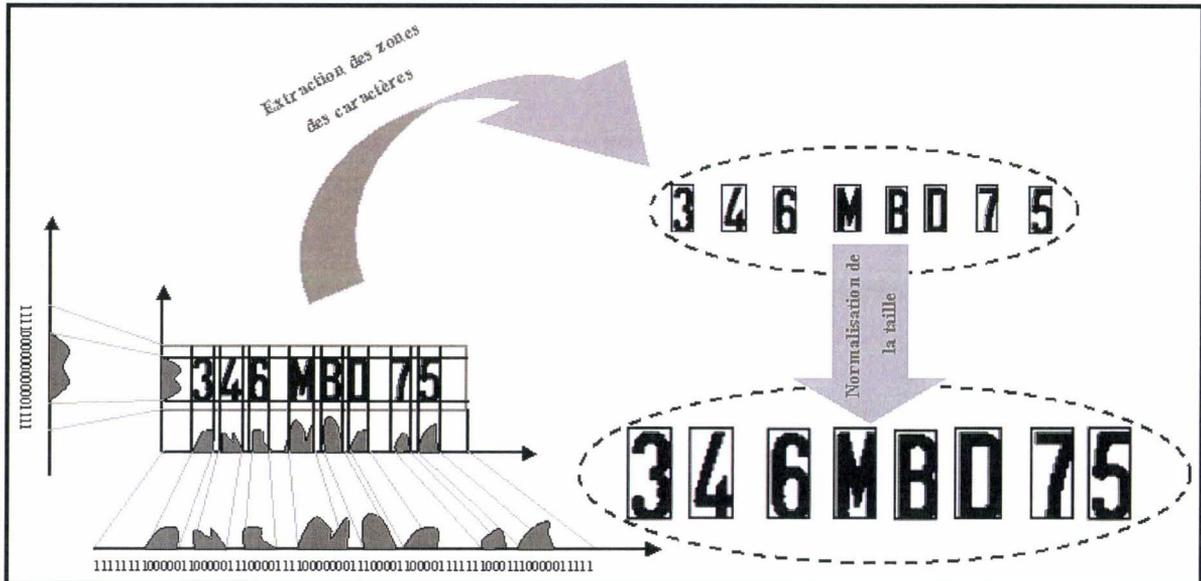


Figure III.34 : Extraction des caractères de l'image binarisée et normalisation de leur taille.

Après avoir effectué l'extraction des caractères constituant le code d'identification du véhicule suivi, nous décrivons dans la section suivante la technique neuronale utilisée dans le cadre de ces travaux pour la lecture des caractères.

III.3.6.2. RECONNAISSANCE DES CARACTERES

Comme nous l'avons précisé lors de l'introduction de la problématique de la lecture de la plaque d'immatriculation, la technique que nous avons utilisée pour la reconnaissance des caractères est bâtie autour d'un réseau de neurones mis en place et « entraîné » pour cette reconnaissance de forme spécifique.

Un « neurone artificiel » est un algorithme ou dispositif physique qui incorpore un modèle mathématique inspiré du comportement des neurones biologiques. Dans le cadre de la description simplifiée du neurone biologique, celui-ci accumule les signaux reçus des autres neurones. Si la somme de l'accumulation de ces signaux est supérieure à certain seuil, le neurone concerné transmet à son tour son propre signal aux autres neurones.

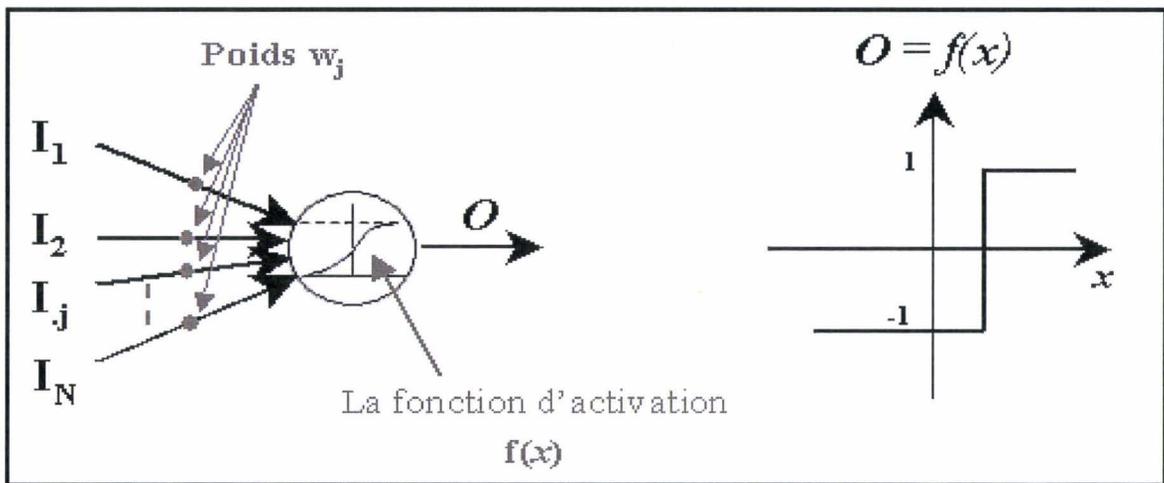


Figure III.35 : Modèle du neurone artificiel (le perceptron). Les connections des entrées au neurone sont pondérées par les poids w_j . La valeur de sortie du neurone est notée O .

De façon générale, n'importe quel modèle mathématique se basant sur l'utilisation d'entrées multiples afin de produire une seule sortie peut être exploitée pour la reconnaissance des formes. Ces modèles sont à la base du neurone artificiel. Qui plus est, si l'influence de chaque entrée peut être modifiée, le modèle correspondant supportera l'apprentissage.

L'utilisation des réseaux de neurones artificiels pour la reconnaissance date du début de leur développement. Le perceptron de Rosenblatt, est le premier réseau de neurones qui représentait une modélisation simple du système visuel [Ros.58]. Cependant, les limitations de ce premier système neuronal furent vite atteintes. Minsky et Papert [Min.69] publièrent des travaux mettant en exergue quelques limitations théoriques du Perceptron, notamment l'impossibilité de traiter des problèmes des séparations non linéaires ou de connexité. L'arrivée de l'algorithme de rétropropagation dans les années 80, a levé ce type de limitations et permis l'explosion des réseaux de type perceptrons multicouches (PMC). Ces derniers, basés sur l'algorithme de rétropropagation, sont actuellement les plus utilisés et représentent des extensions des modèles historiques que sont le *Perceptron* et *Madaline* [Ros.56].

C'est justement un PMC que nous avons utilisé afin de mener à bien la phase d'identification de la plaque d'immatriculation. Ce choix a été principalement basé sur le fait que ces réseaux, une fois leur apprentissage effectué, ont un temps de traitement bien plus rapide que les autres architectures envisageables (par exemple RBF) pour traiter ce même problème [Lec.98]. De ce point de vue, un PMC répond donc mieux que d'autre aux contraintes temporelles qui pèsent sur notre application.

Dans le détail, notre réseau s'appuie sur une structure de type « feed-forward » (voir figure III.37). Le réseau dispose d'une couche d'entrée, d'une couche cachée et d'une couche de sortie. Les neurones de la couche d'entrée reçoivent les composantes du vecteur descriptif de la forme à reconnaître (en l'occurrence ici : les caractères). La couche cachée permet la réalisation de l'opération de classification. En pratique, nous nous sommes contentés de onze

neurones sur la couche de sortie dans la mesure où nous avons jugé que la reconnaissance des chiffres inscrits sur une plaque, combinée à leurs emplacements respectifs dans la chaîne de caractères du code d'identification, est suffisante pour enlever l'ambiguïté sur l'identité du véhicule suivi. La méthode d'apprentissage utilisée est une méthode dite *supervisée*. En effet, pendant la phase d'apprentissage exploitant l'algorithme de rétropropagation du gradient [Rum86], l'adaptation des différents poids synaptiques du réseau s'effectue grâce à la connaissance *a priori* de la réponse que doit fournir le réseau pour chacune des entrées présentées.

La *rétropropagation* est une procédure mathématique qui commence par évaluer l'erreur à la sortie du réseau neuronal et la «propage» ensuite dans le sens inverse du flux informationnel normal (qui va habituellement de la couche d'entrée vers la couche de sortie). La contribution de chaque neurone à l'erreur commise est évaluée et les poids synaptiques du réseau sont ajustés de sorte à minimiser cette dernière.

L'apprentissage supervisé permet la détermination de l'erreur produite sur la couche de sortie en comparant la réponse du réseau avec la réponse (vecteur de sortie) désirée pour un élément donné de l'ensemble d'apprentissage. L'erreur est donnée par :

$$E = \frac{1}{2} \sum_{j=1}^N (D_j - O_j)^2 \quad (3.69)$$

Avec,

- D_j : la composante d'indice j du vecteur de sortie désiré D .
- O_j : la composante d'indice j vecteur de sortie du réseau O .

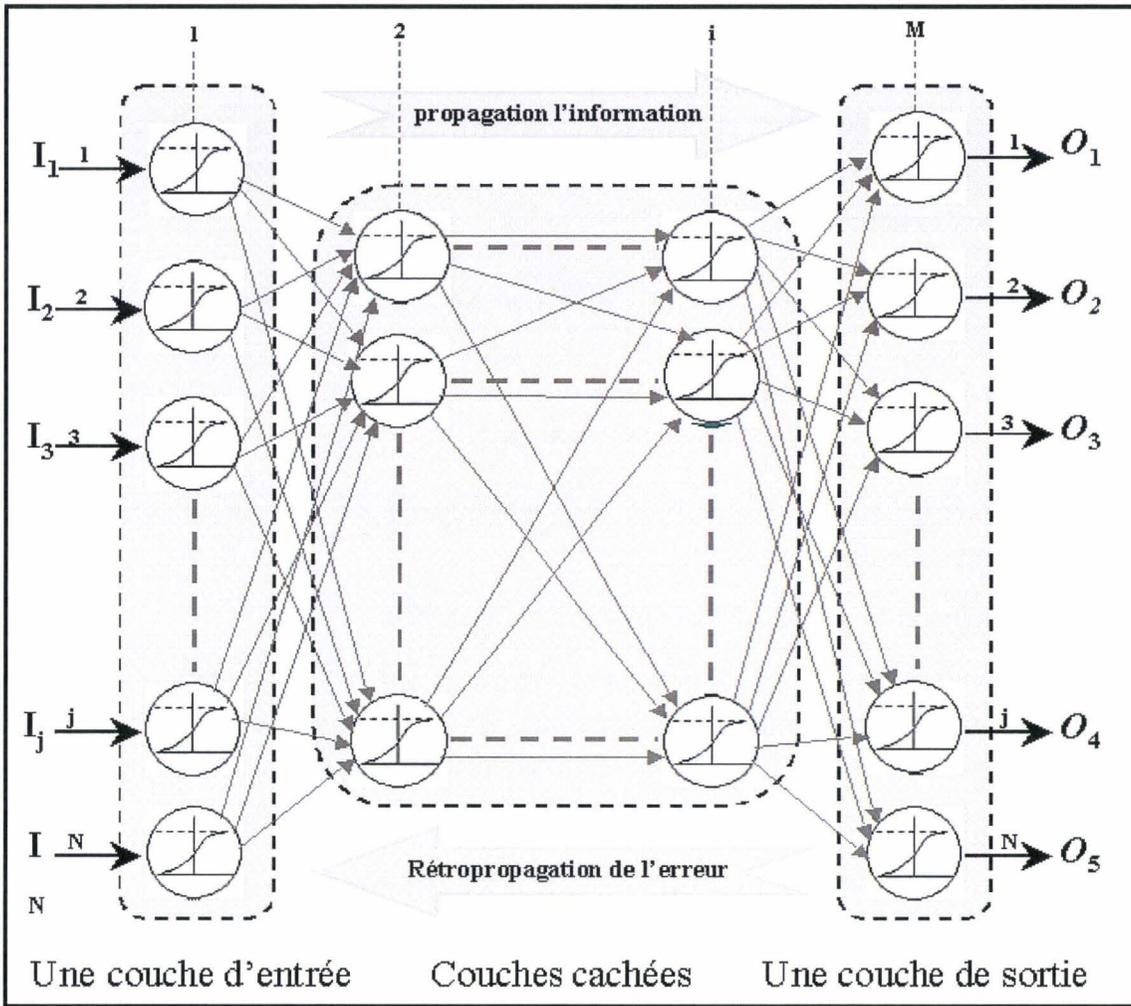


Figure III.36 : Architecture du perceptron multicouches.

L'équation (3.70) exprime l'erreur de la couche de sortie. Grâce à un calcul différentiel (voir, [Lec.87]), on peut exprimer l'erreur à la sortie de chaque neurone, à partir des erreurs de la couche précédente. Supposons un neurone d'indice j d'une couche d'indice i , l'erreur au niveau de ce neurone est exprimée ainsi :

$$E_j^i = O_j^i \cdot \sum_k w_{kj}^{i+1} \cdot E_k^{i+1} \quad (3.70)$$

Avec,

- w_{kj} : poids de la connexion entre le neurone j de la couche i et le neurone k de la couche $i+1$.
- O_j^i : sortie du neurone j de la couche i .

L'erreur ainsi exprimée, les poids peuvent être modifiés selon la règle suivante :

$$\Delta W_{jl}^i = -\lambda \cdot E_j^i \cdot O_l^{i-1} = -\alpha \cdot O_j^i \cdot \sum_k w_{kj}^{i+1} E_k^{i+1} O_l^{i-1} \quad (3.71)$$

Avec, λ : Le coefficient d'apprentissage.

L'ajout d'un terme de lissage permet d'améliorer l'algorithme précédent. En effet, les variations instantanées des poids provoquent l'instabilité du réseau, l'ajout d'un terme appelé le *momentum* μ permet de conserver une certaine dynamique de la modification précédente et ainsi préserver la stabilité du réseau.

$$\Delta W_{ji}^i = -\lambda \cdot E_j^i \cdot O_i^{i-1} + \mu \cdot \Delta W_{ji}^{i-1} \quad (3.72)$$

Plusieurs règles d'apprentissage ont été proposées. L'amélioration de la convergence en constitue le but principal, il s'agit principalement de :

- S'assurer que le réseau converge effectivement vers un minimum lors de l'apprentissage [Bab.89].
- Améliorer la rapidité de la convergence [Mas.92].

L'utilisation du réseau en classification est basée sur la construction d'une fonction discriminante grâce à une phase d'apprentissage. La classification ainsi effectuée est une classification par séparation ; le réseau générant les frontières séparant les différentes classes (dans notre cas les chiffres).

A) CONSTRUCTION DU CLASSIFIEUR DES CHIFFRES

Notre PMC comporte $L \times C$ neurones sur sa couche d'entrée, avec L et C les dimensions en pixels (nombre de lignes et de colonnes respectivement) des imagerie binaires des caractères normalisés à reconnaître. L'activité d'un neurone de coordonnée q (tel que $q = i + (j \cdot L)$) de la couche d'entrée correspond alors au niveau binaire du pixel de coordonnées (i, j) de l'imagerie en cours d'analyse. La couche de sortie contient autant de neurones qu'il y a de caractères à classifier, à savoir dix chiffres et un neurone de plus pour les imagerie qui ne représentent pas des chiffres. Ainsi, le neurone de la couche de sortie produisant la valeur la plus grande et suffisamment proche de l'unité détermine la classe à laquelle sera associée l'imagerie présentée à l'entrée du réseau. Une seule couche cachée est exploitée. Celle-ci utilise la classique fonction sigmoïde $f(x) = \frac{1}{1 + e^{-x}}$ pour fonction d'activation. Le nombre de neurones constituant cette couche cachée est la racine carrée du produit du nombre des neurones de la couche d'entrée par le nombre des neurones de la couche de sortie [She.90], cette organisation permettant une convergence plus rapide du réseau.

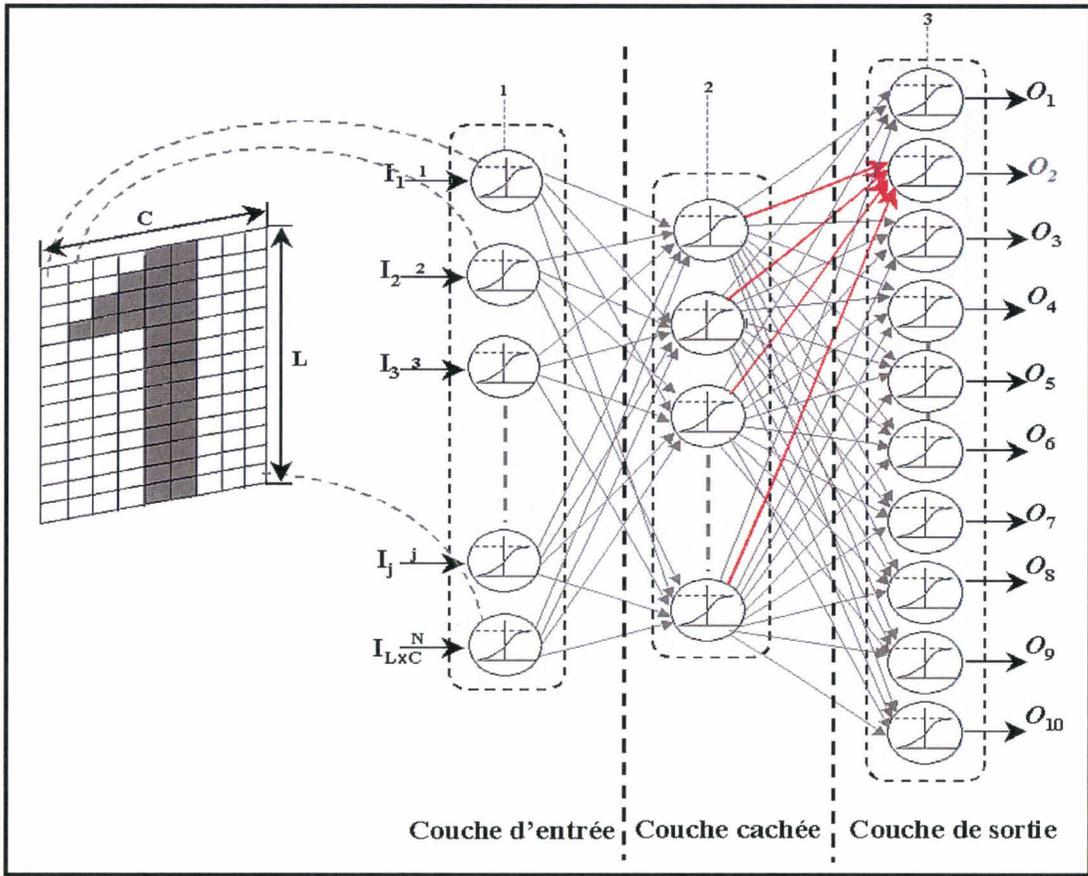


Figure III.37 : Architecture du réseau de neurones utilisé pour la reconnaissance des chiffres et exemple de fonctionnement : L'image du chiffre 1 active la sortie deux (O_2).

Comme nous l'avons déjà précisé, l'opération d'apprentissage du réseau est effectuée de manière supervisée. Dans le détail, pour chaque vecteur modèle $M = (m_0 \ m_1 \ \dots \ m_{L \times C})$ de la base d'apprentissage, nous associons un neurone k de la couche de sortie, de telle manière que nous construisons le vecteur de sortie désiré. Schématiquement, l'apprentissage modifie les poids de connexion du neurone k avec la couche cachée par les composantes du vecteur d'apprentissage.

En phase de reconnaissance, si un vecteur X est identique au vecteur M appris et mémorisé dans les coefficients synaptiques (les poids W), la sortie du neurone k sera maximale. Nous pouvons ainsi estimer le degré de similitude entre le modèle présenté et le modèle appris.

Etant donné le cadre de notre application, nous avons volontairement limité le nombre de classes à 11, puisque d'une part, la majorité des caractères composant le code d'immatriculation est constitué (en général) par des chiffres et, comme nous l'avons précisé, la prise en compte de la position de ces chiffres dans le code d'immatriculation rajoute suffisamment de contraintes pour ôter toute ambiguïté en pratique. Ce choix impose un nombre de 10 neurones dans la couche de sortie, tel que l'activation des neurones de $j = 1$ à

10 représente respectivement les chiffres de 0 à 9 et l'état où tous les neurones de sortie sont désactivés indique qu'il s'agit d'un autre caractère qu'un chiffre.

En ce qui concerne la taille de la couche d'entrée, celle-ci doit respecter un compromis entre la qualité de la reconnaissance et la taille du réseau. En effet, chaque neurone de cette couche est associé à une composante du vecteur à classifier. Il s'agit donc de choisir le vecteur le plus petit possible afin d'aboutir à l'architecture de réseau la plus réduite possible et ainsi diminuer considérablement le temps d'apprentissage et le temps de traitement en exploitation. Cependant, il s'agit de choisir une taille minimale permettant de conserver suffisamment d'informations pour distinguer correctement toutes les classes.

A.1. Détermination de la taille de la couche d'entrée

Pour déterminer la taille optimale de la couche d'entrée du réseau, nous avons procédé à une étude des performances et du comportement de ce dernier en fonction de la taille des imageries. Le problème de classification des caractères considéré ici consiste à distinguer les différentes classes pour un vecteur d'entrée de dimension bien connue. Cette dimension influence inévitablement la qualité de la reconnaissance. Ainsi, une taille trop réduite du modèle présenté à l'entrée du réseau ne prodigue pas assez d'informations pour séparer convenablement les différents chiffres. A l'inverse, si la résolution est trop importante, le classifieur peut déterminer précisément les différentes classes mais exigera un temps de traitement plus important. L'étude des performances de cinq différents réseaux dont la taille de la couche d'entrée varie de 6×10 à 14×18 nous a permis de dégager un bon compromis.

Pour cela, nous avons fait varier la hauteur et la largeur des imageries du modèle à identifier de 10 à 18 et de 6 à 14 avec un pas de 2. On a ainsi élaboré les cinq bases d'apprentissage qui ne diffèrent que par la résolution des modèles qui les composent. Afin de limiter les causes des mauvais fonctionnements et d'évaluer le plus correctement possible les performances des différents réseaux correspondants, nous avons gardé les mêmes paramètres pour tous (taux d'apprentissage, ...) et la même procédure d'extraction des caractères (décrite plus haut).

L'évaluation des performances des cinq réseaux consiste à étudier, lors de la phase d'apprentissage, l'évolution de l'erreur sur la base d'apprentissage, du taux de reconnaissance et du temps de traitement. Nous avons choisi de ne pas tenir compte du temps d'apprentissage puisque cette opération s'effectue hors ligne et n'influence par conséquent pas les performances temporelles du système en exploitation. Cela n'empêche pas de préciser que l'augmentation du nombre des neurones de la couche d'entrée du réseau augmente considérablement le temps d'apprentissage, comme on peut s'en douter.

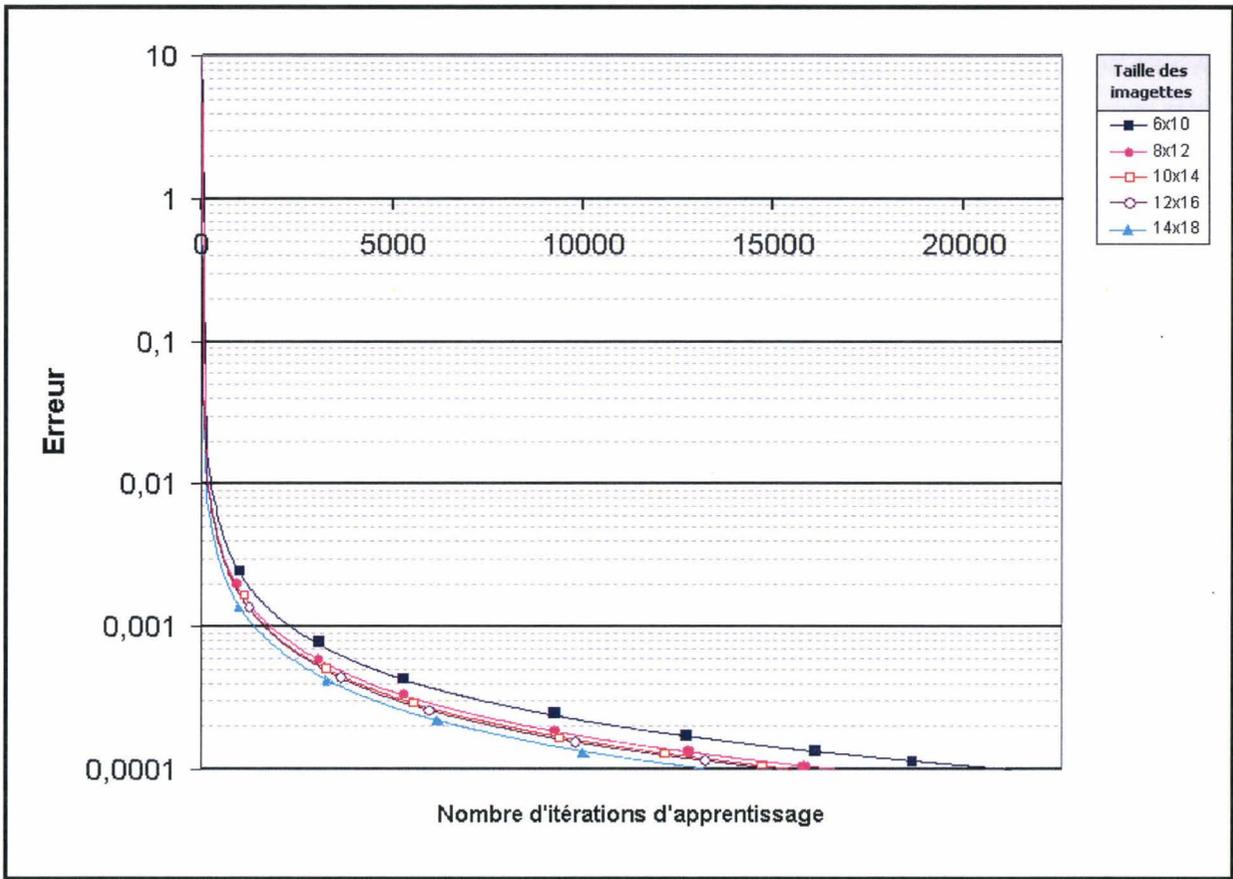


Figure III.38 : Evolution de l'erreur d'apprentissage en fonction du nombre d'itérations, pour différentes tailles de la couche d'entrée.

La Figure III.38 traduit, l'évolution de l'erreur sur la base d'apprentissage en fonction du nombre d'itérations lors de la phase d'apprentissage. Malgré que les erreurs continuent leur décroissance, nous avons limité l'apprentissage à une erreur égale à 10^{-4} . L'analyse du graphique de la Figure III.38 dévoile le lien existant entre la taille du réseau et le nombre d'itérations nécessaires pour converger vers la même erreur. En effet, l'augmentation de la taille de la couche d'entrée impliquera forcément une diminution du nombre d'itérations nécessaires pour l'apprentissage.

Afin, d'évaluer le taux de reconnaissance des cinq réseaux de neurones, nous avons procédé à l'identification d'un ensemble de test constitué de 184 imagettes pour chaque taille de couche d'entrée du réseau. Le tableau suivant montre les taux de reconnaissance pour chaque taille de réseau :

Taille de la couche d'entrée (<i>neurones</i>)	Nombre total de l'ensemble de test	Reconnaissance		Taux de reconnaissance
		Vrai	Faux	
60 (6×10)	184	147	37	79.86 %
96 (8×12)	184	152	32	82.6 %
140 (10×14)	184	152	32	82.6 %
192 (12×16)	184	153	31	83.15 %
252 (14×18)	184	158	26	85.86 %

Tableau III.7: Les taux de reconnaissance obtenus sur 184 imagettes pour les différentes tailles de la couche d'entrée du réseau.

L'examen du tableau précédent montre que les taux de reconnaissance sont proches les uns des autres. Ce critère ne constituant pas une caractéristique discriminante, le choix de la taille de couche d'entrée sera déterminé grâce aux performances temporelles des différents réseaux testés. Afin d'évaluer les temps de traitement des cinq réseaux de la manière la plus proche possible des conditions d'exploitation, nous avons mesuré le temps nécessaire à la reconnaissance d'un caractère, 10 fois consécutives pour chacune des 2660 imagettes constituant l'ensemble de test. Le temps de traitement retenu est la moyenne des temps de traitement mesurés. Le tableau III.8 présente les résultats obtenus ainsi que le temps nécessaire pour construire et initialiser le réseau pour une taille donnée :

Taille de la couche d'entrée (<i>neurones</i>)	Le temps de traitement par imagette (<i>ms</i>)	Le temps nécessaire pour le chargement et l'initialisation du réseau (<i>ms</i>)
60 (6×10)	0.86	46.49
96 (8×12)	1.88	90.93
140 (10×14)	3.36	175.44
192 (12×16)	5.28	294.96
252 (14×18)	8.31	465.42

Tableau III.8 : Temps de traitement, de construction et d'initialisation, nécessaires pour chacune des tailles de réseau testées (avec un pentium III cadencé à 1GHz).

Comme on pouvait s'y attendre, il apparaît clairement sur le tableau ci-dessus que la taille de la couche d'entrée fait augmenter considérablement le temps de traitement nécessaire pour la reconnaissance des formes. Les deux derniers réseaux nécessitent un temps de calcul très important. A des fins de comparaison, le deuxième réseau identifiera quatre caractères

contre un seul pour le cinquième réseau dans un même laps de temps. Compte-tenu des contraintes temporelles associées à l'attelage virtuel, l'utilisation des réseaux 4 et 5 poserait problème, en particulier si on se rappelle qu'une plaque d'immatriculation contient en pratique au moins six caractères à identifier.

Par conséquent, le choix se restreint entre les trois premières configurations. Vue la différence des taux de reconnaissance en faveur du deuxième et du troisième réseau, nous avons opté pour l'utilisation d'une couche d'entrée constituée de 96 neurones, correspondant à des caractères organisés sous la forme de matrices 8×12 .

III.3.6.3. EVALUATION DES PERFORMANCES DE LA PROCEDURE DE LECTURE DE LA PLAQUE

L'évaluation des performances globales de la procédure de lecture de la plaque, à savoir la chaîne de traitement que constitue les procédures d'extraction et de reconnaissance de caractères (voir Figure III.28), nécessite la réalisation d'expériences sur des images de plaques réelles. Nous avons ainsi effectué plusieurs tests sur des images de plaques d'immatriculation. La Figure III.39 montre quelques exemples de résultats obtenus.

Image originale	Image binarisée	Extraction des caractères	Résultat de la reconnaissance
981 WWH 94	981 WWH 94	981 WWH 94	981 XXX 94
1867 VH 64	1867 VH 64	1867 VH 64	1867 XX 64
3065 WWA 34	3065 WWA 34	3065 WWA 34	3065 XX 34
346 MBD 75	346 MBD 75	346 MBD 75	346 XX0 75
1153 VR 64	1153 VR 64	1153 VR 64	1153 XX 64
981 WWH 94	981 WWH 94	981 WWH 94	981 XXX 94
7768 WL 02	7768 WL 02	7768 WL 02	7768 XX 02

Figure III.39 : Quelques exemples issus de la procédure de lecture des plaques d'immatriculation (le symbole X signifie que le caractère identifié ne représente pas un chiffre).

Nous avons pu constater que cette procédure de lecture de plaque minéralogique est efficace dans la majorité des cas et présente une certaine robustesse en ce qui concerne les inclinaisons de la plaque (voir, Figure III.39 image 7 du haut vers le bas). Cependant, certains problèmes peuvent se poser lors de la différenciation de caractères morphologiquement proches tels que le *D* et le *0* ou le *B* et le *8*, problèmes que peuvent

justement accentuer certaines orientations. (voir la 4^{ème} image de la Figure III.39). Pour pallier ce défaut, certaines heuristiques peuvent être employées. Ainsi, pour une plaque française, le troisième caractère de la droite vers la gauche ne peut pas présenter un chiffre.

III.3.7. L'ALGORITHME DE TRACKING

III.3.7.1. INTERET DE LA DEMARCHE

L'approche développée dans la partie précédente nous permet de détecter la plaque minéralogique du véhicule tracteur en traitant l'ensemble de l'image, ce qui est coûteux en terme de temps de calcul. La mise en œuvre de notre système d'attelage virtuel repose sur la robustesse et la rapidité de l'algorithme de « tracking » destiné à suivre en temps-réel la plaque d'immatriculation.

Pour diminuer le temps consacré à la détection de la plaque, nous proposons une technique de suivi qui nous permet d'effectuer des prédictions sur la future position de la plaque au sein de la prochaine image acquise. Nous réduisons de cette façon considérablement la zone de recherche (appelée la « région d'intérêt ») dans laquelle la procédure de détection de la plaque s'effectuera. La technique consiste à estimer la position du centre de la plaque dans l'image suivante, en utilisant un filtre de Kalman [Wel.95], basé sur un modèle simple du déplacement du véhicule. Le tout prend en compte les bruits (supposés ici gaussiens) engendrés dans les phases précédentes.

III.3.7.2. LA DEMARCHE GLOBALE

Comme indiqué, la démarche retenue ici consiste à tenter de prévoir la position du centre de la plaque dans la prochaine image. Nous considérons ici le mouvement apparent du pixel correspondant dans le champ de la caméra. Compte-tenu de l'application envisagée (attelage virtuel entre deux véhicules), les principales grandeurs d'influence affectant ce mouvement apparent sont *a priori* les suivants :

- Trajectoire du véhicule « tracteur ».
- Trajectoire du véhicule « suiveur ».
- Mouvements verticaux des caisses de chacun de ces deux véhicules par rapport à leurs châssis respectifs.
- Variation des paramètres de la caméra (extrinsèques ou intrinsèques).

Cette liste (forcément incomplète) montre que les causes possibles de ce mouvement sont pour le moins très nombreuses et que leur prise en compte exhaustive nécessiterait des capteurs supplémentaires et des traitements éventuellement complexes, qui compliqueraient l'obtention d'un algorithme présentant des caractéristiques « temps-réel » tout en restant relativement « généraliste » (au sens où il serait applicable en l'état à tout type de véhicule).

On prend ici le parti de ne s'appuyer que sur les seules images consécutives de la scène observée pour effectuer le « tracking ». Compte-tenu de la modélisation effectuée, rien n'empêchera par la suite la prise en compte de capteurs supplémentaires pour affiner l'estimation. L'algorithme détaillé ci-après, exploitant le formalisme du filtrage de Kalman [Wel.95], repose sur les hypothèses suivantes :

- (A) le mouvement d'un pixel dans l'image, sur un intervalle de temps suffisamment petit, peut être considéré comme uniformément accéléré.
- (B) Pour un pixel donné, les déplacements en abscisse et en ordonnée sont indépendants.

Cette hypothèse découle des suivantes :

- (1) L'acquisition des images est effectuée avec une fréquence respectant le théorème de Shanon relativement à la dynamique des phénomènes observés (une fréquence d'échantillonnage de 10 Hz respecte cette contrainte dans le cadre d'une conduite « normale »).
- (2) La trajectoire apparente d'un pixel dans l'image est continue, dérivable et de dérivée continue.
- (3) L'erreur de linéarisation commise est faible.

Les équations données ci-après correspondent à la mise en œuvre du filtre dans une direction donnée (en abscisse, ou en ordonnée). L'hypothèse (B) impliquant le « découplage » des mouvements suivant les axes u et v , la trajectoire complète du pixel sera « prédite » grâce à la mise en œuvre de deux filtres (un pour chacune des deux directions).

A) LE MODELE CINEMATIQUE DU DEPLACEMENT DES PIXELS

Notre modèle s'appuie sur le fait que le mouvement apparent des pixels de la plaque le long d'une courte séquence est uniformément accéléré, et que les mouvements horizontaux et verticaux sont indépendants. Dans ce qui suit, nous décrivons les calculs appliqués pour l'estimation de l'évolution des abscisses seulement (l'évolution des ordonnées s'effectuant de la même manière).

Ainsi, l'évolution de l'abscisse u peut être modélisée par un « double intégrateur » prenant pour entrée l'accélération apparente $\Gamma(z)$. La fonction de transfert discrète correspondante peut être écrite comme suit :

$$H_u(z) = \frac{U(z)}{\Gamma(z)} = \frac{T_e^2}{(z-1)^2} = \frac{T_e^2}{z^2 - 2z + 1} \quad (3.73)$$

Avec T_e est la période d'échantillonnage.

Selon (3.73), les échantillons $\{u_k\}_{0 \leq k \leq N}$ de l'abscisse du pixel considéré peuvent être calculés depuis les valeurs de l'accélération $\{\gamma_l\}_{0 \leq l \leq N}$ grâce à la récurrence suivante :

$$u_{k+1} = 2u_k - u_{k-1} + T_e^2 \gamma_{k-1} \quad (3.74)$$

Bien que l'accélération soit censée être constante, au moins sur un certain horizon, il est commode par la suite de la considérer plutôt comme « lentement variable », ce qui peut se traduire par :

$$\gamma_{k+1} = \gamma_k + \varepsilon_k, \text{ Avec } \varepsilon_k \text{ très petit par rapport à } \gamma_k.$$

L'idée consiste alors à prendre en considération l'accélération courante en tant que composante du vecteur d'état à estimer, ce qui nous permettra d'avoir une idée de ses variations et de les intégrer à la prédiction ultérieure. Ainsi, le vecteur d'état considéré s'exprimera ainsi :

$$X_k = [u_k \quad u_{k-1} \quad \gamma_k]^T \quad (3.75)$$

L'évolution du vecteur d'état dans ce cas est assujettie à la relation suivante :

$$X_{k+1} = \begin{bmatrix} u_{k+1} \\ u_k \\ \gamma_{k+1} \end{bmatrix} = \underbrace{\begin{bmatrix} 2 & -1 & T_e^2 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_F \cdot X_k \quad (3.76)$$

L'observation correspond ici à l'abscisse du pixel et l'équation correspondante présente la forme suivante :

$$Y_k = u_k = \underbrace{[1 \quad 0 \quad 0]}_H \cdot X_k \quad (3.77)$$

L'équation (3.76) sert de base à la construction d'une prédiction, notée $X_{k+1/k}$:

$$X_{k+1/k} = F \cdot X_{k/k}^* \quad (3.78)$$

B) LES EQUATIONS DU FILTRE DE KALMAN

Les équations établies ci-dessus nous ont permis d'implémenter un filtre de Kalman. La figure III.40 représente l'algorithme prédicteur – correcteur de Kalman :

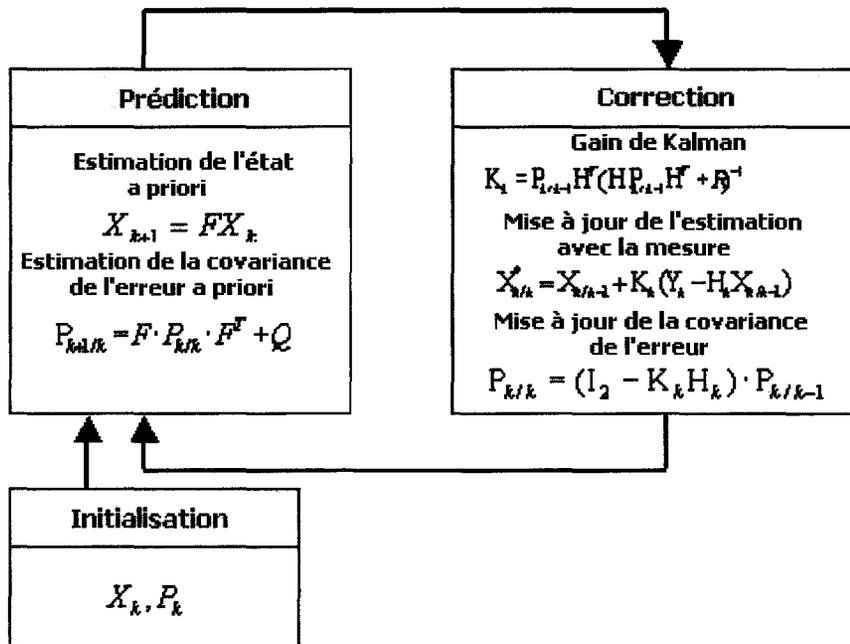


Figure III.40 : l’algorithme du filtre de Kalman

Nous disposons des éléments nécessaires à la mise en œuvre du filtre qui exploitera :

- La matrice de covariance de l’erreur de prédiction *a priori* : $P_{k+1/k} = F \cdot P_{k/k} \cdot F^T + Q$
- L’estimation « optimale » au sens de Kalman du vecteur d’état : $X_{k/k}^* = X_{k+1/k} + K_k (Y_k - H_k X_{k+1/k})$
- La matrice de covariance de l’erreur de prédiction *a posteriori* : $P_{k/k} = (I_2 - K_k H_k) \cdot P_{k+1/k}$

K_k le gain du filtre de Kalman : $K_k = P_{k+1/k} H^T (H P_{k+1/k} H^T + R)^{-1}$. Q et R sont respectivement les matrices de covariance du modèle de prédiction et celui de l’observation³⁴, I_2 est la matrice identité 2×2 .

III.3.7.3. MISE EN ŒUVRE DU FILTRE

Ce filtre de Kalman nous permet d’améliorer significativement les performances temporelles du système puisqu’il fournit dynamiquement une fenêtre de recherche restreinte pour l’entité poursuivie. Nous avons procédé à des tests sur des scènes « en situation » d’une centaine d’images consécutives dont les résultats sont présentés sur la Figure III.41. Ce graphique met en évidence le comportement satisfaisant de l’estimation par rapport à l’évolution réelle de la position, ce qui semble valider *a posteriori* l’hypothèse du mouvement apparent « localement » uniformément accéléré.

³⁴ L’observation étant un scalaire, est donc directement la variance de l’observation (scalaire).

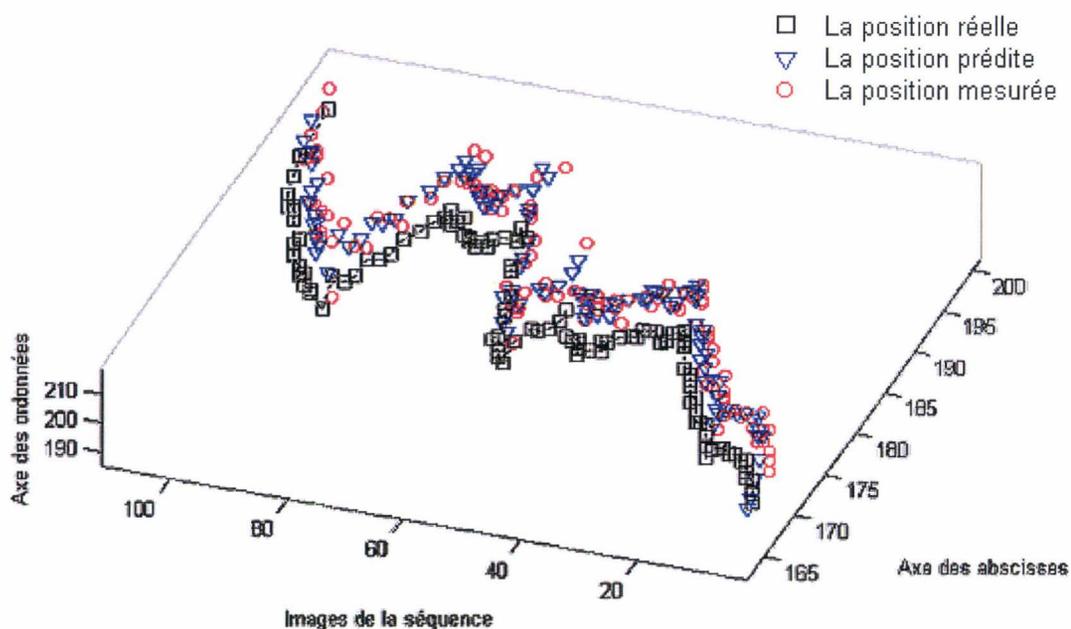


Figure III.41 : Evolution des positions réelles, prédites et estimées du centre de la plaque d'immatriculation (à travers une séquence de 100 images).

La procédure décrite précédemment nous procure une technique rapide et robuste pour la détection de la plaque d'immatriculation sur les images. Nous sommes donc capables d'extraire en temps réel dans une séquence d'image des éléments caractéristiques « fiables » associés au véhicule suivi, permettant d'effectuer sa localisation 3D. Les sections qui suivent ont pour objectif de décrire la façon dont cette dernière tâche est menée à bien.

III.3.8. EXTRACTION DES INFORMATIONS NECESSAIRES AU SUIVI DU VEHICULE

On identifie deux modes de fonctionnement du capteur stéréoscopique développé, à savoir le mode normal (« canonique ») et le mode dégradé. Le premier consiste à utiliser le capteur stéréoscopique dans le cas « habituel » où les deux caméras qui le composent sont parfaitement fonctionnelles et l'objet traqué (la plaque d'immatriculation) se trouve dans le champ commun des deux caméras, sous des conditions de prise de vue favorables (de jour, absence du brouillard, etc...). Dans ce cas, la reconstruction métrique (ou plus exactement l'extraction des paramètres nécessaires à l'attelage) s'accomplit en se basant sur l'information fournie par les deux caméras et les techniques de la reconstruction tridimensionnelle conventionnelles peuvent être appliquées (par exemple : la triangulation passive).

Dans un second temps, il ne faut bien sûr pas exclure que, pour une raison ou une autre, l'une des deux caméras tombe en panne (défaillance du matériel) ou que l'objet traqué sort du champ de vision commun des deux caméras (ce qui peut se produire en cas de virage serré effectué à basse vitesse, comme en ville par exemple). La question qui se pose dans ce cas est la suivante : peut-on assurer la fonction d'attelage virtuel avec une seule caméra ?

Dans un premier temps, nous exposons donc les détails de fonctionnement de notre système dans le cadre du mode normal. Dans un second temps, nous apportons une solution originale au problème de la perte d'une des deux images en nous basant sur des connaissances *a priori* de la géométrie de la scène d'une part et sur des informations métriques extraites lors de la phase du fonctionnement normal du capteur d'autre part. Nous avons ainsi mis en place une stratégie de fonctionnement en mode dégradé, qui s'avère plus que nécessaire vu la criticité du domaine des véhicules intelligents !

III.3.8.1. LE MODE NORMAL DU FONCTIONNEMENT DU CAPTEUR STEREOSCOPIQUE

Comme nous l'avons évoqué, le capteur stéréoscopique dans le mode normal permet de fournir les coordonnées des projections d'un point de l'espace tridimensionnel (de coordonnées *3D* inconnues) sur la rétine des deux caméras. Le modèle géométrique de ces dernières étant supposé connu, nous pouvons estimer les coordonnées du point ayant engendré les projections. En effet, l'estimation des coordonnées d'un point de la scène appartenant au champ commun des deux caméras du capteur stéréoscopique consiste à effectuer une triangulation passive impliquant les centres de projection des deux caméras et les projections du point de la scène par rapport à un repère donné (par exemple : repère de la caméra gauche). Toutefois, la reconstruction 3D (la triangulation passive dans notre cas) est fortement affectée par les « bruits d'extraction » d'une part et les « bruits d'estimation des paramètres » des modèles des caméras d'autre part (l'estimation des paramètres intrinsèques et extrinsèques). Par conséquent, les droites formées ne se croisent presque jamais, ce qui exclut l'emploi d'un simple calcul d'intersection de droites.

A) LA RECONSTRUCTION 3D

La « reconstruction 3D » d'un point de l'espace, c'est à dire la détermination de ses coordonnées dans l'espace euclidien relativement à un repère connu, nécessite l'observation de ce point de l'espace de deux points de vue ou plus. Ces techniques de reconstruction sont désignées dans la littérature par le nom de technique de « triangulation ». On distingue à ce titre deux types de méthodes:

- La triangulation active
- La triangulation passive

Le premier type de triangulation, à savoir la triangulation active, nécessite généralement l'utilisation d'une source lumineuse destinée à « illuminer » un point donné de la scène. De cette manière, il devient relativement facile d'extraire ce point particulier dans les deux (ou plus) images de la scène en utilisant ses propriétés d'apparence (telle que sa couleur). Cette technique facilite, pour ne pas dire qu'elle résout, le problème de la mise en correspondance. Cependant, l'utilisation d'une source lumineuse pilotée n'est pas sans posé de problèmes. En premier lieu, le nécessaire « balayage » de la scène que cela impose conduit à une relative lenteur du procédé. Par ailleurs, le faisceau lumineux peut subir diverses réflexions qui

empêchent l'obtention d'une mesure valide. Pour finir, il s'agit malgré tout d'un dispositif de mesure actif, et par conséquent sensible aux phénomènes d'interférence avec des instruments basés sur des techniques comparables.

A contrario, la triangulation passive ne nécessite pas l'utilisation d'éléments d'éclairage particuliers. Cependant, elle présente de réels défis techniques tels que ceux que nous avons évoqué (mise en correspondance temps réel, robustesse vis-à-vis des inévitables erreurs de mise en correspondance, sensibilité aux conditions d'observation, calibration, etc.). Elle demeure toutefois une technique extrêmement flexible, avec un pré-requis nul relativement aux infrastructures.

L'une et l'autre des techniques précédemment évoquées utilisent strictement les mêmes fondements simples de la géométrie pour effectuer la « reconstruction 3D » à partir d'une paire de points mis en correspondance. Ce calcul repose en effet sur la détermination de l'intersection de deux lignes dans l'espace euclidien.

Dans le cadre de cette thèse nous nous intéressons exclusivement à la reconstruction 3D en utilisant la triangulation passive compte-tenu des contraintes qui sont les nôtres.

Il est à noter qu'on distingue classiquement entre deux problématiques de la reconstruction, à savoir la reconstruction euclidienne d'une part et la reconstruction projective et affine d'autre part. Ces problématiques sont, bien entendu, fortement liées à l'espace de reconstruction choisi.

En ce qui nous concerne, nous allons passer en revue les méthodes classiques de reconstruction à partir de vues calibrées, qui constituent des méthodes de reconstruction euclidienne.

1) INVARIANCE DES TRANSFORMATIONS

Même si elles ne sont pas l'objet de notre discours, il faut signaler qu'un intérêt particulier a été porté au sujet des reconstructions projectives ou affines [Stu.97], dans le cadre desquelles la scène est reconstruite à une transformation inconnue du groupe donné près. Généralement, dans une telle situation, on ne dispose pas des vraies matrices de projections mais plutôt de l'image de ces dernières par certaines transformations.

Une caractéristique que doit présenter une méthode de reconstruction est son invariance aux transformations du groupe considéré (affine, projectif ou euclidien). Une méthode de triangulation Ω reconstruisant un point 3D P à partir de ses projections p_1 et p_2 et des matrices de projection correspondantes (notées respectivement P_1 et P_2) sera notée:

$$P = \Omega(p_1, p_2, P_1, P_2) \quad (3.79)$$

Ω sera dite invariante à une transformation T si et seulement si :

$$\Omega(p_1, p_2, P_1, P_2) = T^{-1}\Omega(p_1, p_2, P_1T^{-1}, P_2T^{-1}) \quad (3.80)$$

L'équation (3.80) implique que suite à un changement de repère (transformation des matrices de projection), la méthode doit donner comme résultat l'image de P par la transformation correspondante. Autrement dit, si nous nous trouvons dans le cas d'une reconstruction projective ou affine, les matrices de projection sont connues à une transformation projective ou affine près. Il est par conséquent recommandable d'utiliser une méthode de triangulation invariante aux transformations projectives ou affine, suivant le cas considéré.

2) LES METHODES DE TRIANGULATION

Jusqu'à un temps assez proche, les travaux relatifs à la reconstruction à partir de deux vues s'inscrivaient dans le contexte de la stéréovision, sous l'hypothèse de caméras calibrées. Le problème principal qui y était considéré était généralement la détermination du mouvement entre les deux caméras [Adi.89, Ais.90, Bro.86, Bru.88, Fan.84, Fau.87, Kar.94, Wen.87, Zhu.86]. La partie « reconstruction » n'était pas l'objet d'une attention excessivement soutenue.

La méthode du Point Milieu est la plus utilisée. En effet puisque les deux rayons de projections ne se croisent pas forcément, cette technique consiste à prendre le point milieu de la perpendiculaire commune aux deux rayons de projections [Bea.94, Che.94, Tos.87a, Wen.88, Wen.87, Yak.78] comme indiqué sur la Figure III.42.

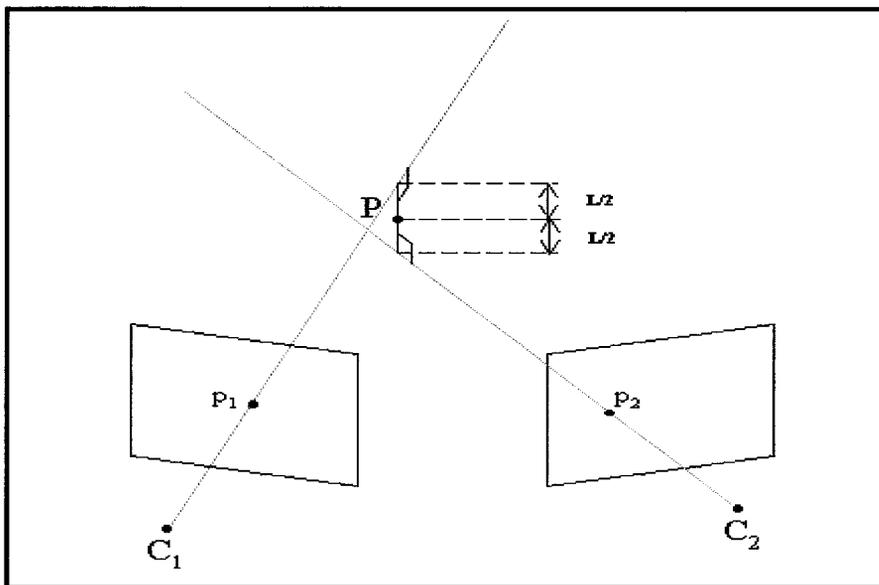


Figure III.42 : Le point milieu de la perpendiculaire aux rayons de projection

Toscani et Faugeras ont proposé deux solutions alternatives [Tos.87b]. La première consiste à trouver le point 3D qui minimise l'erreur de projection. Cette technique a été adoptée par nombre d'autres chercheurs, davantage dans le but d'ajuster les faisceaux que d'opérer une reconstruction 3D. Cette technique, connue sous le nom de « bundle adjustment », est plutôt usitée dans le but d'estimer le mouvement. Cette minimisation est généralement effectuée par des méthodes itératives, tandis que Sturm a donné une solution

directe au sous-problème de la triangulation (voir [Stu.97]). La deuxième solution proposée pour la triangulation [Tos.87b] semble moins « heuristique » que le choix du point milieu mais les résultats obtenus en pratique témoignent de la stabilité de cette technique comparativement à la précédente. En effet, cette technique postule (légitimement) que le point reconstruit appartient inévitablement au plan contenant un des deux rayons de projection. Par conséquent, le point reconstruit est l'intersection de l'un des rayons de projection avec le plan perpendiculaire au plan défini par l'autre rayon de projection et le centre optique contenu dans le premier rayon, et contenant le second rayon de projection (voir Figure III.43).

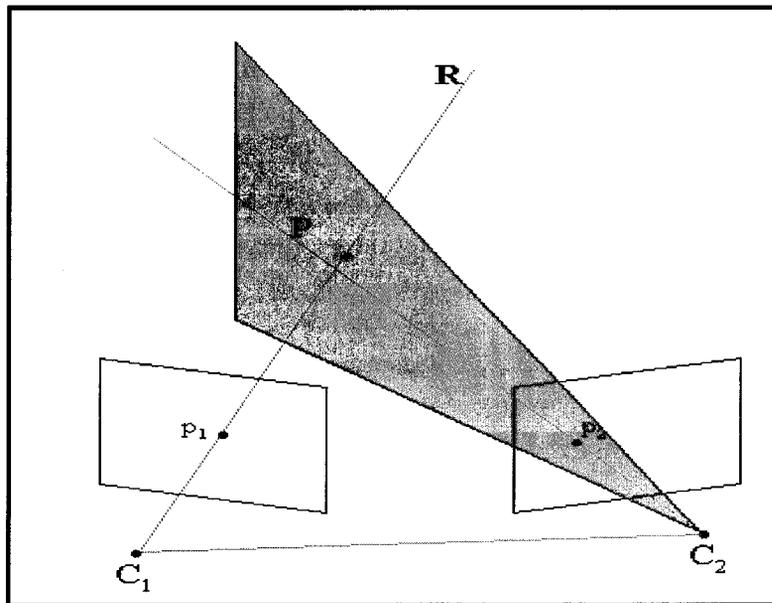


Figure III.43 : La reconstruction proposée par Toscani et Faugeras

Dans le cas d'une reconstruction projective (ou affine), les matrices de projection sont données par rapport à un repère projectif (ou affine). Ici, les concepts euclidiens tels que les angles et la perpendicularité perdent leur sens. En ce qui concerne spécifiquement la reconstruction projective, même les notions affines comme le rapport de distances entre points colinéaires (et donc le concept de milieu) dégénèrent. Comme les méthodes citées ci-dessus sont basées sur les notions de perpendicularité ou de milieu, elles ne sont pas applicables dans un contexte projectif ou affine.

B) LOCALISATION DES POINTS DANS L'ESPACE TRIDIMENSIONNEL

La Figure III.10 illustre le fait que l'emplacement d'un point P_i dans l'espace tridimensionnel, correspond à l'intersection des deux droites $(C_g P_{ig})$ et $(C_d P_{id})$;

Avec C_g, C_d les centres optiques des caméras et p_{ig}, p_{id} les projections de P_i sur les plans des caméras gauche et droite respectivement.

En raison du « bruit » affectant la détection des points p_{ig} et p_{id} , il n'y aura généralement pas d'intersection entre les deux droites et plusieurs techniques ont été développées pour faire face à ce problème. Une comparaison entre ces dernières est disponible dans [Har.97]. Dans le contexte particulier de notre application, à savoir la reconstruction euclidienne métrique, nous exploitons la technique du point milieu, qui est moins sensible aux bruits dans ce cas de figure. Elle consiste à calculer les coordonnées du point – milieu du segment de longueur la moins importante qui joint $(C_g p_{ig})$ et $(C_d p_{id})$, satisfaisant l'équation suivante :

$$ap_{ig} + cw = T + bR^T p_{id} \quad (3.81)$$

Avec : $a, b, c \in \mathcal{R}$,

ap_{ig} : est la droite qui passe à travers C_g et p_{ig} , exprimée dans la référence de la caméra gauche,

$T + bR^T p_{id}$: définit la droite qui passe par C_d et p_{id} , exprimée dans la référence de la caméra gauche

$w = [p_{ig}]_x R^T p_{id}$: est le vecteur orthogonal aux deux droites ($[p_{ig}]_x$ est la matrice antisymétrique de p_{ig}).

En résolvant le système linéaire (3.81) pour (a, b, c) , on peut calculer les coordonnées des deux points des extrémités du segment le plus court liant les deux droites. L'itération de ce processus pour des points caractéristiques d'un même objet nous permettra de déterminer les dimensions de l'objet en question ainsi que son orientation par rapport au système de stéréovision. La poursuite de cette démarche dans « l'espace temps » (une séquence d'images) autorise également l'estimation de sa vitesse relative.

C) LES RESULTATS EXPERIMENTAUX

Dans ce qui suit, nous allons présenter les résultats obtenus grâce à la mise en place d'un capteur stéréoscopique conforme au concept de « capteur intelligent » (intelligence à la fois structurelle et intégrée au sein des traitement appropriés à la tâche de l'attelage virtuel, comme expliqué en II.4).

Afin d'estimer et d'évaluer les performances du concept du capteur mis en place, nous avons opté pour l'utilisation de deux types différents de caméras. Le premier capteur a été réalisé à base de Webcams grand public dont les optiques d'origine ont été remplacées par du matériel de meilleure qualité (tout en restant dans une gamme de prix parfaitement abordable, de l'ordre de 45 €). La qualité des images obtenues a ainsi été grandement améliorée et les résultats fournis par le capteur furent tout à fait encourageants, comme permettent de le

constater les données fournies dans ce qui suit. Le deuxième capteur développé, plus spécifiquement conçu pour être embarqué sur le véhicule, est quant à lui constitué de deux caméscopes d'une qualité d'image nettement supérieure à celle des Webcams. Cette amélioration découle des caractéristiques du capteur CCD employé (meilleure résolution, meilleure sensibilité, dynamique plus rapide), de l'optique (plus lumineuse, affectée par moins d'aberrations et bénéficiant d'une focale variable sur une large gamme) et des traitements intégrés dans les caméras elles-mêmes (auto-focus, contrôle d'exposition performant et rapide offrant une certaine innocuité aux variations d'éclairage, etc.). C'est ce capteur qui nous a permis de mener à bien des expériences en conditions de fonctionnement réelles. Dans l'un et l'autre cas, les caméras utilisées sont des matériels grand public, nettement moins coûteuses que celles utilisées généralement dans l'industrie. Ce choix délibéré s'inscrit dans la perspective de développer un capteur de vision pour les véhicules intelligents dont le coût permette une accessibilité au plus grand nombre.

1) STÉRÉOSCOPE À BASE DE WEBCAMS

Comme nous venons de l'évoquer, l'utilisation de caméras industrielles en tant que constituants d'un stéréoscope conduit inévitablement à une solution chère, dotée de performances assez variables. Dans l'optique d'abaisser le coût de revient du capteur utilisé, le remplacement de ce type de caméras « traditionnelles » par du matériel plus largement diffusé s'avérait inévitable. La première étape a consisté à évaluer ce qu'il était possible d'attendre de la part des dispositifs de prise de vue sans doute les moins chers du marché que constituent les Webcams, basées sur des capteurs CMOS, dans le cadre d'applications ayant pour objectif la mesure tridimensionnelle [Zay.04a]. Bien plus que la « rétine » en elle-même, la cause première de la mauvaise qualité des images que fournissent en général ces caméras est liée à l'optique. En conséquence, la première adaptation que nous avons faite a consisté à remplacer les optiques fournies par des optiques à vis de « type – C », ces dernières présentant de bien meilleures caractéristiques optiques (de moindres distorsions) que ce qui est fourni d'origine. Evidemment, ce changement d'optiques a nécessité le ré-empaquetage total des caméras. Cela nous a donné l'occasion de concevoir, à partir de ces « *Webcams modifiées* » des dispositifs simples, robustes et bon marché, qui conviennent parfaitement aux exigences des applications d'extérieur.

Dans ce qui suit, nous allons présenter quelques résultats issus des expériences effectuées avec le capteur stéréoscopique en question.

Après avoir appliqué à ce système la procédure de calibration décrite plus haut, nous sommes en mesure de tester le capteur de stéréovision. Les premiers tests ont été menés avec le système que vous pouvez voir sur la figure ci-dessous (Figure III.44) :

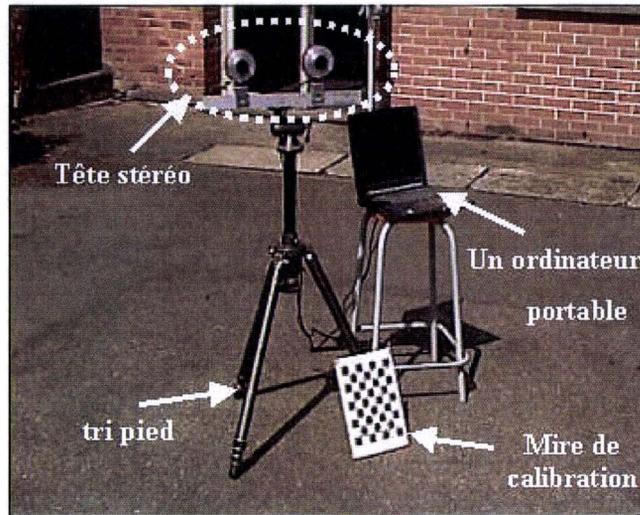


Figure.III.44: Le système de stéréovision basé sur des Webcams modifiées.

La figure III.45 détaille la tête stéréo utilisée, et plus spécialement le type d'objectifs qui peut être utilisée avec les nouvelles caméras :

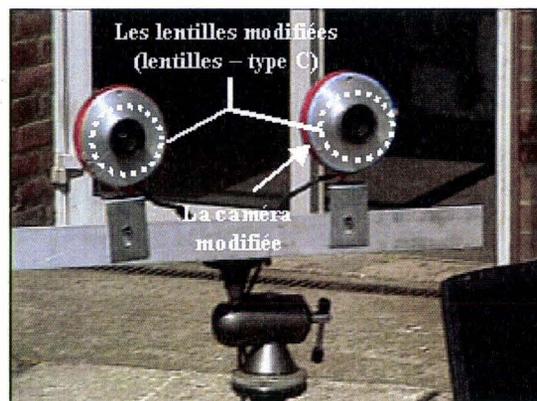


Figure. III.45: Vue détaillée de la tête stéréo.

La première expérience que nous présentons ici a consisté à estimer le déplacement d'un cube sur le sol. Nous avons effectué une série de séquences comprenant chacune plus de 300 déplacements. A chaque fois, nous avons mesuré la distance réelle de façon précise (erreur inférieure au demi millimètre). Les Figures III.46 – a et III.46 – b montrent la zone de déplacement du cube. Les Figures III.47 et III.48 représentent les déplacements estimés contre les déplacements réels.

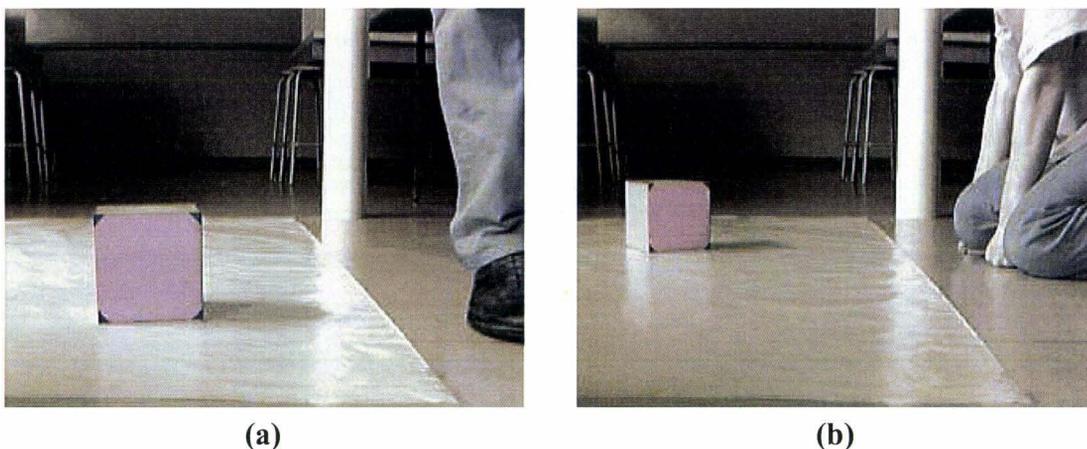


Figure. III.46 : La position la plus proche (a) et la plus éloignée (b) du cube.

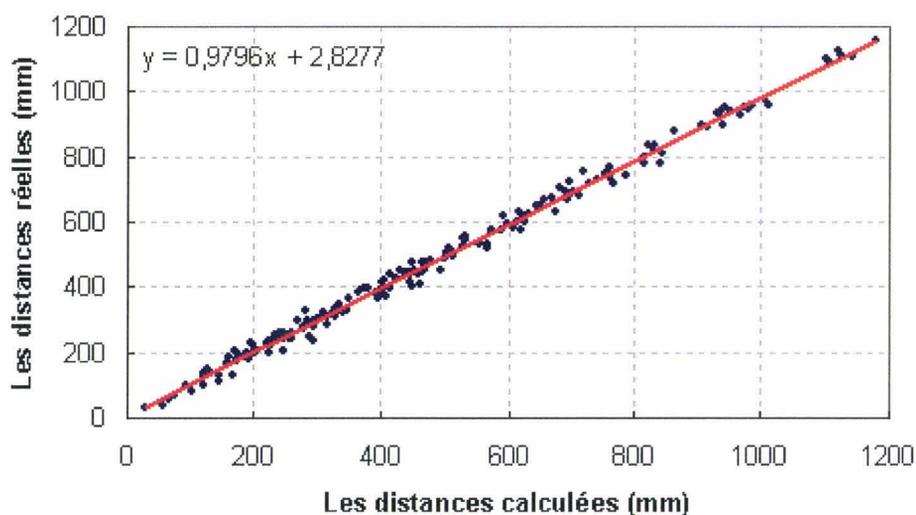


Figure.III.47 : comparaison entre déplacements estimés et déplacements réels (1^{ère} série)

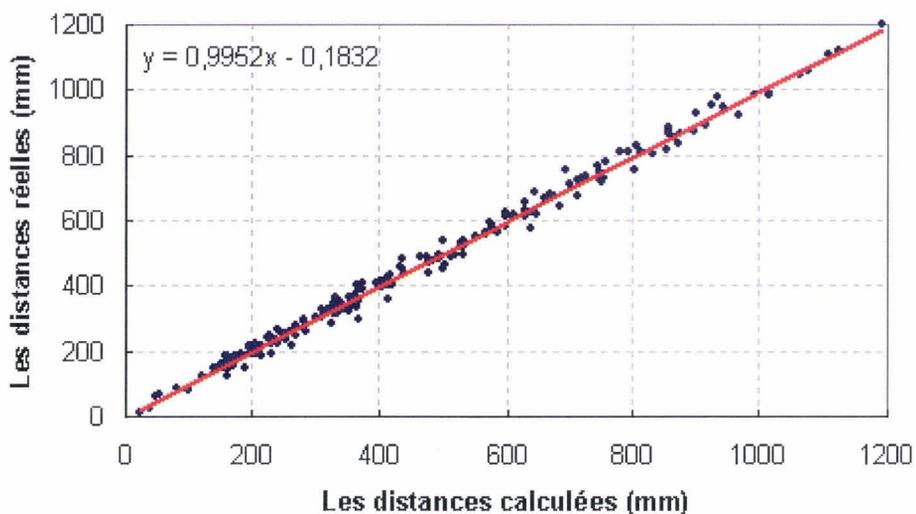


Figure.III.48: Comparaison entre déplacements estimés et déplacements réels (2^{ème} série).

Comme attendu, les déplacements estimés sont très proches des vrais déplacements. La caractéristique du capteur de stéréovision est linéaire avec une très bonne approximation (cela se vérifie au travers d'une très bonne corrélation entre les distances calculées et réelles et par la faible variance affectant l'erreur d'estimation). Notons que l'erreur d'estimation semble ne pas être corrélée à la distance elle-même (au moins sur l'espace d'évolution considéré ici). Ce premier exemple démontre qu'il est possible de mesurer des déplacements dans des régions restreintes (entre 0 et 2 mètres) avec une bonne précision (de l'ordre du millimètre), suffisante pour certaines applications (localisation d'objets se trouvant dans une zone de travail automatisée, dans une zone d'accident, de fouille, de crimes, etc..) et ceci avec du matériel très peu onéreux.

A l'aide du même dispositif, La seconde application présentée ici consiste à estimer les dimensions d'objets extérieurs, associés par exemple à des constructions ou des bâtiments (voir, Figure III.49 – a et III.49 – b).

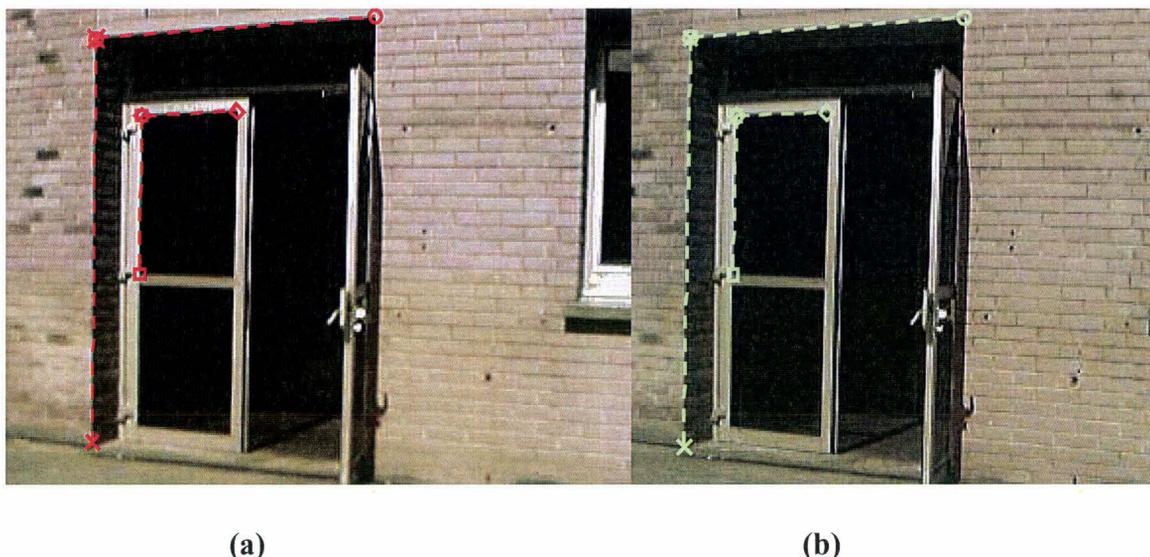


Figure.III.49 : Les images, gauche (a) et droite (b) prises par le capteur

Dans le contexte de cette application, les points dans l'image de gauche et leurs correspondants dans l'image de droite ont été extraits à la main, mais cette procédure de mise en correspondance peut être automatisée (pour plus d'informations voir [Zay.03a, Zha.94]). La largeur réelle de l'ouverture murale dans laquelle la porte est encastrée est de 1m 800 mm, et la largeur estimée est de 1m790mm. La hauteur de cet élément est de 2 m 400 mm et celle estimée est de 2 m 404 mm. En ce qui concerne la fenêtre sur le côté gauche de la porte, les dimensions réelles sont de $0.630m \times 0.945m$ et ceux issues du système sont $0.661m \times 0.945m$.

En fonction des applications, cette précision peut être suffisante. Par ailleurs, l'adaptation de la configuration du système de stéréovision (la position relative des deux caméras et la focale des objectifs utilisés) à une tâche de mesure spécifique peut améliorer considérablement la précision des mesures.

2) STEREOSCOPE A BASE DE CAMESCOPES

La solution présentée dans la section précédente donne de bons résultats de reconstruction tridimensionnelle (compte-tenu du coût de ce dispositif), et valide directement la partie conceptuelle du capteur stéréoscopique, en particulier pour ce qui touche à ses particularités géométriques. Toutefois, les Webcams que nous avons utilisées présentent d'importantes lacunes en terme d'acquisition d'image. En particulier, celles-ci ne s'adaptent pas facilement aux changements brusques de l'éclairage, ce qui détériore considérablement la qualité du signal recueilli et influence ainsi négativement la partie « détection de la plaque d'immatriculation » qui est, rappelons le, la « pierre angulaire » de notre chaîne de traitement pour l'attelage virtuel.

Fort de ces constats, nous avons effectué des tests sur un capteur stéréoscopique conceptuellement identique au premier, mais utilisant des caméscopes à la place des Webcams. Cette nouvelle configuration matérielle a amélioré considérablement la qualité des images obtenues (résolution plus importante, des couleurs plus contrastés, des temps d'acquisition plus rapides, utilisation du IEEE1394 au lieu de USB 1.1, etc). Comme nous l'avons évoqué, un des autres intérêts des caméscopes est leur capacité à gérer par eux-même et de façon efficace les variations de luminosité affectant la scène observée, à l'inverse des Webcams utilisées dans notre première expérience. Par ailleurs, une fonctionnalité supplémentaire de « vision nocturne » est envisageable, après avoir constaté que la plaque d'immatriculation est particulièrement visible dans le domaine infrarouge (longueur d'onde utilisée par la « torche » des caméscopes), ce qui faciliterait la détection de la plaque de nuit. Cette piste reste encore à explorer à l'heure actuelle. Les paragraphes qui suivent décrivent les différentes expériences effectuées avec ce second type de stéréoscope, embarqué cette fois-ci sur notre véhicule électrique expérimental.

Comme nous l'avons évoqué précédemment, les résultats présentés ci-dessus supposent que la calibration du capteur, a été déjà effectuée. Les tests ont été menés avec le système de la figure III.50 :

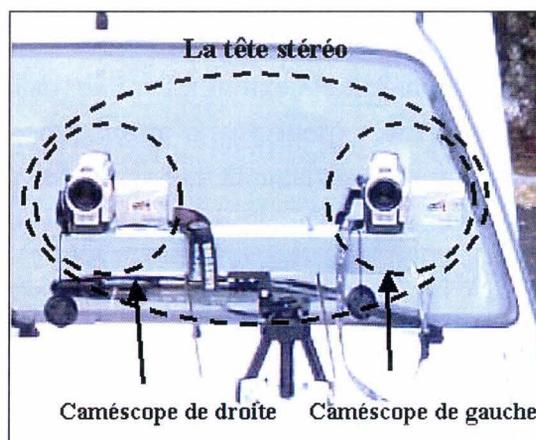


Figure III.50 : Système de stéréovision basé sur des caméscopes, embarqué sur le véhicule expérimental.

Comme on peut le constater sur la figure ci-dessus, la tête stéréo est composée de deux caméscopes de même type, ayant de ce fait les mêmes spécifications techniques. Celle-ci est montée et orientée sur le véhicule de telle sorte qu'on puisse effectuer le tracking des mouvements de la plaque d'immatriculation du véhicule suivi dans des conditions correspondant à une vitesse peu élevée (jusqu'à 50 km/h), impliquant une distance suiveur / suivi relativement faible. Notre véhicule de test est un « Renault express électrique » (voir Figure III.51) équipé de la tête stéréo sur son capot. Il dispose d'un ordinateur mettant en œuvre les traitements précédemment décrits sur les images acquises par les caméras. Les systèmes de bord bénéficient d'une autonomie électrique de plus de trois heures, grâce à un pack de batteries indépendantes de celles utilisées pour la propulsion. Dans un futur proche, celle-ci sera équipée de capteurs proprioceptifs (tels que des accéléromètres) qui permettront de prendre en considération l'influence de la dynamique du châssis du véhicule sur le processus de l'acquisition d'images.



Figure.III.51 : Le véhicule électrique expérimental de l'école des mines de douai, équipé d'une tête stéréo montée sur son capot et d'une unité de traitement embarquée.

L'une des expériences a consisté à suivre un véhicule (jouant le rôle de « tracteur ») pendant une vingtaine de minute en effectuant l'acquisition des paires stéréoscopiques correspondantes. Le parcours suivi était très riche et représentatif des cas réels au sens où il se

composait de sections planes, de bosses, de montée et de virages plus ou moins serrés, etc. La figure.III.52 représente quelques résultats de reconstruction 3D obtenus lors de tels essais :

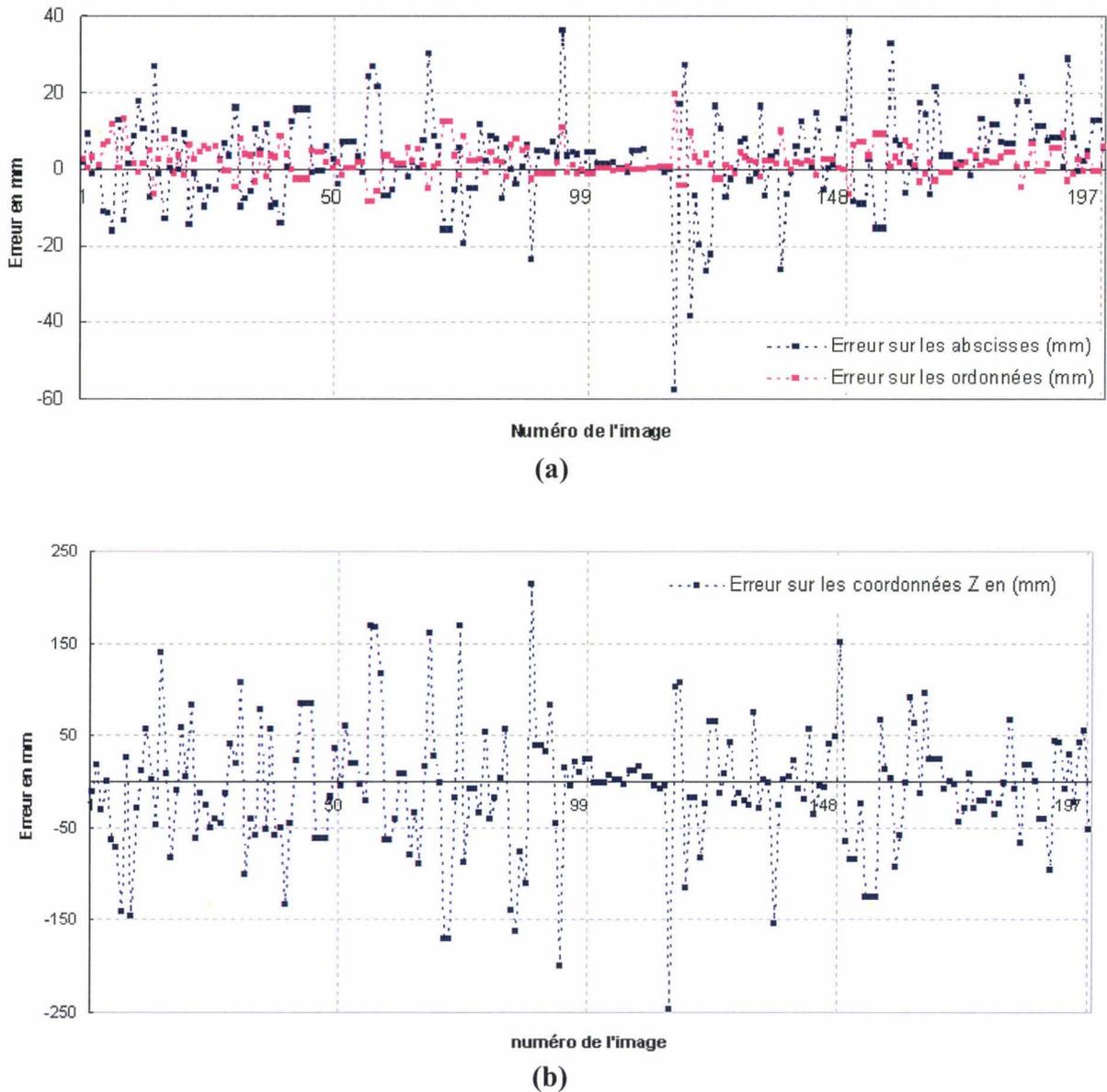


Figure III.52 : Les diagrammes (a) et (b) représentent les différences entre la reconstruction 3D directe du centre de la plaque d'immatriculation et l'estimation de ce dernier en tant que barycentre des coins de la plaque. Ces données constituent un des indices de « déformation » de la plaque à l'issue de la phase de reconstruction.

Etant donné que nous ne disposons pas, dans ce cas de figure précis, de mesures redondantes (mesures obtenues en utilisant un radar par exemple) permettant de vérifier la justesse de la reconstruction 3D effectuée avec le stéréoscope, nous avons estimé les erreurs commises de façon indirecte, en évaluant les « déformations » affectant la plaque après reconstruction. Ont été ainsi évaluées les « erreurs de planéité » et les erreurs dimensionnelles affectant la cible. L'estimation des erreurs de planéité a consisté à reconstruire tout d'abord

les cinq points caractéristiques de la plaque, à savoir ses quatre coins et son centre de gravité, de façon indépendante pour chacun d'entre eux. Ensuite, nous recalculons la position 3D du centre de la plaque (calcul barycentrique) en utilisant l'estimation précédente des coordonnées des quatre coins. En l'absence d'erreur, les coordonnées obtenues doivent être identiques à celles issues de la reconstruction 3D.

Les erreurs d'estimation commises sur les abscisses et les ordonnées en adoptant la procédure ci-dessus sont moins importantes que celles touchant la profondeur, mais d'une manière générale, les erreurs d'estimation sont tout à fait acceptables, puisque celles-ci ne dépassent pas les 4% des valeurs réelles dans les pires des cas.

La seconde technique consiste à retrouver les dimensions de l'entité suivie (la plaque d'immatriculation, de taille connue). Afin de vérifier que celles-ci n'ont pas subi de variations trop importantes, nous calculons quatre coefficients liés directement aux dimensions estimées de la plaque d'immatriculation d'une part et à ses dimensions réelles d'autre part. Ces coefficients (correspondant à des erreurs dimensionnelles relatives) s'expriment de la façon suivante :

$$\left\{ \begin{array}{l} C_1 = \frac{|l_1 - l_2|}{l_1} \\ C_2 = \frac{|h_1 - h_2|}{h_1} \\ C_3 = \frac{|l_1 - l_0|}{l_0} \\ C_4 = \frac{|h_1 - h_0|}{h_0} \end{array} \right. \quad (3.82)$$

l_1, l_2 : Représentent respectivement les largeurs « haute » et « basse » de la plaque d'immatriculation.

h_1, h_2 : Représentent respectivement les hauteurs « gauche » et « droite » de la plaque d'immatriculation.

l_0, h_0 : Représentent respectivement la largeur et la hauteur réelles de la plaque d'immatriculation.

Dans le cas d'une reconstruction 3D idéale, chacun de ces quatre coefficients doit être nul. La figure ci-dessous montre l'évolution des valeurs de ces coefficients en fonction de la distance relative estimée.

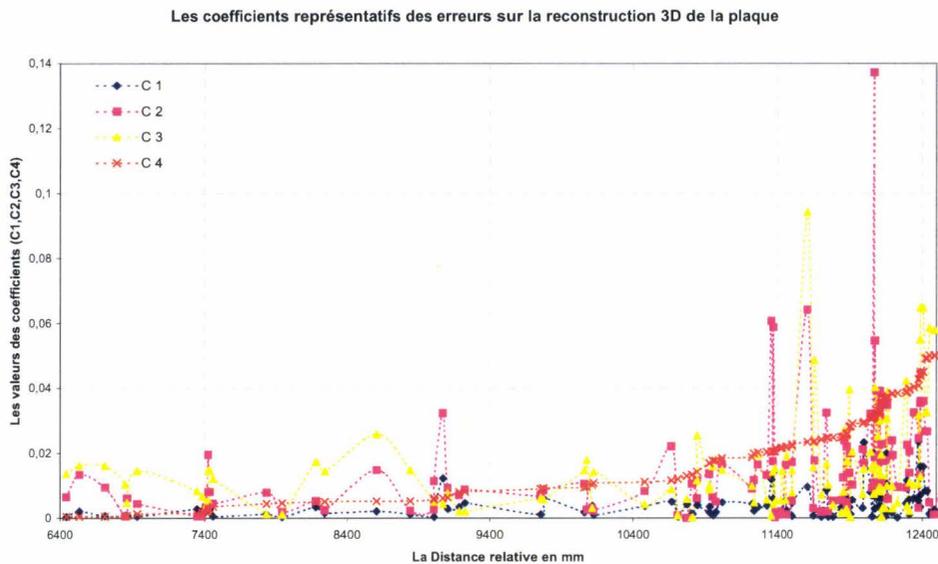


Figure III.53 : Evolution des coefficients (C_1 , C_2 , C_3 , C_4) caractéristiques de la reconstruction 3D en fonction de la distance relative.

En effet, C_1 et C_2 permettent de vérifier que la plaque d'immatriculation a bien conservé ses proportions après reconstruction. Par ailleurs, les coefficients C_3 et C_4 se réfèrent plutôt aux dimensions réelles de la plaque afin de donner une idée plus précise sur la justesse de la reconstruction dimensionnelle globale (affectant donc aussi l'estimation de la distance au véhicule suivi). D'une façon générale, le graphique de la figure III.53 nous confirme une fois de plus que les erreurs de la reconstruction effectuée sont assujetties principalement à la distance qui sépare l'objet observé de la tête stéréo, mais que cependant l'erreur relative commise reste tout à fait acceptable dans le domaine de fonctionnement du capteur (approximativement 3% d'erreur relative au pire à une distance estimée de 11 mètres).

La richesse de cette expérience nous a permis de souligner les points forts et faibles du système mis en place et les améliorations que nous pouvons amener par la suite dans le cadre du fonctionnement normal du capteur.

III.3.5.2. LE MODE DU FONCTIONNEMENT DEGRADE DU CAPTEUR STEREOSCOPIQUE

L'hypothèse majeure faite jusqu'ici, est que la plaque d'immatriculation appartient bien au champ de vision commun des deux caméras. Hélas, cette hypothèse n'est pas toujours vérifiée. En effet la plaque d'immatriculation peut, dans plusieurs cas de figure, sortir de ce champ commun et n'être plus visible que par une seule caméra à la fois (soit la caméra gauche ou la caméra droite). Un tel « incident » interdit bien évidemment l'usage des techniques de reconstruction 3D conventionnelles. Dans ce cas de figure, correspondant à un fonctionnement en mode dégradé, nous proposons une solution palliative, permettant d'extraire l'information nécessaire à l'attelage virtuel à partir d'une seule caméra et cela en

prenant en considération les aspects géométriques caractéristiques de la scène observée [Zay.04b].

A) LA LOCALISATION 3D DE LA PLAQUE

Dans cette partie, on propose un algorithme utilisant une seule caméra et des informations connues *a priori* de la scène observée pour la localisation tridimensionnelle de la plaque.

1) LE MODÈLE GÉOMÉTRIQUE DE LA CAMÉRA

Dans ce qui suit, nous considérons un modèle à *sténopée* de la caméra (Figure. III.7). Ce modèle relie l'espace 3D à sa projection perspective (c'est à dire, son image) sur la rétine de la caméra. Cette transformation est linéaire dans l'espace projectif et elle est décrite par l'intermédiaire des paramètres intrinsèques et extrinsèques de la caméra.

On rappelle ici que les paramètres intrinsèques de la caméra (tels que, la distance focale (f), la taille d'un pixel (k_u, k_v), et les coordonnées image $(u_0, v_0)^T$ du centre de projection) décrivent une transformation affine représentée par une mise à l'échelle, une rotation et une translation entre la caméra et la référence de la caméra.

Les paramètres extrinsèques, quant à eux, décrivent la transformation rigide du repère du monde vers le repère de la camera. Cette transformation est entièrement définie par une matrice de rotation 3×3 , et un vecteur de translation 3×1 . En utilisant les coordonnées homogènes dans l'espace de projection, on obtient :

$$\begin{pmatrix} x \\ y \\ s \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_A \begin{pmatrix} R & T \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (3.83)$$

Avec $(x \ y \ s)^T$, les coordonnées homogènes de la projection du point P sur l'image,

$\alpha_u = \frac{f}{k_u}$ et $\alpha_v = \frac{f}{k_v}$. (avec k_u et k_v la largeur et la hauteur des pixels, respectivement),

R la matrice de rotation,

$(X \ Y \ Z \ 1)^T$ les coordonnées homogènes du point P dans le repère de la scène,

T le vecteur de translation.

A la matrice de la projection perspective (cette matrice exprime la transformation entre le repère de la caméra et le repère de l'image).

L'ensemble des paramètres de ce modèle est issu d'une étape de calibration, déjà largement abordée ici.

Dans l'absolu, la localisation d'un point dans l'espace Euclidien tridimensionnel nécessite que celui-ci soit observé de différents points de vue. Cette exigence sera remplie, soit en observant un point de l'espace à l'aide de deux caméras différentes, soit en déplaçant une caméra selon une trajectoire connue.

2) LE PROFIL DE LA GEOMETRIE D'UNE PLAQUE D'IMMATRICULATION

Dans le contexte particulier de notre application, la plaque d'immatriculation est un rectangle définie par ses quatre coins, qui définissent également sa largeur et sa hauteur (dans notre cas, 520 x 110 mm).

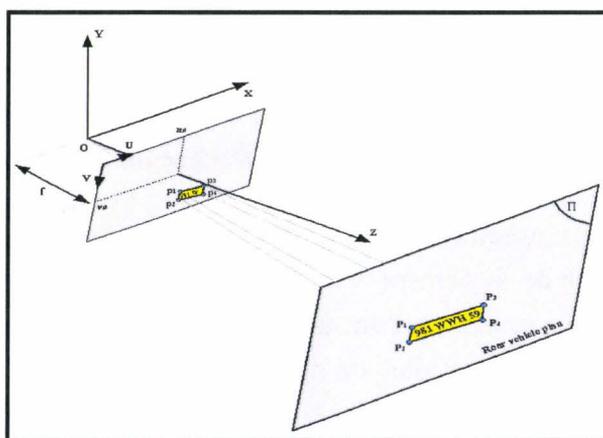


Figure III.54: La plaque telle qu'elle sera projetée sur la rétine de la caméra

La figure ci-dessus, décrit le processus de projection de la plaque d'immatriculation sur le plan rétine de la caméra. Dans le cas du contrôle automatique d'un véhicule, il sera raisonnable de supposer que le plan de la plaque observée et le plan de la caméra sont, avec une assez bonne approximation, perpendiculaires au plan de la route. Dans ce cas, la projection d'une ligne verticale dans l'espace 3D restera toujours verticale dans le repère des images [Zay.03b].

Dans ce qui suit, nous examinons la projection de la plaque d'immatriculation et sa projection sur le plan OYZ (figure.III.55 (a)) et le plan OXZ (figure.III.55 (b)), respectivement.

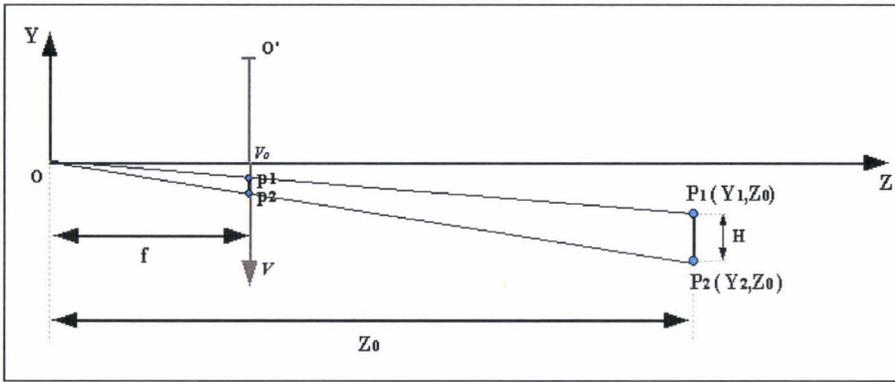


Figure III.55-a: Projection de la plaque sur le plan OYZ .

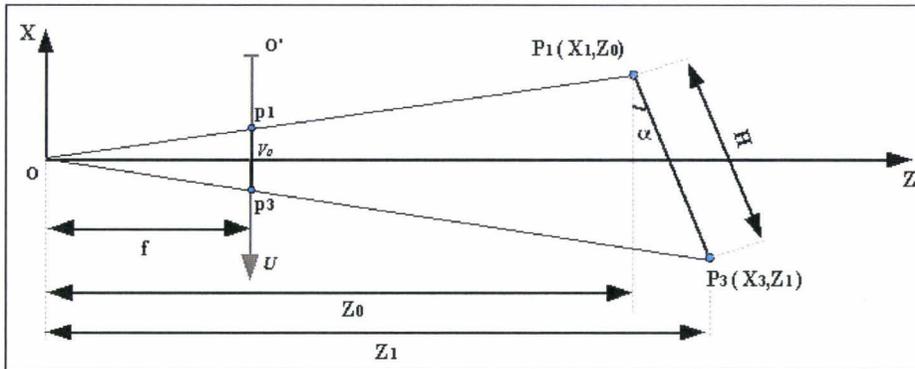


Figure III.55-b: Projection de la plaque sur le plan OXZ .

D'après l'équation (3.83) et en effectuant la triangulation passive, les coordonnées tridimensionnelles des quatre coins définissant la plaque sont obtenues grâce aux équations suivantes (3.84) :

$$\left\{ \begin{array}{l} X_1 = \frac{\alpha_v \cdot (u_0 - u_1)}{\alpha_u \cdot (v_2 - v_1)} \cdot H \\ X_3 = \frac{\alpha_v \cdot (u_0 - u_3)}{\alpha_u \cdot (v_4 - v_3)} \cdot H \\ Y_1 = \frac{(v_0 - v_1)}{(v_2 - v_1)} \cdot H \\ Y_2 = \frac{(v_0 - v_2)}{(v_2 - v_1)} \cdot H \\ Z_0 = \frac{\alpha_v}{|v_2 - v_1|} \cdot H \\ Z_1 = \frac{\alpha_v}{|v_4 - v_3|} \cdot H \end{array} \right. \quad (3.84)$$

H est la hauteur de la plaque, u_i et v_i avec $i > 0$ sont les abscisses et les ordonnées des projections des coins de la plaque sur le plan de la rétine de la caméra, en pixel.

Ces équations sont suffisantes pour localiser les coordonnées des coins de la plaque dans l'espace Euclidien tridimensionnel (voir, Figure (III.55)), relativement au repère de la caméra (ou également par rapport au repère du véhicule suiveur).

Par ailleurs, l'utilisation des équations (3.84) et l'hypothèse réaliste consistant à supposer connue la largeur de la plaque d'immatriculation (W valant généralement 510mm dans ce cas - voir annexe pour le format des plaques d'immatriculation en France) nous permettent de calculer l'angle α , correspondant à la différence de cap entre véhicule suiveur et véhicule suivi, sur le plan OXZ (Figure.III.55-b) grâce à :

$$\alpha = \arcsin \left[\left(\frac{\alpha_v H}{W} \right) \left(\frac{1}{(v_4 - v_3)} - \frac{1}{(v_2 - v_1)} \right) \right] \quad (3.85)$$

3) PREUVE ALGEBRIQUE ET EXTRACTION DE LA VITESSE RELATIVE

Soit un point de l'espace euclidien tridimensionnel P_i de coordonnées homogènes $(X_i \ Y_i \ Z_i \ 1)^T$. Ce point est observé par une caméra de centre de projection C tel que le repère du monde et le repère de la caméra soient confondus. Ceci se traduit mathématiquement par le fait que la matrice de rotation est réduite à l'identité et que le vecteur de translation reliant ces deux repères est nul. Dans ces conditions, les équations reliant le point P à sa projection s'expriment de la manière suivante :

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix}$$

Il est alors possible d'exprimer respectivement l'abscisse et l'ordonnée de la projection du point P_i , grâce à :

$$\begin{cases} u_i = \alpha_u \cdot \frac{X_i}{Z_i} + u_0 \\ v_i = \alpha_v \cdot \frac{Y_i}{Z_i} + v_0 \end{cases} \quad (3.86)$$

Ce qui implique que :

$$\begin{cases} X_i = \frac{(u_i - u_0)}{\alpha_u} \cdot Z_i \\ Y_i = \frac{(v_i - v_0)}{\alpha_v} \cdot Z_i \end{cases} \quad (3.87)$$

Supposons alors deux points P_1 et P_2 de l'espace, situés à la même distance Z du repère de la caméra (dont on rappelle qu'il est ici repère du monde). En s'appuyant sur les équations (3.86) et (3.87), nous pouvons déduire la distance Z en fonctions des coordonnées images des projections de P_1 et P_2 et de $\Delta Y = Y_2 - Y_1$ suivant :

$$Z = \frac{\alpha_v}{(v_2 - v_1)} \cdot \Delta Y \quad (3.88)$$

Cas particulier de la plaque d'immatriculation :

La plaque d'immatriculation d'un véhicule se déplaçant sur la route est supposée être contenue dans un plan perpendiculaire à celui de la route. De la sorte, les coins de la plaque qui appartiennent au même segment vertical forment une ligne parallèle au plan de la rétine de la caméra équipant le véhicule tracté (l'un des segments relie donc entre eux les coins de gauche, notés P_1 et P_2 , tandis que l'autre relie entre eux les coins de droite, notés P_3 et P_4). Si nous nous restreignons au plan OXZ , cela signifie que les coins de la plaque P_1 et P_2 sont à la même distance Z de l'origine, tandis que les points P_3 et P_4 sont à distance Z' de ce même point (voir la figure III.39-b). En nous ramenant à ce cas simple, nous obtenons les relations suivantes :

$$\begin{cases} X_i = \frac{u_i - u_0}{\alpha_u} \cdot Z \\ Y_i = \frac{v_i - v_0}{\alpha_v} \cdot Z \end{cases} \text{ avec : } i \in [1 \quad 2] \text{ et } Z = \frac{\alpha_v}{v_2 - v_1} \cdot H \quad (3.89)$$

$$\begin{cases} X_i = \frac{u_i - u_0}{\alpha_u} \cdot Z' \\ Y_i = \frac{v_i - v_0}{\alpha_v} \cdot Z' \end{cases} \text{ avec : } i \in [3 \quad 4] \text{ et } Z' = \frac{\alpha_v}{v_4 - v_3} \cdot H \quad (3.90)$$

Calcul de la différence de cap entre suivi et suiveur:

La Figure.III.55-b, permet d'établir que l'angle entre le plan de la plaque d'immatriculation du véhicule suivi et le plan rétine de la caméra du véhicule suiveur, qui correspond à la différence de cap entre ces deux véhicules, est donné par :

$$\sin(\alpha) = \frac{Z - Z'}{L} \quad (3.91)$$

En remplaçant Z et Z' par leurs expressions, nous obtenons :

$$\alpha = \arcsin \left[\left(\frac{\alpha_v \cdot H}{L} \right) \cdot \left(\frac{1}{(v_4 - v_3)} - \frac{1}{(v_2 - v_1)} \right) \right] \quad (3.92)$$

Calcul de la vitesse relative du véhicule suivi :

D'après (3.86) nous pouvons écrire :

$$v_i - v_0 = \alpha_v \cdot \frac{Y_i}{Z_i} \quad \Rightarrow \quad \frac{\partial(v_i - v_0)}{\partial Z} = -\alpha_v \cdot \frac{Y_i}{Z^2} \quad (3.93)$$

Ainsi :

$$\frac{\partial(v_i - v_0)}{\partial Z} = -\frac{(v_i - v_0)}{Z} \quad (3.94)$$

En dérivant (3.94) par rapport au temps, nous obtenons :

$$\frac{\partial v_i}{\partial t} = -\frac{(v_i - v_0)}{Z} \cdot \frac{\partial Z}{\partial t} \quad (3.95)$$

Ainsi, la composante suivant Z de la vitesse relative du véhicule suivi par rapport au véhicule suiveur est donnée par :

$$V_{z \text{ véhicule}} = -\frac{Z}{(v_i - v_0)} \cdot V_{v \text{ pixel}} \quad (3.96)$$

En adoptant le même raisonnement pour les abscisses des projections d'un point de l'espace, nous pouvons écrire les autres composantes du vecteur vitesse relative du véhicule suivi, qui sont alors :

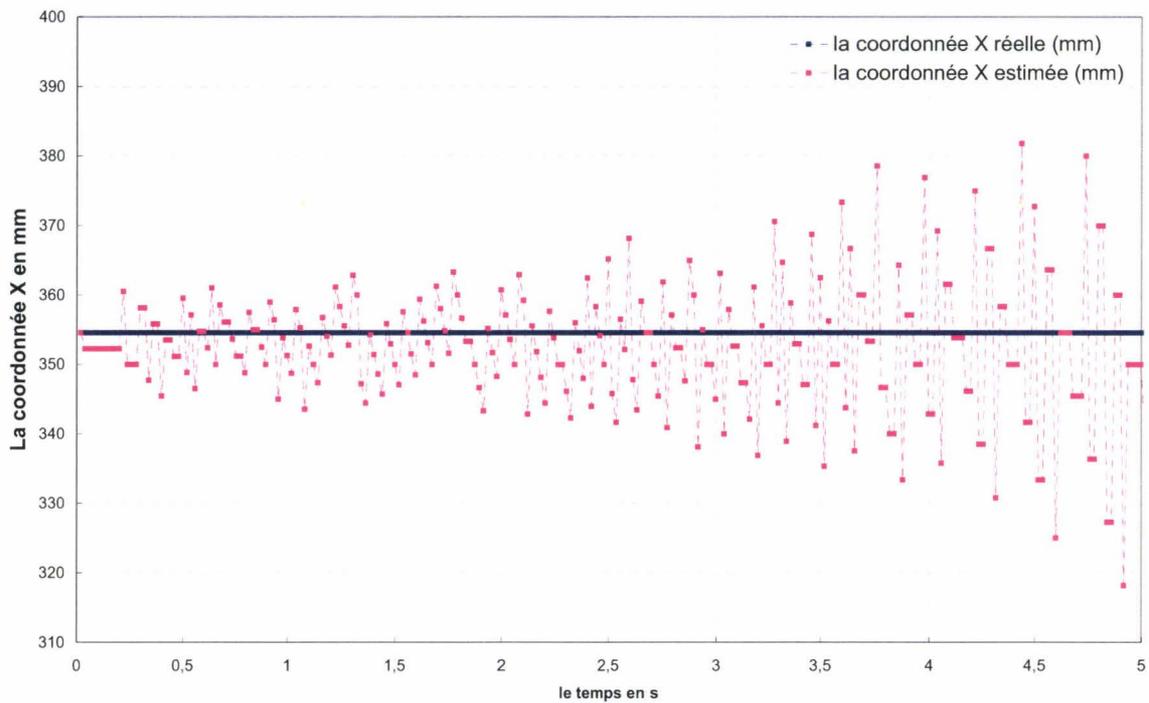
$$V_{x \text{ véhicule}} = \frac{X}{(u_i - u_0)} \cdot V_{u \text{ pixel}} \quad (3.97)$$

$$V_{y \text{ véhicule}} = \frac{Y}{(v_i - v_0)} \cdot V_{v \text{ pixel}} \quad (3.98)$$

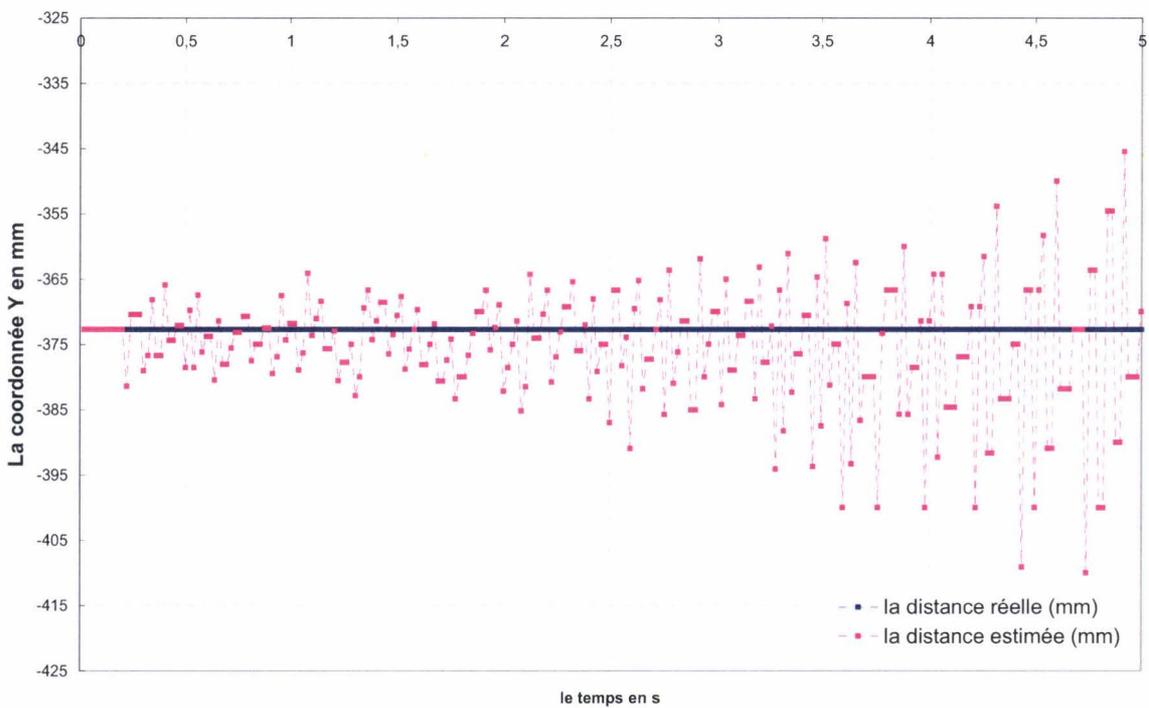
Il est clair que l'estimation de ce vecteur vitesse relative nécessite l'estimation préalable de la position 3D du point auquel il se rapporte. Le système (3.84) est alors tout indiqué pour fournir une telle estimation. Il découle de ce qui précède que si les dimensions de la plaque d'immatriculation sont connues et sous réserve que soient respectées les contraintes d'orientation entre la plaque d'immatriculation, le plan rétine de la caméra et la route (du moins en première approximation), les équations précédentes sont à même de fournir à la fois la position, la différence de cap et la vitesse relative du véhicule suivi à l'aide d'une unique caméra.

B) RESULTATS EXPERIMENTAUX

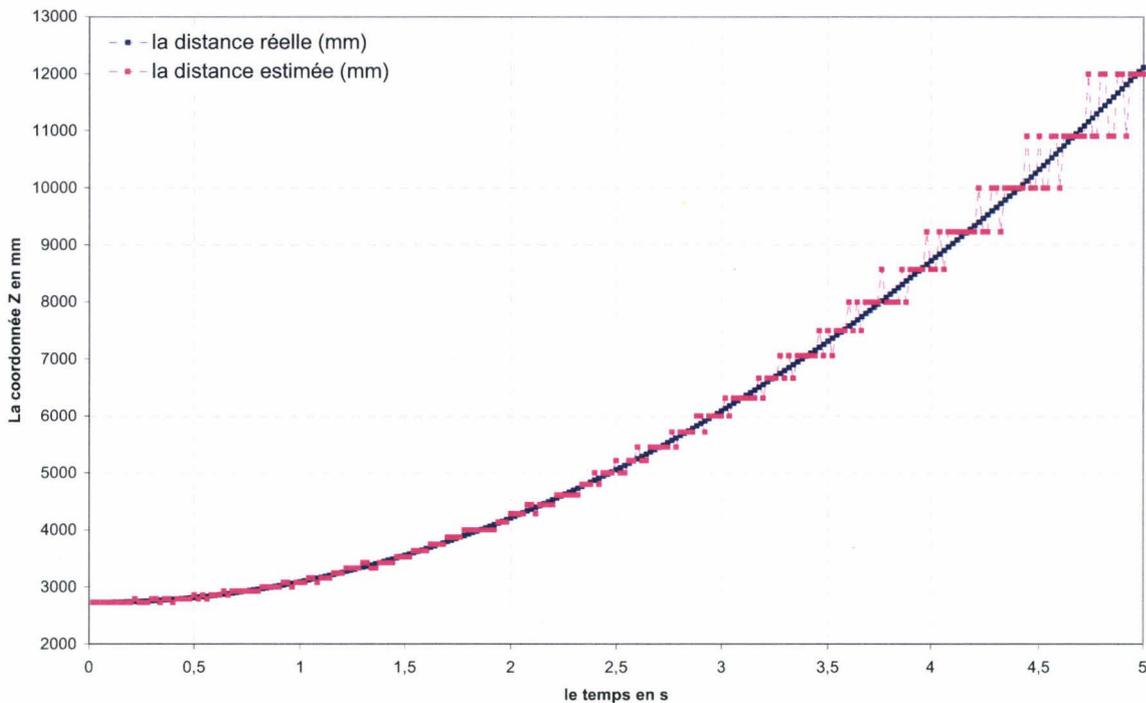
L'algorithme précédent a tout d'abord été testé par le biais d'une simulation. Dans un premier temps, nous avons donc procédé à la simulation du suivi du véhicule tracteur, supposé effectuer un mouvement uniformément accéléré. Bien que l'image de la plaque observée par la caméra soit « calculée » par le simulateur, la chaîne de traitement mis en œuvre (intégrant les phases de détection / tracking et extraction des primitives d'intérêt) est rigoureusement celle qui est mise en œuvre dans les cas réels. Au niveau des paramètres de simulation, le véhicule tracteur subit une accélération définie par le vecteur $(0.8 \ 0 \ 0)^T$ en m/s^2 . Les paramètres intrinsèques utilisés sont quant à eux : $\alpha_u = 1200$, $\alpha_v = 1200$, $u_0 = 320$ et $v_0 = 240$. Il est important de noter que dans le cadre de cet essai, le véhicule « suiveur » reste immobile et « regarde » le véhicule « suivi » partir au loin !



(a)



(b)



(c)

Figure III.56 : Coordonnées 3D réelles et estimées du centre de la plaque d'immatriculation du véhicule suivi en fonction du temps (a) Comparaison entre abscisse réelle et estimée, (b) Comparaison entre ordonnée réelle et estimée (c) Comparaison entre profondeur réelle et estimée.

Les résultats obtenus à l'issue de la simulation sont présentés sur la figure III.56. Ceux-ci montrent que les estimations des coordonnées du point de référence (point « d'accroche visuel ») correspondent à leur évolution effective, imposée par la trajectoire du véhicule suivi relativement au véhicule suiveur. On constate que l'erreur commise sur la distance augmente avec la distance séparant les deux véhicules. Ceci est dû principalement au fait qu'à mesure que la plaque d'immatriculation s'éloigne, sa taille relative dans l'image diminue, induisant un « bruit de quantification » de plus en plus important relativement aux pixels constituant l'image de la plaque. Une façon de pallier cet inconvénient consisterait, par exemple, à utiliser une focale variable (zoom) de sorte à pouvoir conserver une image de la plaque suffisamment grande, même lorsque celle-ci est éloignée, de sorte à conserver un bon « rapport signal/bruit ». Bien sûr, une telle approche n'est pas sans poser des problèmes spécifiques que des projets tels que RAVIOLI³⁵ ont pour ambition de résoudre.

Forts de nos résultats de simulation, nous avons procédé au traitement de scènes réelles. Les deux figures suivantes sont issues d'une séquence de suivi temps réel : le rectangle autour

³⁵ RAVIOLI : Radar et Vision Orientables Lidar

de la plaque d'immatriculation délimite la région d'intérêt dans laquelle la plaque a été localisée.

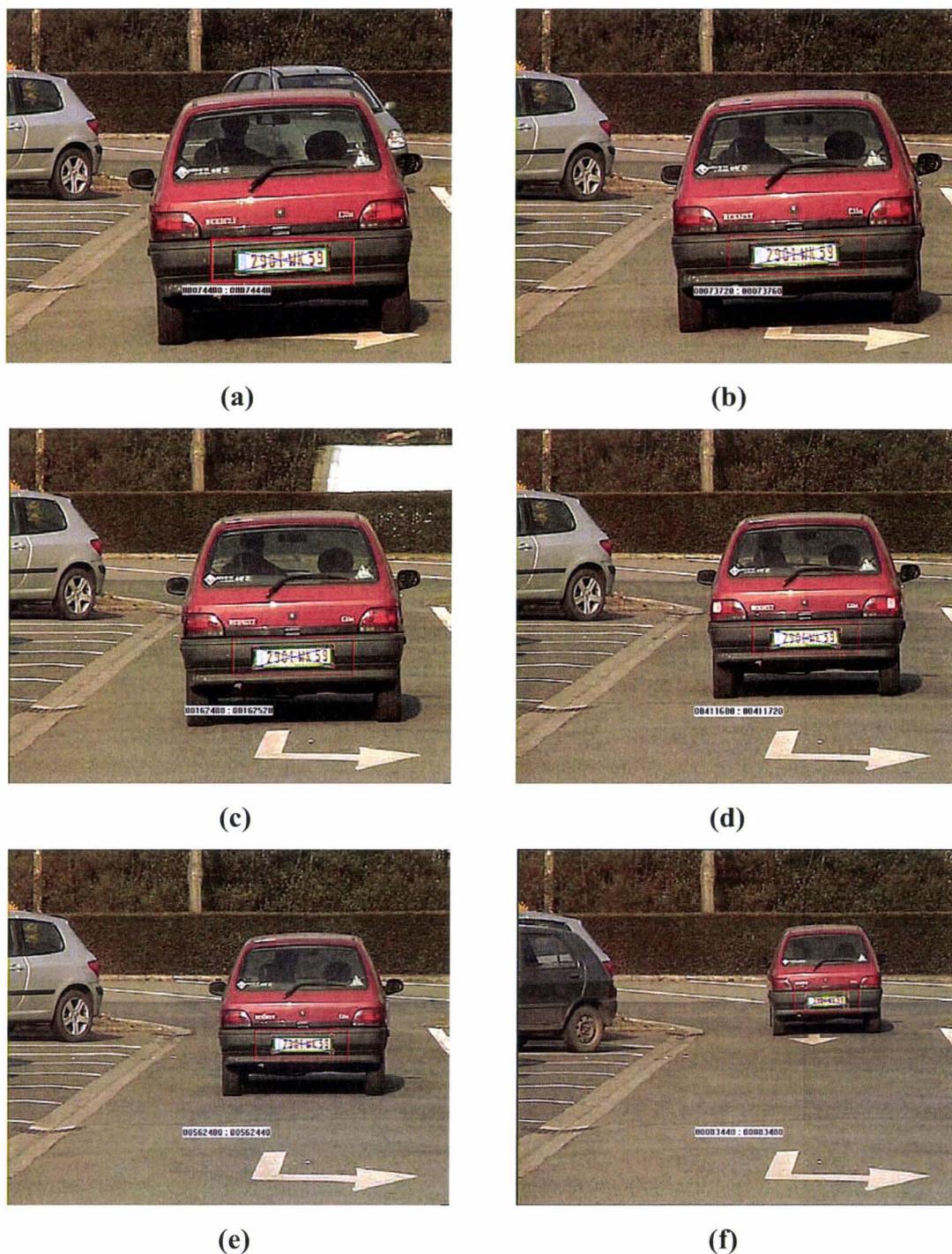


Figure III.57-a-f: Des exemples des régions d'intérêt (le rectangle rouge autour de la plaque d'immatriculation).

La Figure III.57 représente la distance estimée entre les véhicules en utilisant une caméra de focale 42 mm. La distance réelle varie entre 16 et 28 mètres; elle a été mesurée en utilisant un télémètre à infra rouge très précis. L'estimation demeure exploitable pour tous les essais effectués.

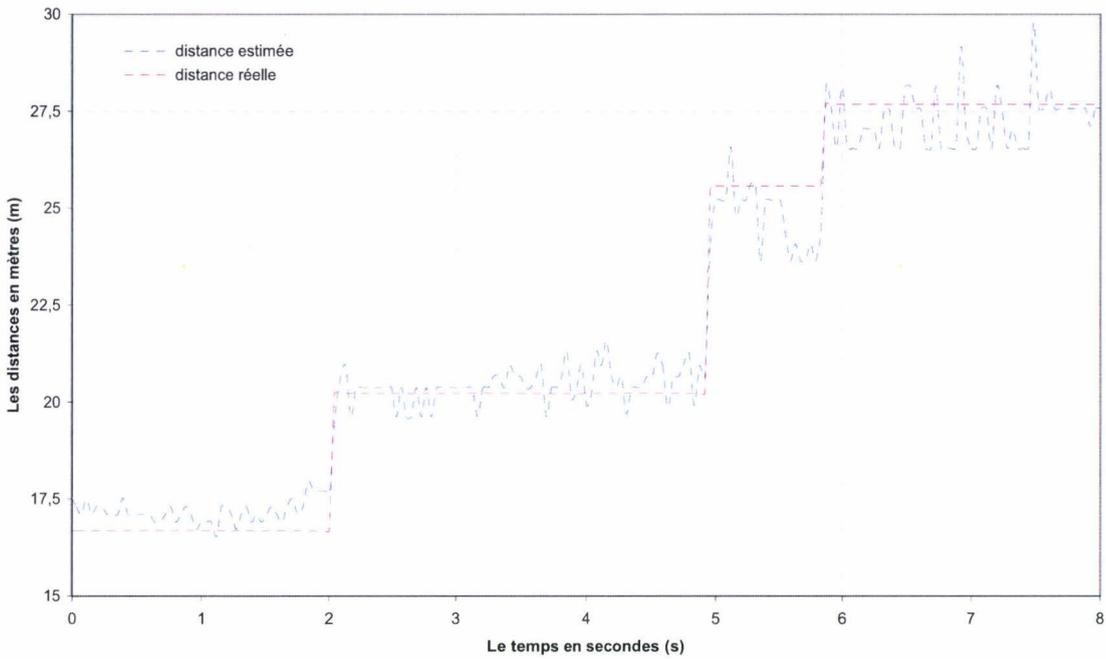


Figure III.58: Le suivi de la distance relative réelle entre deux véhicules.

La Figure III.58 représente la distance estimée par rapport à la distance réelle. On peut voir qu'elles sont fortement corrélées, démontrant ainsi l'efficacité du processus global. L'erreur est proportionnelle à la distance, pour les raisons déjà évoquées dans le cas des données simulées.

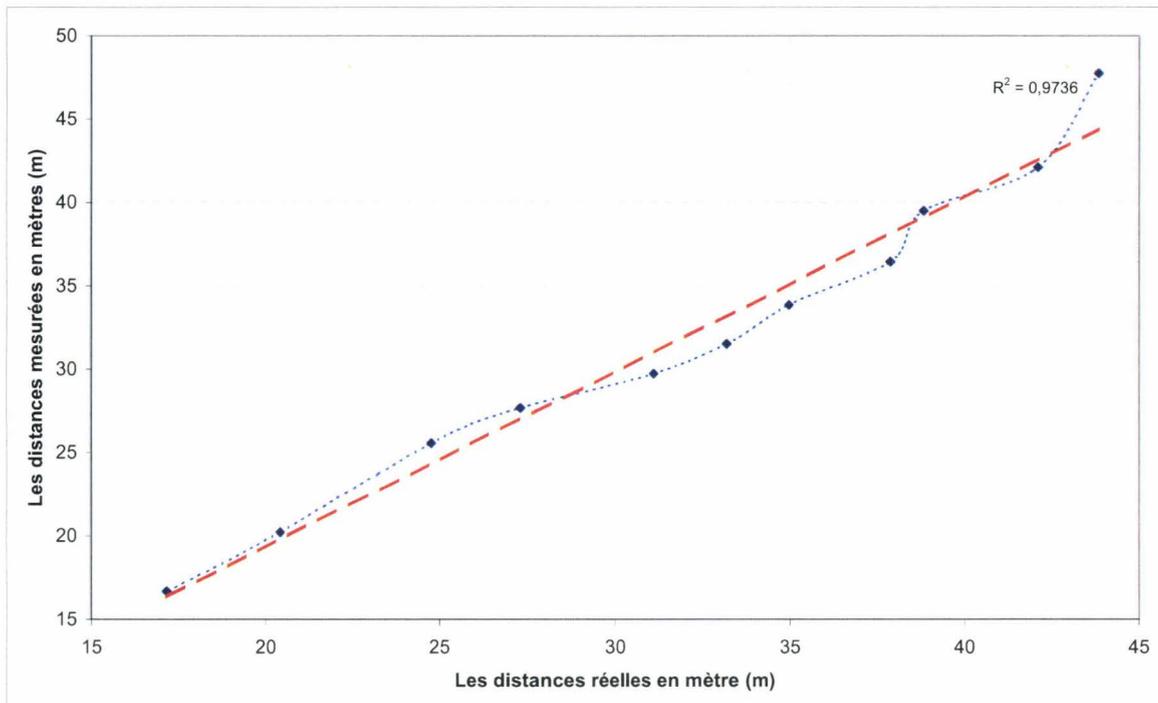


Figure III.59: Comparaison entre distance mesurée et distance réelle.

Une fois encore la figure III.59, fait apparaître clairement la forte corrélation existant entre les valeurs des distances réelles et mesurées (le coefficient de corrélation est égal à 0.9867), la courbe de tendance (en pointillé rouge sur le graphique) exhibe également une pente de droite pratiquement égale à 1, cela signifie que les distances mesurées reflètent avec une grande fidélité la distance réelle relative de la plaque d'immatriculation du véhicule tracteur, et ceci malgré l'utilisation d'une seule caméra. Ces données expérimentales valident les algorithmes exploités en mode dégradé pour la localisation du véhicule tracteur.

III.4.CONCLUSION

Les résultats que nous avons obtenus montrent qu'il est parfaitement possible d'estimer la distance relative avec un véhicule en mouvement en temps réel grâce à des caméras et un ordinateur standards. Les variations qu'on peut observer autour d'une valeur estimée sont dues essentiellement aux petites variations de la position de la plaque sur l'image. Le travail à venir consistera à réduire ce « bruit de détection » en fusionnant les données visuelles avec les mesures proprioceptives (accélération) et en calculant des « signatures visuelles » additionnelles, dans la région d'intérêt, ceci afin de renforcer l'étape de la détection.

Nous avons exposé dans cette partie une approche de suivi de véhicule basée sur l'observation d'une entité inhérente à celui-ci, à savoir sa plaque d'immatriculation. Pour y parvenir, notre algorithme commence par la détection de la plaque en adoptant soit une approche de décision statistique, soit le calcul d'un critère, l'un et l'autre basés sur l'analyse de la couleur des pixels dans l'image. Dans tous les cas, cette opération est conduite en temps réel. Afin d'améliorer encore le temps de réponse du système, une phase de *tracking* exploitant un filtre de Kalman basée sur un modèle de déplacement apparents de la plaque dans l'image est mis en œuvre. Le bénéfice de ce filtre est de permettre de réduire drastiquement la région d'intérêt ou se trouve probablement (au sens propre) le « point d'accroche visuel » que constitue la plaque d'immatriculation. Les expérimentations menées permettent d'affirmer qu'une telle approche permet en effet de rester « verrouillé » sur la région d'intérêt à une cadence de 10 Hz à l'aide de matériel tout à fait « grand public », qu'il s'agisse des caméras ou du calculateur utilisé. L'estimation de la position relative du véhicule suivi est effectuée par triangulation passive lorsque les points caractéristiques ont pu être isolés dans les deux images (mode normal), mais celle-ci peut être aussi effectuée en utilisant une seule caméra et des connaissances *a priori* sur les dimensions de la plaque, sous couvert d'hypothèses réalistes relativement à la disposition de celle-ci par rapport à la route et le plan rétine de la caméra utilisé. Les performances affichées par notre système sont probantes et répondent aux contraintes que nous nous étions fixées au début de ce travail. L'approche consistant à vouloir assurer l'attelage virtuel entre deux véhicules en utilisant la vision comme seule source de données extéroceptives paraît donc tout à fait réaliste. Cependant, certains aspects de notre chaîne de traitement nécessitent encore d'être renforcés et des approches complémentaires méritent d'être explorées. Quoiqu'il en soit, le système développé constitue une base solide sur laquelle il est désormais possible de s'appuyer afin d'aboutir, dans un

avenir très proche, à un système pleinement opérationnel et fiable dans toutes les conditions d'exploitation possibles.

CONCLUSION GÉNÉRALE

La première partie de ce mémoire se voulait une description précise de la problématique attaché à « l'attelage virtuel » dans le domaine, plus large, des « véhicules intelligents ». L'un des objectifs de l'introduction présentée était de fournir les éléments permettant d'effectuer une distinction claire entre la « conduite automatique » au sens large et l'attelage virtuel. Cette dernière approche se veut en effet plus « restrictive », au sens où le problème à résoudre, à savoir la mise en place d'un lien virtuel entre deux véhicules, se veut moins large que celui consistant à doter un engin mobile d'une complète autonomie. L'approche « attelage virtuel » apparaît bien davantage comme un complément aux algorithmes plus « généralistes » de conduite automatique que comme une alternative à ces derniers. Ce travail se voulait donc une approche à la fois réaliste et efficace du problème.

L'état de l'art relatif aux véhicules intelligents nous a permis d'affiner la position de nos travaux au sein de la riche contribution de la communauté scientifique, tout en mettant en exergue les potentialités et les faiblesses des différentes approches proposées jusqu'alors. Cette étude bibliographique nous a fourni la matière pour effectuer les choix stratégiques concernant notre capteur, tant en termes scientifiques que techniques et méthodologiques. C'est à ce stade que s'est révélé tout le potentiel d'une approche basée sur la vision stéréoscopique.

Dans ce contexte, l'approche « capteurs intelligents » fourni un cadre formel tout à fait adapté à la phase de conception et / ou de description du système à réaliser, raison pour laquelle ce mémoire aborde les concepts empruntés au domaine nous ayant permis de mener à bien la réalisation du capteur exploité dans le cadre de l'attelage virtuel. Munis de ces outils, les descriptions fonctionnelles et structurelles bénéficient d'un cadre rigoureux. Ainsi, les besoins spécifiques de notre capteur ont pu être décrits sous forme de services organisant les différents traitements à effectuer pour obtenir les informations nécessaires, à savoir la distance intervéhiculaire, la différence de cap entre « suiveur » et « suivi » ainsi que la vitesse relative du véhicule suivi par rapport au véhicule suiveur.

L'un des principaux problème que nous avons eu à résoudre au cours de ce projet consistait à établir dans quelle mesure la stéréovision pouvait être une approche « abordable » dans le domaine des transports intelligents. Le terme « abordable » est à considérer de différents points de vue, à savoir scientifique (sera-t-on réellement capable d'obtenir les informations voulues avec un temps de réponse, une fiabilité et une précision suffisantes ?), économique (le coût du capteur se doit d'être modéré si on souhaite pouvoir l'intégrer sur des véhicules particuliers), mais aussi ergonomique puisque le système doit pouvoir être utilisé par tout un chacun, sans nécessiter de formation ou de connaissances particulières. C'est en gardant à l'esprit cet ensemble de contraintes que nous avons abordé la mise en œuvre de notre capteur.

La troisième partie de ce mémoire a été consacrée à l'étude des différents aspects d'un capteur stéréoscopique, véritable « cœur » de notre système, afin d'identifier et de cerner les problèmes liés à la vision par ordinateur dans le contexte précis des transports routiers. L'étude menée a ainsi permis d'aborder les problématiques liées à la réalisation des services identifiés par le biais de l'approche « capteur intelligent » et de sélectionner des algorithmes basés sur le traitement d'images et la vision par ordinateur afin que notre dispositif fournisse bien les données attendues avec la précision et la fiabilité requises.

Ainsi, le dernier chapitre de ce mémoire permet de présenter dans le détail les différents algorithmes mis en œuvre, ainsi que la structure définitive retenue pour le capteur stéréoscopique embarqué sur le véhicule électrique expérimental. C'est en fait l'intégralité de la chaîne de traitements qui se trouve détaillée, depuis l'acquisition des images jusqu'à l'estimation de la configuration relative des deux véhicules en passant par la lecture de la plaque. Au niveau de l'acquisition des images, nous démontrons que l'utilisation de matériel « grand public » est parfaitement envisageable, ce qui permet de diminuer considérablement le coût matériel du dispositif. Le lecteur intéressé trouvera en annexe certaines précisions quant aux technologies mises en œuvre par ces matériels « bas-coût ». En matière de facilité d'utilisation du système, s'il est une phase qui peut poser problème à l'utilisateur, il s'agit de la phase de calibration de la tête stéréoscopique. La méthode mise en œuvre ici (calibration « hybride ») est d'utilisation aisée (grâce à l'utilisation d'algorithme *a contrario* relativement évolué) et la qualité de l'estimation des paramètres intrinsèques et extrinsèques ainsi déterminés est tout à fait satisfaisante, sans que ne soit nécessaire une lourde infrastructure (mire de calibration liée à la route ou un bâtiment particulier par exemple). Cette étape cruciale de calibration du stéréoscope a fait l'objet d'une étude permettant de définir les conditions aboutissant à des résultats de bonne qualité tout en garantissant une opération rapide. Ceci nous a permis de définir à la fois la taille et la structure de la mire de calibration utilisé, ainsi que le nombre d'images nécessaires à l'opération.

Certaines des informations issues de la procédure de calibration, permettant d'obtenir un modèle du processus de formation des images sur le plan rétine des caméras, seront exploitées afin d'obtenir des informations métriques relatives au « point d'accrochage visuel » choisi, en l'occurrence la plaque d'immatriculation. Il est par conséquent indispensable de pouvoir localiser cette dernière en temps réel dans une suite d'images ainsi que de la « certifier » (afin de répondre à cette simple mais importante question : « Sommes-nous en train de suivre le bon véhicule ? »). Pour mener à bien cette tâche en respectant les contraintes temporelles imposées par l'application, nous mettons en œuvre une séquence de type « détection / tracking / validation ». L'étape de détection s'appuie sur l'analyse de la couleur des pixels, où nous cherchons à isoler les pixels « jaune plaque ». Cette tâche de classification peut s'appuyer soit sur la base d'une décision statistique construite grâce à un ensemble d'apprentissage comprenant un grand nombre de plaque d'immatriculation, soit sur la mise en œuvre d'une « heuristique » exploitant une fonction évaluant le caractère « jaune plaque » d'un pixel. Les résultats expérimentaux obtenus valident l'une et l'autre de ces deux méthodes dont les forces et les faiblesses respectives sont détaillées dans les sections correspondantes du chapitre III.

Comme nous l'avons évoqué, le fait de pouvoir garantir que le suivi s'opère sur le bon véhicule est bien entendu capital ! La technique de « certification » mise en œuvre ici repose tout simplement sur la lecture des chiffres portés par la plaque, ce qui permet en pratique de lever toute ambiguïté. Le module de reconnaissance de caractères s'appuie sur un perceptron à trois couches que différents essais nous ont permis de dimensionner afin de respecter les contraintes pesant sur les performances désirées pour le système. Différents résultats expérimentaux, validant globalement l'approche tout en mettant en lumière certaines de ses déficiences, ont été présentés. Compte-tenu du nombre et de la complexité des traitements mis en œuvre, l'obtention d'un temps de réponse compatible avec la tâche d'attelage virtuel a nécessité l'exploitation d'un algorithme de « tracking », dont le but est de diminuer considérablement la quantité d'informations à traiter afin de localiser et d'identifier la plaque. Dans sa version actuelle, l'algorithme exploite un filtre récursif de Kalman et l'analyse des mouvements apparents des pixels d'intérêt afin de prédire la position future sur l'image de la projection de la plaque d'immatriculation. Les résultats expérimentaux obtenus alors ont permis de constater que le système était alors parfaitement capable d'effectuer l'ensemble des traitements requis à la cadence imposée de 10 Hz. Les bons résultats obtenus en matière de localisation 3D sur la base de la triangulation passive valident alors pleinement l'approche proposée. Cependant, il est relativement « fréquent » que la plaque d'immatriculation sorte du champ commun des deux caméras du stéréoscope. Pour demeurer capables de fournir les informations demandées, nous avons mis en place un algorithme permettant de pallier l'absence de l'information issue d'une seule caméra en exploitant les informations connues *a priori* sur la géométrie du capteur stéréoscopique et des hypothèses réalistes quant aux dispositions relatives de la plaque, de la caméra utilisée et de la route. Dans ce mode dégradé, comme l'attestent les résultats expérimentaux obtenus, le système reste alors parfaitement capable de fournir une estimation des distance, vitesse, et orientation relatives entre véhicule « tracteur » et véhicule « tracté ».

En définitive, le travail présenté ici a montré que la définition, la construction et la mise en œuvre d'un capteur de vision adapté à la tâche d'attelage virtuel est parfaitement envisageable. La solution proposée ici se veut simple d'utilisation (ceci afin de pouvoir être exploitée par le plus grand nombre), flexible (le système est ouvert et peu s'adapter très facilement à d'autres modèles de caméra ou de calculateur), à bas coût (car pouvant se contenter d'un matériel très largement diffusé) et embarquable (de par sa consommation électrique modérée lorsqu'il est mis en œuvre sur des calculateurs « stand alone » dédiés à l'embarqué et à bas coût).

Cependant, malgré les succès rencontrés lors de nos expérimentations, certains points nécessitent d'être renforcés avant d'aboutir à une solution réellement utilisable en toute sécurité par le grand public. Les premiers axes de progrès, actuellement en développement au travers de nouveaux projets de recherche, consistent à rendre la phase de *tracking* plus performante que ce qu'elle est actuellement. En particulier, l'utilisation de données proprioceptives, telles que des mesures d'accélération, permettra de gérer les perturbations induites par les mouvements propres de la caisse du véhicule embarquant le capteur

relativement à son châssis. Ensuite, nous avons évoqué le fait que lorsque le véhicule « tracteur » s'éloigne du véhicule « tracté », le « bruit » sur l'estimation produite augmente. Une solution consisterait à pouvoir « zoomer » dynamiquement sur la plaque d'immatriculation afin de garantir à cette dernière une taille minimale dans l'image (de sorte à opérer à rapport « signal / bruit-de-quantification » à peu près constant). Ceci n'est pas sans poser de problèmes, du fait de la variation des paramètres intrinsèques des caméras qu'une telle opération produit inévitablement, et par la sensibilité accrue aux mouvements des deux véhicules lors de la prise de vue. Il faut alors envisager de pouvoir orienter le dispositif en temps réel en direction du véhicule tracteur et d'avoir accès à des fonctions permettant d'accéder aux paramètres de calibration correspondant à la configuration courante. Le projet de recherche RAVIOLI (RAdar, VIsion Orientable et LIdar), auquel l'équipe participe, a pour objectif la mise au point d'un stéréoscope doté de caractéristiques similaires. Pour terminer, la richesse des informations auxquelles la vision par ordinateur permet d'avoir accès constitue une porte grande ouverte vers des myriades d'applications et de services qui restent pour la plupart à inventer !

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Abe.96] S. Abe, M.S. Lan and R. Thawonmas, « Tuning of fuzzy classifier. Derived from Data » International Journal of Approximation. Reasoning, vol. 14, Issue 1, January 1996.
- [Adi.89] G. Adiv. « Inherent ambiguities in recovering 3D motion and structure from a noisy flow field ». IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(5) :477-489, May 1989.
- [Afn.91] AFNOR NF-X 50-151 « Analyse de la valeur, analyse fonctionnelle. Expression fonctionnelle du besoin et cahier des charges fonctionnel », décembre 1991.
- [Afn.96] AFNOR NF-Z 68-901 « Génie automatique. Représentation des systèmes de contrôle et de commande des systèmes automatisés de production », novembre 1996.
- [Ais.90] J.Aisbett. « An iterated estimation of the motion parameters of a rigid body from noisy displacement vectors. IEEE Transactions Analysis and Machine Intelligence, 12(11):1092-1098, November 1990.
- [Alh.98] H. Al haddad. « Contrôle par vision du mouvement d'un robot mobile en environnement naturel ».Thèse de doctorat, Université Paul Sabatier, laboratoire LAAS, Nov. 1998.
- [Arc.03] ARCOS 2003, « Rapport : Etude de definition de l'APD », Edition publique, avril 2002.
- [Aub.91] J.P. Aubert, P. Dixneuf, «Conception et programmation par objet : techniques, outils et applications», Edition Masson, Paris, 1992.
- [Bab.89] N.Baba, «Anew approach for finding the global minimum error function of neural networks», Neural Networks, vol.2, pp367-373, 1989.
- [Bay.95] M. Bayart, A.L. Gehin and M. Staroswiecki, «Smart Intelligent Design», MMAR'96: Methods and Models in Automation and Robotics, Miedzyzdroje, Pologne, pp. 727-732, 1996.
- [Bea.94] P. Beardsley, A. Zissermann, and D. Murray. «Navigation using affine structure from motion». In Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden, pages 85-96, 1994.

- [Ber.98] M. Bertozzi, A. Broggi, GOLD: “A parallel real-time stereo vision system for generic obstacle and lane detection”, *IEEE Transactions on Image Processing* 7 (1) (1998) 62– 81.
- [Ber.00] M. Bertozzi, A. Broggi and A. Fascioli, “Vision-based Intelligent vehicles: state of the art and perspectives” , *Journal of Robotics and Autonomous Systems*, 32(1):1-16, June 2000.
- [Bis.97] D. Bishop, “Vehicle-highway automation activities in the United States”, in: *Proceedings of the International AHS Workshop*, US Department of Transportation, 1997.
- [Bla.01] N. Blanc, “CCD versus CMOS – has CCD imaging come to an end?”. p. 131 – 137, 2001.
- [Boe.88] B.W. Boehm, TRW defense systems group, “A spiral Model of software development and enhancement”, *IEEE Computer*, 21 (5), May 1988, pp.61-72.
- [Boo.00] G. Booch, J. Rumbaugh, I. Jacobson, *UML, le guide de l'utilisateur*, Eyrolles, 2000.
- [Bou.98] S. Bougnoux, from projective to Euclidian space under any practical situation, a criticism of self-calibration. In *Proceedings of the 6th International Conference on Computer Vision*, pages 790-796, Jan. 1998.
- [Bri.86] J.E. Brignell, *Sensors in Distributed Instrumentation Systems, Sensors & Actuators*, Vol.10, pp.249, 1986.
- [Bro.95] A. Broggi, S. Bertè, “Vision-based road detection in automotive systems: A real-time expectation-driven approach”, *Journal of Artificial Intelligence Research* 3 325–348, 1995.
- [Bro.97] A. Broggi, G. Conte, F. Gregoretti, C. Sansoè, L.M. Reyneri, “The evolution of the PAPRICA system”, *Integrated Computer-Aided Engineering Journal (special issue on Massively Parallel Computing)* 4 (2), pages 114–136, 1997.
- [Bro.99a] A. Broggi, M. Bertozzi, A. Fascioli, “The 2000 km test of the ARGO vision-based autonomous vehicle”, *IEEE Intelligent Systems*, pages 55–64, 1999.
- [Bro.99b] A. Broggi, M. Bertozzi, A. Fascioli, G. Conte. « Automatic Vehicle Guidance : The Experience of the ARGO Autonomous Vehicle ». *World Scientific Publishing*, 1999.
- [Bro.86] T.J. Broida and R. Chellappa. “ Estimation of object motion parameters from noisy images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(1):90-99, January 1986.
- [Bro.71] D.C. Brown, Close-range camera calibration. *Photogrammetric Engineering*, 37(8):855-866, 1971.

- [Bru.88] A.R. Bruss and B.K.P. Horn. "Passive navigation". *Computer Vision, Graphics and Image Processing*, 42(1):130-132, April 1988.
- [Cab.99] F. Cabestaing, R. Yang, J.L. Bruelle et J.-G. Postaire, "Real Time Processing of Images Sequences : Application ti Introsion Detection in Secured Areas", *Proceedings of International Conference On Quality Control by Artificial Vision, QCAV'99, Trois_Rivières, Québec, Canada, 1999*, pp 101-106,.
- [Cha.98] P. Charbonnier, P. Nicolle, Y. Guillard, J. Charrier, "Road boundaries detection using color saturation", in: *Proceedings of the Ninth European Signal Processing Conference '98*, September 1998.
- [Che.76] P.P.S. Chen, "The entity-relationship model: toward a unified view of data", *ACM transaction on database systems*, Vol. 1, n°1, mars 1976, pp. 9-36.
- [Che.94] Y.Q. Cheng, R.T. Collins, A.R. Hanson, and E.M. Riseman. "Triangulation without correspondences". In *Proceedings of ARPA Image Understanding Workshop, Monterey, California, USA, pages 993-1000, November 1994*.
- [Coa.91] P. Coad, E. Yourdon, "Object-Oriented Analysis", Prentice Hall, Eaglewood Cliffs, New Jersey, 1991.
- [Cod.97] A. Coda, P.C. Antonello, B. Peters, "Technical and human factor aspects of automatic vehicle control in emergency situations", in: *Proceedings of the IEEE International Conference on Intelligent Transportation Systems '97, 1997*.
- [Cou.02] P. Coulombeau, C. Laugeau, "Vehicle yaw, pitch, roll and 3D Line shape recovery by vision", in: *Proceedings of the IEEE Intelligent Vehicle Symposium, Versailles, Paris, France, June 2002*.
- [Cri.91] J.D. Crisman, C.E. Thorpe, UNSCARF, "a color vision system for the detection of unstructured roads", in: *Proceedings of the IEEE International Conference on Robotics and Automation, Sacramento, CA, pages. 2496– 2501, April 1991*.
- [Del.95] D. Delfieu, « Expression et Validation de Contraintes Temporelles pour la Spécification des Systèmes Réactifs », Thèse de l'université Paul Sabatier, Toulouse, Janvier 1995.
- [Den.94] S. Denasi, C. Lanzone, P. Martinese, G. Pettiti, G. Quaglia, L. Viglione, "Real-time system for road following and obstacle detection", in: *Proceedings of the SPIE on Machine Vision Applications, Architectures, and Systems Integration III, pp. 70–79, October 1994*.

- [Dic.95] E.D. Dickmanns, “Expectation-based multi-focal vision for vehicle guidance”, in: Proceedings of the Eighth European Signal Processing Conference '95, Trieste, Italy, pp. 1023–1026, September 1995.
- [Dou.73] D. H. Douglas and T. K. Peucker. Algorithms for the reduction of the number of points required to represent a line or its caricature. *The Canadian Cartographer*, 10(2):112--122, 1973.
- [Eas.00] Eastman Kodak Company. “Kodak CMOS image sensors – white paper”, Rochester, 2000.
- [Esa.89] European Spacial Agency, “HOOD Reference Manual-HOOD User Manual”, September and December 1989.
- [Fai.75] W. Faig, Calibration of close-range photogrammetry systems: Mathematical formulation. *Photogrammetric Engineering*, 41(12): 1479-1486, 1975.
- [Fan.84] J.Q. Fang and T.S. Huang. “Solving three dimensional small rotation motion equations: Uniqueness, algorithms, and numerical results”. *Computer Vision, Graphics and Image Processing*, 26(2):183-206, May 1984.
- [Fau.86] O.D. Faugeras and G. Toscani, “The calibration problem for stereo”. *Proceedings of the IEEE Computer Vision and Pattern Recognition*, pp. 15-20, 1986.
- [Fau.87] O.D. Faugeras, F. Lustman, and G. Toscani. “Motion and structure from point and line matches”, In *Proceedings of the 1st International Conference on Computer Vision*, London, England, pages 25-34. IEEE Computer Society Press, June 1987.
- [For.03] D.A. Forsyth, J. Ponce. “Computer Vision: A modern approach”, Prentice Hall, 2003.
- [Fra.98] U. Franke, D. Gavrilu, S. Görzig, F. Lindner, F. Paetzold, C. Wöhler, “Autonomous driving goes downtown”, in: *Proceedings of the IEEE Intelligent Vehicles Symposium '98*, Stuttgart, Germany, pp. 40–48, October 1998.
- [Fun.95] B.V. Funt, G.D. Finlayson, “Color Constant, color Indexing”, *IEEE Transactions on Pattern analysis and machine intelligence*, 1995.
- [Geh.94] A.L. Gehin, “Analyse fonctionnelle et Modèle Générique des Capteurs Intelligents: Application à la Surveillance de l’Anesthésie », Thèse de l’université de Lille 1, Janvier 1994.
- [Gev.01] T. Gevers. “Robust Histogram Construction from Color Invariants”. University of Amsterdam, 2001.
- [Gia.86] J.M. Giachino, *Smart Sensors, Sensors & Actuators Vol.10*, pp.239, 1986.

- [Gin.04] D. Gingras, W. Johnson and M. St.-Pierre, “Véhicules Intelligents et Systèmes Intelligents Véhicule-Route”, Rapport du projet ATALNTIC, Work Group 2.2; Centre de recherche sur les transports, Québec, Canada, avril 2004.
- [Gol.98] J. Goldbeck, D. Graeder, B. Huertgen, S. Ernst, F. Wilms, “Lane following combining vision and DGPS”, in: Proceedings of the IEEE Intelligent Vehicles Symposium '98, Stuttgart, Germany, pages. 445–450, October 1998.
- [Gol.96] G. Golub and C. van Loan. Matrix Computations. The John Hopkins University Press, Baltimore, Maryland, 3rd edition. 1996.
- [Goo.XX] J. Goodman, Introduction To Fourier Optics, McGraw-Hill Physical and Quantum Electronics Series.
- [Gra.99] P. Grattoni, G. Pettiti and M.L. Rastello. “Experimental Set-up for Characterization of Automated Number Plate Recognizers”, Elsevier Measurement N° 26, 1999.
- [Hal.82] E.L. Hall, J.B.K Tio, C.A. McPherson and F.A. Sadjadi. Measuring curved surfaces for robot vision, Comput. J. 15 : 42-54, 1982.
- [Har.85] D. Harel, A. Pnueli, “On the Development of Reactive Systems, in Logic and Models of Concurrent Systems”, NATO Asi Series in Computer Science, pp. 447-489, 1985.
- [Har.88] C. Harris and M. Stephens, “ A combined corner and edge detector”, Fourth Alvey Vision Conference. pp. 147-151, 1988.
- [Har.94] R.I. Hartley. An algorithm for self calibration from several views. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 908-912, Seattle WA, June. 1994.
- [Har.95] R. I. Hartley. In Defence of the 8-point Algorithm. IEEE In Proceedings of the 5th International Conference on Computer Vision and Pattern Recognition. Pages 908-912, Seattle, WA, June 1995.
- [Har.97] R. Hartley., P. Sturm, “ Triangulation”, Computer Vision and Image Understanding number 2 vol. 68 p.146-157, 1997.
- [Har.99] R.I. Hartley. Kruppa’s Equations Derived from Fundamental Matrix. IEEE Transactions On Pattern Analysis and Machine Intelligence. 1999.
- [Har.00] R. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press. 2000.
- [Har.04] M. Hariti, « Une Méthode Rapide D’appariement d’Images Stéréoscopiques : Application à la Perception de l’Environnement d’Un Véhicule Routier », Thèse de l’Université de Technologie de Belfort-Montbéliard, décembre 2004.

- [Hat.90] D.J. Hatley, I.A. Pirbhai, « Stratégies de spécification des systèmes temps-réel (SA-RT) », Editions Masson, 1990.
- [Hel.03] H. Helmers, M. Schellenberg, “CMOS vs. CCD sensors in speckle interferometry”. *Optics & Laser Technology*, vol. 35 p.587 – 595, 2003.
- [Her.92] J. Hershberger and J. Snoeyink. “Speeding up the Douglas-Peucker line simplification algorithm”. In *Proc. 5th Intl. Symp. Spatial Data Handling. IGU Commission on GIS*, pages 134--143, 1992.
- [Hod.91] R. Hodgon, “The X-Model : a processe model for object-oriented software development”, *Proc. Of Le genie logiciel & se applications*, Toulouse, 1991.
- [Hor.96] R. Horaud et O. Monga. *Vision par ordinateur : outils fondamentaux*, Deuxième édition. Editions Hermès, 1996.
- [Hsi.02] J.W. Hsieh, S.H. Yu and Y-S. Chen, “Morphology-Based License Plate Detection from Complex scene”. *16th International Conference on Pattern Recognition (ICPR’02)*, vol. 03, 2002.
- [Hua.89] T.S. Huang and O.D. Faugeras. Some Properties of the E-matrix in two view motion estimation. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 1310-1312. 1989.
- [Igl.88] I.G.L. Technology, “SADT un langage pour communiquer”, Eyrolles, 1998.
- [Iss.04] H. Issa, « Mise En Correspondance Stéréoscopique Par Algorithmes Génétiques : Nouveaux Codages », Thèse de l’université des sciences et technologie de Lille, février 2004.
- [Ito.91] M. Ito. Robot vision modelling – camera modelling and camera calibration, *Adv. Robotics 5* : 321-335, 1991.
- [Jan.02a] JR, Janesick, “Dueling detectors – CMOS or CCD ?” *SPIE’s OE Magazine*, p. 30 – 3. 2002
- [Jan.02b] J. Janesick, “Lux transfer: complementary metal oxide semiconductors versus charge-coupled devices”. *Opt Eng* 2002; 41(6):1203 – 15, 2002.
- [Joc.93a] T.M. Jochem, S. Baluja, “Massively parallel, adaptive, color image processing for autonomous road following”, in: H. Kitano (Ed.), *Massively Parallel Artificial Intelligence*, AAAI Press/MIT Press, Cambridge, MA, 1993.
- [Joc.93b] T.M. Jochem, D.A. Pomerleau, C.E. Thorpe, “MANIAC: A next generation neurally based autonomous road follower”, in: *Proceedings of the Third International Conference on Intelligent Autonomous Systems*, Pittsburgh, PA, February 1993.
- [Joc.98] T.M. Jochem, S. Baluja, “A massively parallel road follower”, in: M.A. Bayoumi, L.S. Davis, K.P. Valavanis (Eds.), *Proceedings of the IEEE*

- Computer Architectures for Machine Perception, New Orleans, LA, pp. 2–12, December 1998.
- [Jod.94] J.F. Jodouin, « Les réseaux neuromimétiques », Hermes, Paris, 1994.
- [Jun.01] K. Jung, “Neural network-based text location in color images”. Elsevier, Pattern Recognition Letters vol. 22, 2001.
- [Kan.96] M.S. Kankanhalli, B.M. Mehtre and J.K Wu “Cluster Based Color Matching for Image Retrieval”. Pattern Recognition, vol.26, N°4, 1996.
- [Kar.94] A. Kara, D.M. Wilkes, and Kawamura. “3D structure reconstruction from point correspondences between two perspective projections”. Computer Vision, Graphics and Image Processing: Image Understanding, 60(3):392-397, November 1994.
- [Kav.00] S. Kavadis, B. Dierickx, D. Scheffer, A. Alaerts, D.U. Waerts and, J. Bogaerts, “A logarithmic response CMOS image sensor with on-chip calibration”. IEEE Solid-State Circuits 2000;35(8):1146 – 52, 2000.
- [Kim.95] K.I. Kim, S.Y. Oh, S.W. Kim, H. Jeong, C.N. Lee, B.S. Kim, C.S. Kim, “An autonomous land vehicle PRV II: Progresses and performance enhancement”, in: Proceedings of the IEEE Intelligent Vehicles Symposium '95, Detroit, MI, pp. 264–269, September 1995.
- [Kim.02a] S. Kim, D. Kim, Y. Ryu and G. Kim, “A Robust License-Plate Extraction Method under Complex Image Conditions”, 16th International Conference On Pattern Recognition, 2002.
- [Kim.02b] K.I. Kim, K. Juang and J.H. Kim, “Color Texture-Based Object Detection: An application to License plate Localization”. 1st International Workshop SVM 2002, Niagara Falls, Canada, 2002.
- [Kit.82] L. Kitchen and A. Rosenfeld. Gray level corner detection. Pattern Recognition Letters. 1:95-102, 1982.
- [Kol.95] D. Koller, J. Malik, Q.-T. Luong, J. Weber, “An integrated stereo-based approach to automatic vehicle guidance”, in: Proceedings of the Fifth International Conference on Computer Vision, Boston, MA, pp. 12–20, 1995.
- [Kru.95] W. Kruger, W. Enkelmann, S. Rossle, “Real-time estimation and tracking of optical flow vectors for obstacle detection”, in: Proceedings of the IEEE Intelligent Vehicles Symposium '95, Detroit, MI, pp. 304–309, September 1995.
- [Kwa.02] H. Kwasnicka, B. Wawrzyniak, “License plate localization and recognition in camera pictures”, AI-METH 2002 Artificial Intelligence Methods November 13-15, Gliwice, Poland, 2002.

- [Lab.02] R. Labayrade, D. Aubert and J.Ph. Tarel, "Real Time Obstacle Detection in Stereo Vision on non Flat Road Geometry through "V-Disparity" Representation", in: Proceedings of the IEEE Intelligent Vehicle Symposium, Versailles, Paris, France, June 2002.
- [Lab.04] R. Labayrade, « Détection D'obstacles Par Stéréovision », Thèse de l'Université Paris VI, Janvier 2004.
- [Lam.97] S. Lampierre-Couffin, "De la verification de cahiers des charges de systèmes à événements discrets à la validation des spécifications décrites en Grafcet », thèse de l'Ecole Normal Supérieure de Cachan, Toulon, janvier 1997.
- [Lan.01] J. Langheim, "CARSENSE – New environment sensing for advanced driver assistance systems", IEEE Intelligent Vehicles Symposium, National Institute of Informatics, Tokyo, Japan, May 2001.
- [Lec.87] Y. Lecun, « Modèles connexionnistes de l'apprentissage », Thèse de doctorat, Université de Paris IV, 1987.
- [Lec.98] S. Lecoeuche, « Reconnaissance de caractères industriels par application d'un système de réseaux de neurones à boucle de rétroaction », Thèse de doctorat de l'université des sciences et de la technologie de Lille, novembre 1998.
- [Lee.94] E.R. Lee, P.K. Kim, H.G. Kim, "Automatic recognition of a car license plate using color image processing", Proceedings of International Conference on Image Processing, pp.301-305, 1994.
- [Lem.99] E. Lemaire, "Spécification et Vérification Fonctionnelle et Comportementale d'un Équipement Intelligent", Thèse de l'université des sciences et technologies de Lille, décembre 1999.
- [Lob.03] J. Lobo, C. Queiroz and J. Dias, "World feature detection and mapping using stereovision and inertial sensors ", Robotics and Autonomuous Systems vol. 44 p. 69-81, 2003.
- [Luo.92] Q.-T. Luong. Matrice Fondamentale et calibration Visuelle sur l'Environnement – Vers une plus grande autonomie des systèmes robotiques. PhD thesis, Université de Paris-Sud, Centre d'Orsay, Dec. 1992.
- [Luo.93] Q.-T. Luong and O.D. Faugeras. Self-calibration of a stereo rig from unknown camera motions and point correspondences. INRIA, RR – 2014, Juillet 1993.
- [Luo.97] Q.-T. Luong and O.D. Faugeras. Self-calibration of a moving camera from point correspondences and fundamental matrices. The International Journal of Computer Vision, 22(3):261-289, 1997.



- [Lut.97] D. Luttenbacher D., « Modélisation du concept capteur intelligent par une approche orientée objet : application à un capteur intelligent de température », Thèse de l'université Henri Poincaré, Nancy, février 1997.
- [Lüt.98] M. Lützeler, E.D. Dickmanns, "Road recognition with MarVEye", in: Proceedings of the IEEE Intelligent Vehicles Symposium '98, Stuttgart, Germany, pp. 341–346, October 1998.
- [Mag.03] P. Magnan, "Detection of visible photons in CCD and CMOS: A comparative view". Nuclear Instruments and Methods in Physics Research A 504, p.199 – 212, 2003.
- [Mas.92] M.H. Masson, « Réseaux de neurones et reconnaissances de formes », Thèse de doctorat, U.T.C Compiègne, 1992.
- [Mau.96] M. Maurer, R. Behringer, F. Thomanek, E.D. Dickmanns, "A compact vision system for road vehicle guidance", in: Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria, 1996.
- [May.92] S. J. Maybank and O. D. Faugeras. A theory of self-calibration of a moving camera. The International Journal of Computer Vision, 8(2):123-152, Aug. 1992.
- [Mic.97] S.L. Michael Beuvais, C. Kreucher, "Building world model for mobile platforms using heterogeneous sensors fusion and temporal analysis", in: Proceedings of the IEEE International Conference on Intelligent Transportation Systems '97, Boston, MA, p. 101, November 1997.
- [Min.69] M. Minsky, S. Papert. "Perceptrons", MIT Press, Cambridge, MA, 1969.
- [Mor.77] J. More. The levenberg-marquardt algorithm, implementation and theory. In G.A. Watson, editor, Numerical Analysis, Lecture Notes in Mathematics 630. Springer-Verlag, 1977.
- [Par.03] M. Parent, G. Gallais, « Cybercars : review of first projects », International Conference on People Movers APM03. Singapore, September 2003.
- [Par.96a] JR. Parker, P. Federl, "An approach to License Plate Recognition", Computer Science Technical Report (1996-591-11), 1996.
- [Par.96b] J.R. Parker, "Algorithms for image processing and computer vision", Wiley Computer Publishing, 1996.
- [Per.05] M. PEREZ-PATRICIO, "Stéréovision Dense Par Traitement Adaptatif temps réel : Algorithmes et Implantations », Thèse de l'Université des Sciences et de la Technologie de Lille, février 2005.
- [Pol.99] M. Pollefeys. Self-calibration and metric reconstruction from uncalibrated image sequences. Ph.D. Thesis. Katholieke Univesiteit Leuven. May 1999.

- [Pom.96] D.A. Pomerleau, T.M. Jochem, "Rapidly adapting machine vision for automated vehicle steering", IEEE Expert 11 (2), 1996.
- [Pos.87] J.G. Postaire, "De l'image à la décision", Dunod, 1987.
- [Pre.88] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling. Numerical Recipes : The Art of Scientific Computing. Cambridge University Press. 1988.
- [Ram.04] A. Rambaldini, G. Alessandretti, G. Irato, P. Metetti, E. Nodari, A. Toffeti and C. Zoldane, "Iterative Design of a New On-Board System for public Vehicles: From the idea to prototype", FISITA 2004, Berceclona, May 2004.
- [Ros.58] F. Rosenblatt, "The Perceptron: a Probabilistic Model for Information Storage and organization in the Brain", Psychological Review, volume 65, pp. 386-408, 1958.
- [Rov.99] S. Rovetta, R. Zunino, "License-plate localization by using vector quantization", International Conference on Acoustics, Speech and Signal Processing, 1999.
- [Roy.70] W.W. Royce, "Managing the development of large software systems", Proc. WESTCON, California., USA, 1970.
- [Rum.86] D. Rumelhart, G. Hinton & R.J. Williams, "Learning internal representations by error propagation", in Rumelhart&McClelland Editions Parallel distributed processing – Explorations in microstructures of cognition. Cambridge, MA: MIT Press, 1986.
- [Sal.02] J. Salvi, X. Armangué and J. Batlle. A comparative review of camera calibrating methods with accuracy evaluation. The Journal of The Pattern Recognition Society. 35 : 1617-1635, 2002.
- [Sch.98] C. Schmid, R. Mohr and C. Bauckhage, Comparing and Evaluating Interest Points, Proceedings International Conference on Computer Vision, Norosa Publishing House, pp. 230-235, 1998.
- [She.90] R.N. Shepard, "Neural Nets For Generalization and Classification: Comment on Staddon and Reid (1990)", Psychological Review, Volume 97, Issue 4, pages 579-580, 1990
- [Smi.92] S. Smith. "Feature based image sequence understanding". Ph.D. thesis, Oxford: Department of Engineering Science, University of Oxford, 1992.
- [Smi.95] S.M. Smith, J.M. Brady, ASSET-2: "Real-time motion segmentation and shape tracking", IEEE Transactions on Pattern Analysis and Machine Intelligence 17 (8) 814–829, 1995.

- [Sta.94] M. Staroswiecki et M. Bayart, “Actionneurs intelligents », Edition Hermes, 1994.
- [Ste.02] B. Steux, C. Laugeaux, L. Salesse and D. Wautier, “Fade : A vehicle detection and tracking system featuring monocular color vision and radar data fusion”, in: Proceedings of the IEEE Intelligent Vehicle Symposium, Versailles, Paris, France, June 2002.
- [Stu.97] P.F. Sturm. « Vision 3D non calibrée : contributions à la reconstruction projective et études des mouvements critiques pour l’auto-calibrage ». Ph.D Thesis. Institut National Polytechnique de Grenoble. Décembre 1997.
- [Str.95] M.A. Stricker, M. Orengo, “Similarity of color images”, in Storage and Retrieval for Image and Video Databases, Proc 2420, 1995.
- [Tay.99] C.J. Taylor, J. Kořecká, R. Blasi, J. Malik, “A comparative study of vision-based lateral control strategies for autonomous highway driving”, International Journal of Robotics Research 18 (5) 442–453, 1999.
- [Tok.97] H. Tokuyama, “Asia–Pacific projects status and plans”, in: Proceedings of the International AHS Workshop, US Department of Transportation, 1997.
- [Tos.87a] G. Toscani and O.D. Faugeras. « Structure and motion from two noisy perspective views ». In Proceedings of IEEE International Conference on Robotics and Automation, Raleigh, North Carolina, USA, pages 221-227. IEEE Computer Society Press, March 1987.
- [Tos.87b] G. Toscani and O.D. Faugeras. “Structure from motion using the reconstruction & reprojection technique”. In Proceedings of the IEEE Computer Society Workshop on Computer Vision, pages 345-348. IEEE Computer Society Press, November 1987.
- [Tos.87c] G. Toscani, Systèmes de calibration et perception du mouvement en vision artificielle, Thèse de doctorat, Université Paris Sud, 1987.
- [Tou.04] P-F. Toulotte, S. Delprat, T-M. Guerra et J. Boonaert, “Robust fuzzy control with pole placement in LMI region : Application to an automated vehicle”. IEEE VTS-VPP 04, Paris, France, October 2004.
- [Tru.98] E. Trucco, A. Verri, “Introductory techniques for 3-D Computer Vision”, Prentice Hall, 1998.
- [Tsa.87] R.Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses. IEEE Journal of Robotics and Automation, 3(4): 323-344, Aug. 1987.

- [Van.92] J. Van der Spiegle, Computational Sensors of the 21st Century, Extended Abstracts of International Symposium on Sensors in the 21st Century, pp.51, Tokyo 1992.
- [Van.00] N. Vandenbroucke, "Segmentation d'images couleur par classification de pixels dans des espaces d'attributs colorimétriques adaptés. Application à l'analyse d'images de football ", Thèse de l'université des sciences et de la technologie de Lille, 2000.
- [War.86] P. Ward, S. Mellor, "Structured Development of Real-Time Systems". Englewood Cliffs, Yourdon Press, 1986.
- [Wel.95] G. Welch, G. Bishop. "An Introduction To Kalman Filter". Technical Report TR95-041, University of North Carolina at Chapel Hill, 1995.
- [Wen.87] J. Weng, T.S. Huang, and N. Ahuja. "Error Analysis of motion parameter estimation from image sequences". In Proceedings of 1st International Conference on Computer Vision, London, England, Pages 703-707. IEEE Computer Society Press, June 1987.
- [Wen.88] J. Weng, N. Ahuja, and T.S. Huang. "Closed-form solution + maximum likelihood: A robust approach to motion and structure estimation". In Proceedings of the Conference on Computer Vision and Pattern Recognition, San Diego, California, USA, pages 381-386, June 1988.
- [Wen.92] J. Weng, P. Cohen and M. Herniou. « Camera calibration with distorsion models and accuracy evaluation". IEEE Trans. Pattern Anal. Mach. Intell. 14 : 965-980, 1992.
- [Wil.94] R. Willson, Modelling and calibration of Automated Zoom Lenses, Ph.D. thesis, Department of Electrical and Computer Engineering, Carnegie Mello University, January 1994.
- [Yak.78] Y. Yakimovsky and R. Cunningham. "A system for extracting three-dimensional measurements from a stereo pair of TV cameras". Computer Graphics and Image Processing, 7(2):195-210, 1978.
- [Yam.84] H. Yamasaki, Approaches to Intelligent Sensors, Proc. of the 4th Sensor Symposium, pp.69, 1984.
- [Yam.91] H. Yamasaki, Sensors and Intelligent Systems, Y. Hashimoto and W. Day ed. Mathematical and Control Applications in Agriculture and Horticulture (Proc. IFAC/ISHS Workshop Matsuyama 1991) pp.349, Pergamon Press, 1991.
- [Yam.93] H. Yamasaki and K. Takahashi, An Intelligent Sound Sensing System Using Sensor Fusion, Digest of Technical papers Transducers' 93 pp 2 1993.

- [Zay.03a] M. Zayed., J. Boonaert. , “An effective Stereo-head points matching approach”. The Fourth International Conference AIAI, Montreal, Canada, June 2003.
- [Zay.03b] M. Zayed., J. Boonaert, “Obstacles detection from disparity properties in a particular stereovision system configuration”. The 6th IEEE on Intelligent Transportation Systems, Shanghai, China, p 311-316, October 2003
- [Zay.03c] M. Zayed, A. Rennuit, J. Boonaert, “ Plate Detection and Tracking for Automatic Car Following”. International Crashworthiness and Design Symposium "Innovation Technologique Pour les Transports Terrestres", Lille, France, Decembre 2003.
- [Zay.04a] M. Zayed, J. Boonaert and M. Bayart, “A low cost Stereovision Sensor for Non-Contact Measurement Purpose”, 7th IFAC Symposium on Cost Oriented Automation, Québec, Canada, June 2004.
- [Zay.04b] M. Zayed, J. Boonaert and M. Bayart, “License Plate Tracking for Car Following with a Single Camera”, 7th IEEE International Conference on Intelligent Transportation Systems, Washington DC, USA, October 2004.
- [Zha.94] Z. Zhang, R. Deriche, O. Faugeras and Q.-T. Luong. A Robust Technique for Matching Two Incalibrated Images Through the Recovery of the Unknown Epipolar Geometry. INRIA. RR – 2273. 1994.
- [Zha.00] Z. Zhang, « A Flexible New Technique for Camera Calibration ». IEEE Transactions on Pattern Analysis and Machine Intelligence. 22(11):1330-1334. 2000.
- [Zhu.86] X. Zhuang, T.S. Huang, and R.M. Haralick. “Two-view motion analysis: A unified algorithm”. Journal of the optical Society of America A, 3(9):1492-1500, September 1986.

ANNEXE A

A. L'objectif des caméras

Le rôle de l'objectif d'une caméra est de capter les rayons lumineux issus de la scène observée et de les faire converger vers la « rétine » du capteur d'image. On cherche à faire converger le plus de rayons possible en un seul point. La fonction réalisée par l'objectif se ramène à une opération de projection d'un espace 3D vers un espace 2D.

L'objectif se présente physiquement sous la forme d'un cylindre qui contient un certain nombre de lentilles assemblées d'une façon très précise. Il dispose également d'un certain nombre de « bagues » qui permettent le réglage de ses paramètres. Généralement les objectifs se fixent sur les caméras soit par une liaison de type « baïonnette » (comme les appareils photographiques) ou à vis (monture de type C ou CS). En ce qui nous concerne, nous avons réalisé un capteur stéréoscopique modulaire permettant d'utiliser soit des Webcams modifiées, utilisant des objectifs à vis (monture C), soit des caméscopes du commerce, ceux-ci disposant d'objectifs à baïonnette ou directement intégrés au boîtier.

A.1. Paramètres caractéristiques d'un objectif

Un objectif est caractérisé par les paramètres suivant :

A.1.1. La focale :

La focale est la distance entre le centre optique et le plan focal. La sous-unité en usage est le millimètre (*mm*). Plus la focale est petite et plus l'angle de vue est grand (c'est par exemple le cas des optiques des Webcams du commerce). On trouve principalement trois types d'objectifs :

- Focale fixe
- Focale variable manuelle
- Focale variable automatiquement (auto focus).

A.1.2. Le diaphragme :

Ce dispositif permet de régler la quantité de lumière qui traverse l'objectif, ce réglage peut s'effectuer de deux manières différentes :

- *Manuellement* : En manipulant la bague correspondante.
- *Automatiquement* : Le diaphragme est alors asservi, la grandeur de consigne se ramenant à la quantité de lumière reçue.

Le réglage manuel du diaphragme convient uniquement si la luminosité de la scène est peu ou pas variable, où si le capteur utilisé par ailleurs est à « temps d'intégration variable ». Lorsque la luminosité de la scène est amenée à varier rapidement pour le capteur, le réglage automatique s'impose.

A.1.3. Le Focus :

Le focus permet de faire varier légèrement la distance entre l'objectif et le plan principal. En effet, un objectif ne fait en toute rigueur converger correctement qu'un seul plan de la scène observée par la caméra sur le plan image. Régler le focus revient à changer le plan du monde réel correctement projeté. D'une manière générale, l'espace au-delà et en deçà de ce plan est plus ou moins flou. A l'instar du diaphragme, il existe des objectifs avec focus manuel ou focus automatique (auto-focus).

A.2. La Qualité d'un objectif

La qualité des objectifs est liée principalement aux lentilles qui composent l'objectif. Les lentilles se caractérisent essentiellement par :

A.2.1. La qualité du matériau

On trouve les lentilles en plastique et les lentilles en verre. Les lentilles en verre présentent de bien meilleures qualités optiques (transmission, reflet, déviation,...). Leur prix est par contre nettement supérieur.

La plupart des objectifs sont réalisés à partir de plusieurs lentilles. Le « signal » réfracté, de lentille à lentille, subit aussi des réflexions partielles ainsi que des aberrations sphériques et chromatiques. Pour cette raison, certaines lentilles sont traitées pour l'antireflet et les combinaisons de dioptries parfois complexes destinées à éviter les aberrations nécessitent un matériau très transparent pour éviter une trop grande perte de luminosité. Bien entendu, le prix est (c'est le cas de la dire) à l'image des performances obtenues !

A.2.2. La réponse spectrale

Celle-ci définit le comportement de l'objectif en fonction de la longueur d'onde de la lumière le traversant. Cette caractéristique est importante dès que l'information « couleur » est utilisée.

A.2.3. La distorsion

La distorsion est responsable de la déformation de l'image. Elle induit des non-linéarités dans l'opération de projection. Ces distorsions sont en général d'autant plus présentes que la focale utilisée est courte.

ANNEXE B

B. Le transfert des images

On distingue entre deux types de transfert des images, l'un est analogique et fait appel aux normes CCIR (norme européenne) ou RS-170 (norme nord-américaine) du monde de la télévision, tandis que l'autre est numérique.

Nous nous intéressons plutôt aux transferts numériques, puisque ceux-ci sont mis en œuvre par un grand nombre de caméras grand public et professionnelles, et ceci à des prix raisonnables (du moins dans le premier cas cité...). C'est pour cette raison que le transfert numérique a eu notre préférence.

B.1. Le transfert numérique

Le transfert numérique des images s'appuie sur deux normes. La première définit la couche physique et la seconde définit quant à elle le protocole.

La technique de développement de ces modes de transfert repose en effet sur la philosophie appliquée aux réseaux informatiques. Cette dernière consiste à décomposer la communication en plusieurs couches, et à définir une norme pour chacune d'entre elles. Le grand avantage d'une telle approche est de pouvoir effectuer des changements dans une couche, sans pour autant remettre en question la compatibilité avec les couches supérieures et inférieures.

A l'heure actuelle, la technologie employée est limitée et répond difficilement aux attentes de certaines applications, qui exigent le transfert d'images de taille importante, codées sur 8 ou 16 bits par pixel à des cadences d'acquisition très élevées (supérieures 50 Hz). L'ordre de grandeur du flux de données nécessaire est alors de 100 Méga Octets par seconde.

Dans ce qui suit, nous allons décrire les normes RS-422, RS-644, IEEE1394 et USB, qui définissent les couches physiques de la transmission, puis la norme CCIR-601, qui définit, quant à elle la couche « protocole de communication ».

B.1.1. La norme RS – 422 ou EIA – 422

C'est une norme électrique de communication de données au-delà de 15 mètres (Les ports séries du Macintosh sont des ports RS-422). Elle a été conçue pour des transferts point à point, avec éventuellement plusieurs points d'arrivée. La norme EIA – 422 permet d'avoir des

connexions plus longues et des débits plus importants que la norme EIA – 232 (Les ports séries du PC standard). Pour atteindre ces performances, les signaux sont transmis en différentiel, pour une meilleure immunité aux bruits.

Afin de transmettre les images de la caméra à la carte d'acquisition vidéo, on utilise une connexion composée de plusieurs liaisons EIA – 422 élémentaires. Une partie des ces liaisons transmet des pixels. En général, il y a autant de liaisons que de bits utilisés pour la quantification. Par exemple, 8 liaisons sont utilisées pour des pixels codés sur 8 bits. Une autre partie des liaisons transmet les signaux de contrôle (horloge, synchronisation ligne et image,...).

B.1.2. La norme RS – 644 ou EIA – 644

La norme EIA – 644 appartient à la famille des LVDS (Low Voltage Differential Signal). Cette norme permet un débit 3 à 9 fois supérieur à la norme EIA – 422.

Les normes EIA – 644 et EIA – 422 présentent une certaine compatibilité entre elles, puisqu'un récepteur EIA – 644 peut recevoir un émetteur EIA – 422.

B.1.3. Le standard IEEE – 1394

La norme IEEE – 1394 a été conçue par « Apple Computer Inc » et développée dans le groupe de travail IEEE 1394. Il s'agit d'une liaison série dont le débit peut atteindre 3200 Mbits/s dans sa dernière version (FireWire b).

Le IEEE – 1394 a été développé pour améliorer les performances des réseaux LAN, en permettant de garantir une bande passante importante à des coûts réduits. Il vise également à remplacer les liaisons parallèles type SCSI, par de meilleurs débits et une plus grande souplesse d'utilisation (au niveau connectiques et longueur des liaisons). Cette norme permet également le *plug-and-play*, ce qui lui procure une très grande flexibilité d'utilisation.

ANNEXE C

C Le modèle de la distorsion des lentilles

Les caméras exhibent généralement des distorsions dues à leurs optiques, la principale d'entre elles étant la distorsion radiale. Les phénomènes de distorsion peuvent être modélisés par :

- - Deux coefficients de distorsion radiale r_1 et r_2
- - Deux coefficients de distorsion tangentielle t_1 et t_2

Soit (u, v) les coordonnées image issues d'une projection idéale, et (\tilde{u}, \tilde{v}) les coordonnées image du même point, issues d'une projection réelle (coordonnées assujetties à la distorsion). De la même manière, (x, y) représentent les coordonnées idéales dans le repère caméra et (\tilde{x}, \tilde{y}) les coordonnées assujetties à la distorsion dans ce même repère caméra. La distorsion subie s'exprime donc par :

$$\begin{cases} \tilde{x} = x + x \cdot \underbrace{[r_1 \cdot k^2 + r_2 \cdot k^4]}_{\text{la distorsion radiale}} + \underbrace{[2 \cdot t_1 \cdot x \cdot y + t_2 \cdot (k^2 + 2 \cdot x^2)]}_{\text{la distorsion tangentielle}} \\ \tilde{y} = y + y \cdot \underbrace{[r_1 \cdot k^2 + r_2 \cdot k^4]}_{\text{la distorsion radiale}} + \underbrace{[2 \cdot t_1 \cdot x \cdot y + t_2 \cdot (k^2 + 2 \cdot y^2)]}_{\text{la distorsion tangentielle}} \end{cases} \quad (\text{C.1})$$

Avec : $k^2 = x^2 + y^2$

Le centre de la distorsion radiale est le point principal. Cela est dû au fait que :

$$\begin{cases} x = (u - u_0) / \alpha_u \\ y = (v - v_0) / \alpha_v \end{cases} \quad (\text{C.2}) \quad \text{et} \quad \begin{cases} \tilde{x} = (\tilde{u} - u_0) / \alpha_u \\ \tilde{y} = (\tilde{v} - v_0) / \alpha_v \end{cases} \quad (\text{C.3})$$

Avec : u_0, v_0, α_u et α_v sont les paramètres intrinsèques de la caméra.

Ainsi le système d'équations (C.1) peut s'écrire sous la forme suivante :

$$\begin{cases} \tilde{u} = u + (u - u_0) \cdot \left[r_1 \cdot k^2 + r_2 \cdot k^4 + 2 \cdot t_1 \cdot y + t_2 \cdot \left(\frac{k^2}{x} + 2 \cdot x \right) \right] \\ \tilde{v} = u + (v - v_0) \cdot \left[r_1 \cdot k^2 + r_2 \cdot k^4 + 2 \cdot t_2 \cdot x + t_1 \cdot \left(\frac{k^2}{y} + 2 \cdot y \right) \right] \end{cases} \quad (\text{C.4})$$

ANNEXE D

D.1. Le détecteur de Harris

La première étape appliquée par l'algorithme consiste à calculer les dérivées de l'intensité pixel $I(u, v)$ suivant u et v dans l'image. A l'issue de cette opération, pour chaque pixel d'abscisse u et d'ordonnée v , une matrice $A(u, v)$ est calculée suivant.

$$A(u, v) = \begin{bmatrix} \left(\frac{\partial I(u, v)}{\partial u} \right)^2 & \frac{\partial I(u, v)}{\partial u} \frac{\partial I(u, v)}{\partial v} \\ \frac{\partial I(u, v)}{\partial u} \frac{\partial I(u, v)}{\partial v} & \left(\frac{\partial I(u, v)}{\partial v} \right)^2 \end{bmatrix} \quad (\text{D.1})$$

Les valeurs propres sont calculées très classiquement en résolvant l'équation: $\text{Det}(A - \lambda I_2) = 0$, où λ représente les valeurs propres et I_2 est la matrice identité. S'agissant d'une équation du second degré, les solutions recherchées s'expriment simplement par :

$$\lambda = \frac{1}{2} \left[\text{Trace}(A) \pm \sqrt{\text{Trace}^2(A) - 4\text{Det}(A)} \right] \quad (\text{D.2})$$

Les quantités $\text{Det}(A)$ et $\text{Trace}(A)$ représentant respectivement le déterminant et la trace de la matrice A .

En terme d'interprétation, si les deux valeurs propres (λ_1, λ_2) de la matrice A sont élevées, cela signifie qu'un petit déplacement dans n'importe quelle direction à partir du pixel considéré engendrera un changement significatif du niveau de gris, ce qui se passe exactement dans le cas des coins. Par conséquent, en analysant les valeurs propres minimums calculées pour chaque pixel, on peut en déduire si ce dernier constitue effectivement un « coin ». Cette extraction de la valeur minimale des deux valeurs propres calculées s'exprime par :

$$f(x, y) = \min(\lambda_1(u, v), \lambda_2(u, v)) \quad (\text{D.3})$$

Etant donné un seuil ζ (calculé généralement à partir de l'histogramme des valeurs propres minimums pour chacun des pixels de l'image), si $f(u, v) > \zeta$, on peut conclure que le pixel de coordonnées $(u, v)^T$ est vraisemblablement un coin. Pour éviter de détecter plusieurs coins liés au coin réel, un candidat n'est considéré comme tel que s'il est suffisamment loin des N coins déjà détectés, ce qui se traduit par l'inégalité suivante :

$$\forall i < N, \|C_i(x, y) - C_N(x, y)\| \geq D \quad (D.4)$$

D étant la distance euclidienne minimum qui doit séparer n'importe quelle paire de coins.

La figure ci-dessous illustre le résultat obtenu après application de la procédure décrite plus haut.



Figure III.5 : Résultat de la procédure de détection des coins (image de Marc Pollefeys).

ANNEXE E

E.1. Le calcul de la matrice fondamentale

Supposons que nous ayons déjà trouvé un certain nombre de correspondances $p \leftrightarrow p'$ entre les deux images. Chaque mise en correspondance fournit une contrainte linéaire sur les coefficients de F :

$$uu' f_{11} + uv' f_{12} + uf_{13} + vu' f_{21} + vv' f_{22} + vf_{23} + u' f_{31} + v' f_{32} + f_{33} = 0 \quad (\text{E.1})$$

Les coordonnées des points p et p' sont notées ici respectivement $(u \ v \ 1)^T$ et $(u' \ v' \ 1)^T$. La combinaison des équations obtenues à partir de chaque mise en correspondance, aboutit à un système linéaire de la forme $Af = 0$, où f est un vecteur contenant les neuf coefficients de la matrice F . Chaque ligne de la matrice A est ainsi construite à partir des coordonnées image des points p et p' d'un seul appariement. Puisque F est seulement définie à un facteur d'échelle près, on peut contraindre la solution à avoir une norme unitaire. Généralement, on a toujours plus de huit paires de correspondance, et on utilise la résolution au sens des moindres carrés pour trouver une solution :

$$\min_{\|f\|=1} \|Af\|^2 \quad (\text{E.2})$$

$$\text{avec, } \|Af\|^2 = f^T A^T A f$$

Ceci revient à trouver le vecteur propre associé à la plus petite valeur propre de la matrice symétrique positive $A^T A$. Plusieurs techniques numériques standards sont envisageables. Cependant, cette formulation n'impose pas la contrainte du rang, ainsi une deuxième étape est ajoutée au calcul pour projeter la solution F sur le sous-espace de rang 2, en effectuant la décomposition en valeurs singulières de F et en égalisant la plus petite valeur singulière à zéro. La décomposition en valeurs singulières, décompose la matrice F comme suit :

$$F = Q D R \quad (\text{E.3})$$

Avec, D une matrice diagonale, et Q et R sont orthogonales. L'affectation de zéro à la plus petite valeur diagonale de D conduit au résultat souhaité.

La méthode décrite ci-dessus est « générique » et son application sans tenir compte des considérations numériques conduit à des résultats instables. Ainsi, la matrice $A^T A$ construite à partir de la matrice A est mal conditionnée³⁶. Une solution simple à ce problème consiste à normaliser les coordonnées des pixels (par exemple passer de $[0, 640]$ à $[-1, 1]$), avant d'appliquer la démarche décrite précédemment. Ceci a pour conséquence de produire une matrice A bien conditionnée. En pratique, un effort considérable doit également être entrepris pour rejeter les faux appariements des données d'entrée ou se montrer numériquement plus « robuste » vis-à-vis de ces derniers. Généralement, les méthodes basées sur les moindres des médianes des carrées produisent de bons résultats.

³⁶ Le nombre de conditionnement est très grand.

ANNEXE F

F.1. Détail de l'extraction des paramètres intrinsèques de la caméra à partir de la matrice fondamentale

L'équation de Longuet-Higgins (F.1), est applicable quand les coordonnées sont normalisées (sans dimension) afin de calibrer la ou les caméras. Si le déplacement entre les deux caméras (ou les deux positions consécutives de la caméra) est donné par une rotation R et une translation T , et si m est le correspondant de m' , alors la contrainte de coplanarité reliant les droites $(C'm')$ et (Cm) est la suivante :

$$m'^T (T \times R)m = m'^T E m = 0 \quad (\text{F.1})$$

La matrice E , définie par le produit d'une matrice orthogonale (la matrice de rotation) par une matrice antisymétrique, est appelée la *matrice essentielle*. Elle dépend uniquement de cinq paramètres. La relation entre les coordonnées caméras et les coordonnées images est donnée par :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ 1 \end{pmatrix} \quad (\text{F.2})$$

Ainsi, on aura : $p = C \cdot m$ et $p' = C' \cdot m'$

On rappelle que les coefficients des matrices C et C' sont les paramètres intrinsèques des deux caméras. Dans le contexte de l'auto calibration d'une unique caméra se déplaçant dans une scène rigide³⁷, il est légitime de les supposer dorénavant égales. En remplaçant dans l'équation (3.31), on obtient :

$$p'^T (C^{-1})^T E (C^{-1}) p = 0 \quad (\text{F.3})$$

De ce qui précède, on peut en déduire que :

$$F = (C^{-1})^T E (C^{-1}) \quad (\text{F.4})$$

³⁷ Cette remarque vaudrait aussi dans le cas d'un système stéréoscopique constitué de deux caméras parfaitement identiques.

La matrice essentielle E dépend seulement de cinq paramètres indépendants. Elle est contrainte par deux équations polynomiales, en plus de $\text{Det}(E) = 0$. Si F est connue, il découle de $E = C^T F C$ que les coefficients de la matrice C sont assujettis aux deux équations polynomiales héritées de E . L'auto calibration consiste alors à utiliser ces contraintes afin d'obtenir les paramètres intrinsèques à partir de la matrice fondamentale. Les équations de Kruppa, détaillées dans ce qui suit, constituent alors un outil efficace.

F.2 Extraction des équations de Kruppa à partir de la transformation épipolaire

Les contraintes mentionnées précédemment expriment le fait que le mouvement de la caméra présente nécessairement un déplacement rigide. Plusieurs formulations sont possibles. La formulation la plus intéressante est l'utilisation des équations de Kruppa, exploitées la première fois dans le domaine de la vision par ordinateur par Faugeras et Maybank afin d'étudier le mouvement, puis ensuite pour développer une théorie d'auto calibration. Les équations de Kruppa sont obtenues à partir d'une interprétation géométrique des contraintes de rigidité de la scène. Cette interprétation se rapporte aux tangentes à l'image ω du cône absolu³⁸ sur les deux vues prises par la même caméra. Les deux images sont liées l'une à l'autre par une transformation épipolaire. La matrice de l'enveloppe du cône (« dual conic », voir annexe D) de ω est la matrice enveloppe³⁹ (« dual ») $K = B^*$, avec B la matrice de ω :

$$B = (C^T)^{-1} C^{-1} \quad (\text{F.5})$$

La ligne épipolaire $l = \langle e, p \rangle$ est tangentielle à ω , si et seulement si :

$$(e \times p)^T K (e \times p) = 0 \quad (\text{F.6})$$

qui est l'équation définissant l'enveloppe du cône absolu. Ici, $K = C \cdot C^T$, e est l'épipole et p est un point appartenant à la ligne épipolaire tangentielle à l'image du cône absolu.

Soit F la matrice fondamentale du système de stéréovision. Effectuons un changement de coordonnées image (dans chaque image) de telle manière que les épipoles se situent aux centres des images et que les lignes épipolaires aient les mêmes ordonnées dans les deux images. De cette façon, la dernière ligne et la dernière colonne de la matrice fondamentale

³⁸ Le cône absolu est un cercle imaginaire de rayon $i = \sqrt{-1}$ dans le plan à l'infini.

³⁹ **Matrice dual**: matrice de cofacteurs.

disparaissent et la matrice fondamentale F' issue de ces transformations aura la forme suivante :

$$F' = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (\text{F.7})$$

On nomme T et T' les matrices 3×3 associées aux transformations appliquées respectivement à la première et à la deuxième image. Les matrices des paramètres intrinsèques C et C' deviennent alors respectivement $T \cdot C$ et $T' \cdot C'$ tandis que les matrices de projection s'écrivent maintenant $T \cdot P$ et $T' \cdot P'$.

Considérons alors un plan passant par les centres de projection des deux caméras et tangentiel au cône absolu. Ce plan est projeté en deux lignes épipolaires, qui sont tangentielles par construction à l'image du cône absolu, dont on rappelle que celle-ci est invariante par rapport à la position de la caméra (En réalité, il y a deux paires de lignes épipolaires qui satisfont cette condition, comme le montre la Figure F.1).

On rappelle que les lignes épipolaires sur les deux images ont des équations identiques suite aux transformations. Ainsi, si $(\lambda, \mu, 0)^T$ est la ligne épipolaire tangentielle à l'image du cône absolu et si D et D' sont les enveloppes des images du cône absolu dans chacune des deux images, il résulte de ce qui précède que :

$$(\lambda, \mu, 0) D (\lambda, \mu, 0)^T = 0 \quad (\text{F.8-a})$$

et

$$(\lambda, \mu, 0) D' (\lambda, \mu, 0)^T = 0 \quad (\text{F.8-b})$$

En écrivant ces équations explicitement, nous obtenons le système suivant :

$$\begin{cases} \lambda^2 d_{11} + 2\lambda\mu d_{12} + \mu^2 d_{22} = 0 \\ \lambda^2 d'_{11} + 2\lambda\mu d'_{12} + \mu^2 d'_{22} = 0 \end{cases} \quad (\text{F.9})$$

Avec $D = [d_{ij}]$ et $D' = [d'_{ij}]$ (les deux matrices sont symétriques).

Puisque les deux tangentes aux images du cône absolu doivent avoir les mêmes équations pour l'une et l'autre des deux images, celles-ci doivent avoir les mêmes solutions pour λ et μ . Cela signifie qu'à un facteur d'échelle près, les deux équations composant le système (F.9) sont les mêmes, ce que nous pouvons traduire par :

$$\frac{d_{11}}{d'_{11}} = \frac{d_{12}}{d'_{12}} = \frac{d_{22}}{d'_{22}} \quad (\text{F.10})$$

Nous allons pouvoir exprimer les coefficients d_{ij} et d'_{ij} en fonction de l'enveloppe des images du cône absolu, K , et des deux transformations T et T' . En effet, étant donnée la transformation T , un point m appartient à une ligne l si et seulement si :

$$l^T \cdot m = 0 \quad (\text{F.11})$$

Ceci qui peut encore s'écrire :
$$l^T \cdot T^{-1} \cdot T \cdot m = 0 \quad (\text{F.12})$$

Cette dernière équation peut s'interpréter en disant que m appartient à l , si et seulement si $T \cdot m$ appartient à $(T^{-1})^T \cdot l$, c'est à dire la transformée de la ligne l .

Par ailleurs, une ligne l appartient à l'enveloppe du cône K si et seulement si :

$$l^T \cdot K \cdot l = 0 \quad (\text{F.13})$$

Ou encore :

$$(l^T \cdot T^{-1}) \cdot (T \cdot K \cdot T^T) \cdot (T^{-T} \cdot l) = 0 \quad (\text{F.14})$$

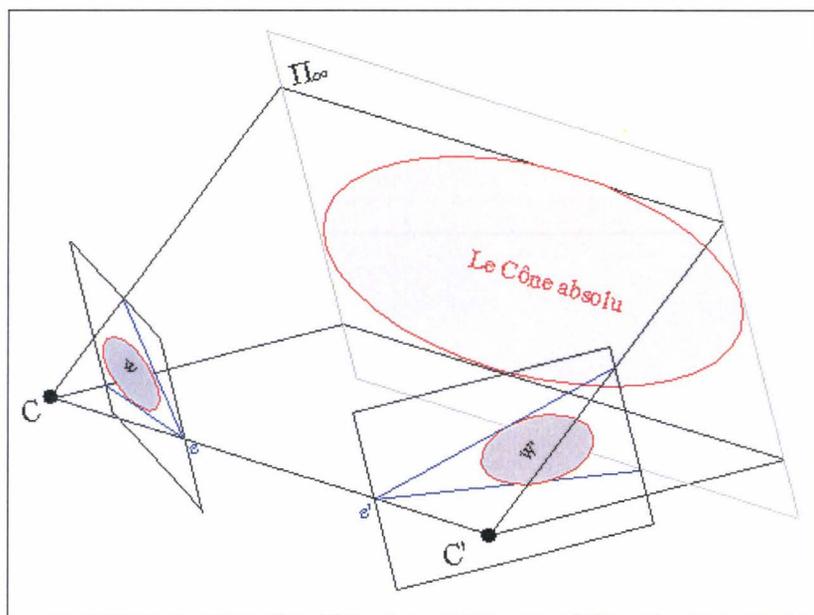


Figure F.1 : Les propriétés du cône absolu et ses images sur les plans des caméras.

Ainsi, la transformation T transforme l'enveloppe du cône K en l'enveloppe $D = T \cdot K \cdot T^T$. De même, un raisonnement parfaitement analogue permettrait de

montrer que T' transforme l'enveloppe de ce même cône K en l'enveloppe $D' = T' \cdot K' \cdot T'^T$. Si nous notons par ailleurs :

$$T = \begin{bmatrix} t_1^T \\ t_2^T \\ t_3^T \end{bmatrix} \quad (\text{F.15})$$

(t_i est la $i^{\text{ème}}$ ligne de la matrice T)

nous aurons alors $d_{ij} = t_i^T K t_j$ et $d'_{ij} = t_i'^T K t_j'$, ce qui permet d'exprimer explicitement (F.10) sous la forme :

$$\frac{t_1^T K t_1}{t_1^T K t_1'} = \frac{t_1^T K t_2}{t_1^T K t_2'} = \frac{t_2^T K t_2}{t_2^T K t_2'} \quad (\text{F.16})$$

qui représente les équations de Kruppa.

Dans ce qui suit, nous allons procéder à l'extraction des matrices de transformations T et T' . Pour y parvenir nous utilisons la décomposition en valeurs singulières de la matrice fondamentale. La matrice F s'écrit alors $F = UDV^T$, avec U et V deux matrices orthogonales et $D = \text{diag}(r, s, 0)$ une matrice diagonale. Ainsi, nous aurons:

$$F = U \underbrace{\begin{bmatrix} r & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_D V^T \quad (\text{F.17})$$

Posons:

$$T' = \begin{bmatrix} r & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{bmatrix} U^T \quad \text{et} \quad T = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} V^T \quad (\text{F.18})$$

Compte-tenu de la notation introduite ci-avant, nous avons bien $F = T'^T \cdot F' \cdot T$ avec T et T' non-singulières. Pour des paires de points mis en correspondance (ce qui est noté ici $p' \leftrightarrow p$), en remplaçant F dans (3.27) il vient :

$$p'^T \cdot T'^T \cdot F' \cdot T \cdot p = 0 \quad (\text{F.19})$$

Posons $\hat{p} = T \cdot p$ et $\hat{p}' = T' \cdot p'$. Il résulte directement de (F.19) que :

$$\hat{p}'^T \cdot F' \cdot \hat{p} = 0 \quad (\text{F.20})$$

Ainsi, T et T' représentent deux transformations satisfaisant l'ensemble des conditions requises. Nous pouvons donc exprimer les matrices T et T' de la façon suivante:

$$T' = \begin{bmatrix} t_1'^T \\ t_2'^T \\ t_3'^T \end{bmatrix} = \begin{bmatrix} ru_1^T \\ su_2^T \\ u_3^T \end{bmatrix} \quad T = \begin{bmatrix} t_1^T \\ t_2^T \\ t_3^T \end{bmatrix} = \begin{bmatrix} v_2^T \\ -v_1^T \\ v_3^T \end{bmatrix}$$

Avec u_i et v_i respectivement les $i^{\text{ème}}$ colonnes de U et V .

De (F.16) on obtient :

$$\frac{v_2^T K v_2}{r^2 u_1^T K u_1} = \frac{-v_2^T K v_1}{rs u_1^T K u_2} = \frac{v_1^T K v_1}{s^2 u_2^T K u_2} \quad (\text{F.21})$$

Ce qui représente les équations explicites de Kruppa.

Puisque les numérateurs et les dénominateurs des équations (F.21) sont des expressions linéaires en fonction des coefficients de la matrice K , les équations résultantes seront quadratiques. La matrice K étant symétrique et définie à un facteur d'échelle près, celle-ci dispose seulement de cinq degrés de liberté. A partir de trois vues différentes, nous aurons trois matrices fondamentales (une par paire d'images). A l'aide de celles-ci, nous pourrions construire un système de six équations quadratiques indépendantes, suffisantes pour déterminer la matrice K .

La détermination des paramètres intrinsèques de la caméra à partir de la résolution de ces équations est possible, bien que difficile (difficulté inhérente aux méthodes de résolution de systèmes non-linéaires redondants). Ce problème peut-être simplifié en supposant le point principal de la caméra connu. Ainsi, en effectuant le changement de repère adéquat, il est possible de ramener ce dernier à l'origine. La matrice des paramètres intrinsèques de la caméra devient alors une matrice diagonale $C = \text{diag}(\alpha_u, \alpha_v, 1)$. Dans ce cas, $K = C \cdot C^T$ et les équations de Kruppa (F.21) sont quadratiques par rapport à α_u^2 et α_v^2 . A partir de deux images, nous obtenons deux équations quadratiques sur ces deux variables, suffisantes pour résoudre le système correspondant en α_u^2 et α_v^2 . Des quatre solutions résultantes pour α_u et α_v , seules les solutions positives seront conservées (étant les seules à avoir une signification par rapport à la définition même de α_u et α_v).

ANNEXE G

G.1. Dimensions des plaques

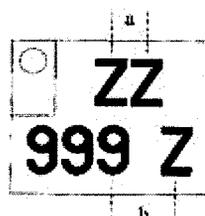
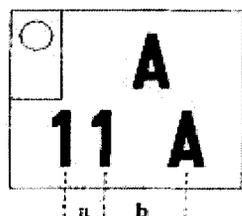
En France, la plaque arrière doit être jaune réfléchissant avec des inscriptions noires et mesurer soit 110 mm x 520 mm, soit 275 mm x 200 mm. La plaque avant doit être blanc réfléchissant avec des inscriptions noires et une taille soit de 110 mm x 520 mm, soit de 100 mm x 455 mm. La police de caractères est aussi homologuée et doit être totalement lisible.

Le numéro d'immatriculation des cyclomoteurs doit être reproduit sur la plaque arrière en caractères noirs non rétro réfléchissants sur fond rétro réfléchissant blanc.

Extrait du Journal Officiel du 27 mars 2004 NOR : EQUS0301814A	COTES en millimètres					
	Cas général			Motocyclettes		Cyclomoteurs
						non carrossés
						à 2 ou 3 roues
	AV	AV & AR	AR	AR		AR
sur 1 ligne	sur 1 ligne	sur 2 lignes	sur 1 ligne	sur 2 lignes	sur 2 lignes	
Hauteur de la plaque (sans bavette)	100 ± 2	110 ± 2	200 ± 2	75 ± 2	130 ± 2	120 ± 2
Hauteur utile de la plaque	90 ± 2	100 ± 2	190 ± 2	65 ± 2	120 ± 2	110 ± 2
Longueur de la plaque (hors tout)	455 ± 2	520 ± 2	275 ± 2	275 ± 2	210 ± 2	140 ± 2
Longueur utile de la plaque	445 ± 2	510 ± 2	265 ± 2	265 ± 2	200 ± 2	130 ± 2
Rayon de raccordement	6 à 12			6 à 10		4 à 6
Hauteur (*) des chiffres et des lettres	70 à 80			42 à 48		42 à 46
Largeur (*) des chiffres et des lettres	32 à 46			19 à 27		19 à 26
(autres que W et M)	40 à 54			24 à 32		20 à 28
W et M						

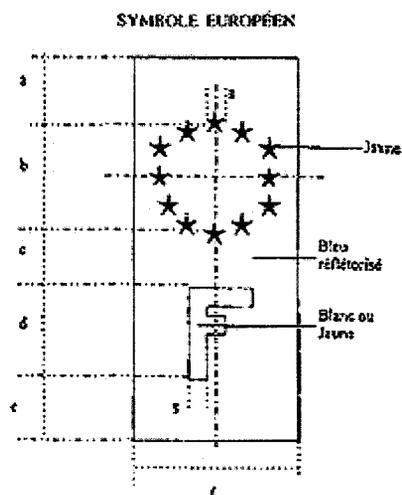
Entraxe entre les chiffres ou entre les lettres (A) (recommandé) (1)	Minimum 45			Minimum 27	Minimum 20
Entraxe entre les blocs (B) recommandé (1)	Minimum 65			Minimum 37	Minimum 25
Largeur (*) du trait	10 à 12			6 à 8	4 à 7
Hauteur du symbole européen	90 ± 1	100 ± 1	90 ± 1	65 ± 1	55 ± 1
Largeur du symbole européen	40 ± 1	45 ± 1	40 ± 1	30 ± 1	26 ± 1
Espacement entre le symbole européen et le bord utile de la plaque	0			0	0
Espacement entre le symbole européen et le premier chiffre	minimum 5			minimum 5	minimum 10
Espacement entre le bord utile de la plaque et le caractère	minimum 5			minimum 5	minimum 4,5
Dimensions des caractères de la date de validité des plaques en série Transit Temporaire					
Entraxe entre le dernier symbole et les chiffres de la date de validité (C)	minimum 45				
Hauteur (*) des chiffres	28 à 32				
Largeur (*) des chiffres	13 à 17				
Largeur (*) du trait	4 à 6				
Entraxe entre les chiffres	minimum 20				
Espacement horizontal entre le bord utile de la plaque et le caractère	minimum 5				

Espacement entre le bord des caractères et le bord supérieur et inférieur de la surface utile de la plaque	égaux et supérieurs à 5	égaux et supérieurs à 5	égaux et supérieurs à 4,5
<p>(A), (B), (C) : voir http://www.code-route.com/plaques_mineralogiques.pdf</p> <p>(*) : Les dimensions (hauteur, largeur, épaisseur) des différents caractères d'une même plaque doivent être uniformes.</p> <p>(1) : Pour permettre l'introduction du symbole européen sur la plaque arrière des véhicules à deux roues comportant les combinaisons MM ou MW, il est possible de descendre au-dessous des valeurs minimales indiquées.</p> <p><i>Nota :</i></p> <p>Les plaques AR d'une remorque d'un PTAC inférieur à 500 kg, ou celles amovibles dont il est fait état à l'article 7 (télécharger ci-dessous), peuvent être de dimensions réduites (100 x 455). Les plaques AR des cyclomoteurs carrossés (voiturettes) ainsi que celles des quadricycles légers à moteur peuvent avoir soit les dimensions prévues pour le cas général, soit avoir des dimensions réduites à 100 x 455, soit avoir les dimensions prévues pour les motocyclettes.</p>			



Cyclomoteurs

Plaque 2 lignes, série normale
(avec symbole européen obligatoire)



Extrait du Journal Officiel du 27 mars 2004 NOR : EQUS0301814A	DIMENSIONS en millimètres			
	Cas général		Motocyclettes	Cyclomoteurs
				non carrossés
	AV et 1 ligne	AV 1 ligne (100 x 455)	AR 1 et 2 lignes	à 2 ou 3 roues
	(110 x 520)	AR 2 lignes (200 x 275)		AR 2 lignes
a	15	13,5	10	8,5
b	30	27	19,5	16,5
c	17	15,5	11	9
d	20	18	13	11
e	18	16	11,5	10
f	45	40	30	26
g	4 à 5	3,5 à 4,5	2,5 à 3,5	2 à 3

