

UNIVERSITÉ BLAISE PASCAL - CLERMONT II

ÉCOLE DOCTORALE
SCIENCES POUR L'INGÉNIEUR DE
CLERMONT-FERRAND

Formation Doctorale :

Composants et Systèmes
pour le Traitement de l'Information
Spécialité : Vision pour la Robotique

THÈSE INTITULÉE :

Conception d'une plate-forme
d'implémentation matérielle dédiée
aux systèmes de vision active basés
sur un imageur CMOS

Rédigé par :

Pierre Chalimbaud

Décembre 2004

Table des matières

Liste des figures	8
Introduction	13
1 Vision Active	15
1.1 Tâche visuelle et vision	16
1.2 Systèmes de vision et niveaux de représentation	19
1.3 Vision active	21
1.3.1 De la vision passive à la vision active	21
1.3.2 Fondements de la vision active	23
1.3.2.1 La vision active de Y. Aloimonos	23
1.3.2.2 La vision Animée de D. H. Ballard	23
1.3.2.3 Le perception active de R. Bajcsy	24
1.3.2.4 Le vision intentionnelle de Y. Aloimonos	25
1.4 Routines visuelles	26
1.4.1 Aspect conceptuel	26
1.4.2 Élaboration des routines visuelles	28
1.5 Systèmes de vision active	29
1.5.1 Têtes robotiques pour la vision active	31
1.5.2 Systèmes intégrés de vision active	33
1.5.3 Systèmes complets de vision active	35

1.5.3.1	Système MEDUSA	36
1.5.3.2	Architecture du système VAP(<i>Vision As Process</i>)	36
1.5.3.3	Système TEA-1	37
1.6	Motivations et objectifs	38
2	Les technologies de perception visuelle	41
2.1	Les technologies d'imagerie	44
2.1.1	La photodétection	45
2.1.1.1	La photodiode	46
2.1.1.2	Le détecteur photoMOS	46
2.1.2	L'imagerie CCD	48
2.1.3	L'imagerie CMOS	51
2.1.3.1	La structure CMOS PPS	52
2.1.3.2	La structure CMOS APS	53
2.1.3.3	Le pixel photodiode en mode courant	54
2.1.3.4	Le pixel photodiode en mode intégration	55
2.1.3.5	Le pixel photoMOS	56
2.1.4	Comparaison technologique	57
2.1.5	Conclusion sur le choix d'un imageur	61
2.2	Les rétines artificielles	62
2.2.1	Architectures des rétines artificielles	63
2.2.2	Classification des rétines	64
2.2.3	Avantages et inconvénients des rétines artificielles	67
2.3	Les caméras dites "intelligentes"	68
2.3.1	Organisation fonctionnelle	69
2.3.2	Les solutions dites "programmées"	71
2.3.2.1	Les microprocesseurs dédiés au traitement du signal(DSP)	72
2.3.2.2	Les machines parallèles	73

2.3.3	Les solutions dites "câblées"	74
2.3.3.1	Les architectures matérielles complètement dédiées	74
2.3.3.2	Les architectures matérielles programmables	75
2.3.3.3	Les solutions mixtes	75
2.3.4	Les méthodologies de conception	76
2.4	Choix technologiques	77
3	Conception d'une plate-forme de recherche	79
3.1	Structure matérielle	81
3.2	La carte de traitement principale	83
3.2.1	Fonctionnalités du circuit programmable	83
3.2.1.1	Structure Matérielle Configurable	84
3.2.1.2	Les blocs de mémorisation internes	86
3.2.1.3	Les fonctions arithmétiques câblées	87
3.2.1.4	Les processeurs paramétrables	88
3.2.2	Les modules de mémorisation externes	90
3.2.3	L'environnement d'implantation matérielle	91
3.3	La carte sensorielle	92
3.3.1	Le dispositif de mesures inertielles	93
3.3.2	Le dispositif d'imagerie	95
3.4	La carte de communication	97
3.4.1	La carte de communication PCI	98
3.4.2	La carte de communication USB2.0	100
3.5	La carte de traitement secondaire	101
3.5.1	Propriétés du calculateur	101
3.5.2	Cohabitation des calculateurs	102
3.6	Conclusion	103

4	Méthodologie d'implantation	105
4.1	Introduction	105
4.1.1	Rétroactions visuelles	106
4.1.1.1	Les paramètres sensoriels	106
4.1.1.2	Les paramètres algorithmiques	107
4.1.2	Une architecture modulaire supervisée	108
4.2	Structure fonctionnelle	109
4.2.1	Détecteur actif	111
4.2.1.1	Définition	112
4.2.1.2	Identification	113
4.2.2	Supervision	114
4.2.2.1	Préemption des procédés de détection	115
4.2.2.2	Interaction système	116
4.3	Squelette d'implantation	118
4.3.1	Les modules génériques	119
4.3.1.1	Le dispositif superviseur	119
4.3.1.2	Le module de communication système	120
4.3.1.2.1	Transfert de données entrantes	120
4.3.1.2.2	Transfert de données sortantes	121
4.3.1.3	Module de commande sensorielle	122
4.3.1.3.1	Module fenêtrage sous échantillonné:	122
4.3.1.3.2	Module de fenêtrage "évolué":	123
4.3.1.4	Module de correction du FPN	126
4.3.2	Espaces de caractérisation du squelette	127
4.3.2.1	Espace d'implantation matériel	127
4.3.2.2	Espace de développement embarqué	128
4.3.2.3	Espace de développement haut niveau	128

4.3.3	Fenêtrage pour implémenter un suivi	129
4.4	Perspectives	129
5	Exemples de détecteurs actifs	131
5.1	Détecteur actif anti-éblouissement	132
5.1.1	Analyse de la chaîne de transduction	133
5.1.2	Le procédé d'asservissement	137
5.1.2.1	Équations d'évolution des tensions analogiques	138
5.1.2.2	Équations d'évolution numériques	139
5.1.3	Correction du FPN	140
5.1.4	Résultats	141
5.2	Détecteur actif de focalisation	143
5.2.1	Description fonctionnelle globale	144
5.2.2	Algorithme d'estimation du déplacement	145
5.2.3	Module de mémorisation	146
5.2.4	Module d'estimation du déplacement	147
5.2.4.0.1	Accumulation des éléments de G et e	148
5.2.4.0.2	Accumulation du vecteur de translation	149
5.2.5	Conclusion	151
5.3	Conclusion	153

Table des figures

1.1	Description de l'œil artificiel de Léonard de Vinci et du cheminement des rayons.	16
1.2	Modèles des routines visuelles. Le traitement est décomposé en deux étapes principales : une ascendante consistant en la création de la représentation de base de bas niveau, la seconde, descendante, appliquant les procédures visuelles sur cette représentation pour extraire les propriétés spatiales nécessaires pour satisfaire les buts.	27
1.3	Exemples de dispositifs robotiques de contrôle du regard: a - la tête robotique PennEyes du laboratoire GRASP de Philadelphie, b - la tête robotique CeDAR du laboratoire RSL en Australie, c - la tête robotique trinoculaire CardEye du laboratoire CVIP de Louisville, d - la tête robotique de l'INRIA de Sophia-Antipolis , e - la tête robotique PANTHER-VZ du laboratoire japonais Shirai.	32
1.4	Plate-forme robotique "embarquable" de l'université d'Oxford.	33
1.5	Architecture du système VAP.	37
2.1	La délocalisation des traitements le long de la chaîne de perception.	42
2.2	Effet photoélectrique dans le Silicium.	45
2.3	Structure d'un pixel photodiode.	46
2.4	Structure de détection photoMOS.	47
2.5	Transfert de charges à partir d'un registre à décalage analogique(CCD).	48
2.6	Transfert de charges vers la capacité de lecture d'une matrice CCD.	49
2.7	Circuit de lecture par intégration.	49

2.8	Structure d'une matrice CCD "inteligne".	51
2.9	Structure d'un imageur CMOS PPS ou MOSXY.	53
2.10	Structure d'un imageur CMOS APS.	54
2.11	Pixel photodiode en mode courant.	54
2.12	Pixel photodiode en mode intégration.	56
2.13	Pixel photoMOS.	57
2.14	(a) - Bruit spatial fixe colonne important du à la présence d'amplificateurs colonnes (Fuga 1000 - Fill Factory), (b) - Image sans correction du bruit spatial fixe (NC1802 - Neuricam), (c) - Image après correction du bruit spatial fixe (NC1802 - Neuricam).	61
2.15	Le contrôle CMOS permet de définir un système de vision active qui exploite l'échantillonnage du signal image.	62
2.16	Délocalisation des traitements au sein du capteur d'image: les rétines artificielles.	65
2.17	Introduction d'éléments de mémorisation pour effectuer un traitement spatio-temporel: a - mémorisation au sein du pixel, b - mémorisation au sein d'une matrice indépendante.	66
2.18	Structure générique d'une caméra "intelligente".	69
3.1	Organisation fonctionnelle de la plate-forme.	79
3.2	Synoptique du système embarqué.	80
3.3	Plate-forme de recherche proposée pour l'expérimentation de système de vision active embarqué.	81
3.4	Structure fagot évolutive constituant le système embarqué.	82
3.5	Composants de la carte de traitement principale.	84
3.6	Organisation structurelle des FPGAs Stratix.	85
3.7	Structure fonctionnelle d'une cellule élémentaire configurable (LE).	86
3.8	Description fonctionnelle d'un microcontrôleur NIOS.	89
3.9	Les capteurs implantés sur la carte Sensorielle.	92
3.10	Dispositif de mesures inertielles.	94

3.11	Dynamique de conversion numérique paramétrable.	96
3.12	Système d'imagerie CMOS permettant une double rétroaction visuelle.	97
3.13	Lien de communication de la caméra "intelligente" avec un système d'exploitation hôte.	97
3.14	Système embarqué muni à gauche de sa carte de communication PCI et à droite de sa carte de communication USB2.0.	98
3.15	Architecture synoptique du microprocesseur de traitement du signal.	102
3.16	Utilisation du calculateur DSP comme co-processeur.	103
4.1	Échantillonnage et transduction électrique de l'éclairage.	106
4.2	Système de vision active bas niveau basé sur une architecture supervisée.	108
4.3	Schéma fonctionnel global.	110
4.4	Décomposition du mécanisme de perception selon la notion de routines visuelles.	111
4.5	Schéma synoptique d'un détecteur actif.	113
4.6	Séquencement des tâches de perception d'un système de vision active.	115
4.7	Synoptique d'implantation.	117
4.8	Architecture générique	118
4.9	Les modules constituant le système de perception sont interfacés comme des périphériques d'un processeur Nios qui assure la fonction de superviseur.	120
4.10	Architecture du module de communication système.	121
4.11	Architecture fonctionnelle du module de commande sensorielle.	122
4.12	Module Fenêtrage sous échantillonné	123
4.13	Tracé d'une droite discrète selon l'algorithme de Bresenham.	124
4.14	Architecture d'adressage d'une fenêtre "évoluée".	125
4.15	Module de correction du bruit spatiale fixe du capteur CMOS.	126
4.16	Suivi d'un motif de niveau de gris (32×32) suivant temps de 720 ms entre deux acquisitions consécutives.	130

5.1	Adaptation de la conversion du signal visuel de chaque fenêtrés d'intérêt.	133
5.2	Chaîne de transduction de l'imageur.	133
5.3	Caractéristique de l'amplificateur analogique de l'imageur.	134
5.4	Caractéristique du convertisseur analogique numérique.	136
5.5	Schéma d'asservissement de la chaîne de transduction de l'imageur.	137
5.6	Modèle de la disparité des pixels de l'imageur CMOS.	141
5.7	Module de correction du FPN suivant les variation de V_{bg}	141
5.8	Scène soumis à un fort éclairément indirect	142
5.9	Scène soumis à un fort éclairément direct	143
5.10	Architecture fonctionnelle du détecteur actif de focalisation.	144
5.11	Mécanisme d'échange permettant au calculateur de déplacement de travailler à sa propre fréquence de traitement.	147
5.12	Architecture fonctionnelle de l'estimateur de translation.	148
5.13	Architecture fonctionnelle de l'estimateur de translation.	149
5.14	Résultats de simulation de l'architecture de convolution.	150
5.15	Résultats de simulation de l'architecture du détecteur actif de focalisation.	152

Introduction

La perception visuelle est une source d'information majeure sur notre environnement. Cependant, la quantité d'information et la complexité de ce signal posent de réelles difficultés d'analyse et de compréhension. Afin d'aborder cette problématique difficile, le concept de vision active propose de ré-évaluer les buts de la vision par ordinateur. L'originalité de cette approche consiste à intégrer la physiologie de l'observateur et les objectifs visuels dans la stratégie de perception. La richesse du signal image nécessite l'implémentation de mécanismes de traitement du signal évolués qui exigent une puissance de calcul d'autant plus élevée que les systèmes de vision imposent de fortes contraintes temporelles. L'analyse des données suivant un processus intentionnel permet alors de focaliser les ressources disponibles sur l'information utile. Cette notion tend vers une l'adéquation du système de vision avec sa cible d'implantation suivant des objectifs spécifiques. Les caméras "intelligentes" constitue une solution intéressante pour mettre en œuvre ce type d'adéquation. En exploitant une architecture embarquée auprès de l'imageur, ces machines de perception visuelle profitent d'un lien privilégié avec l'information à traiter et le dispositif sensoriel. Ainsi, un tel niveau d'interaction capteur/système conforte le postulat de la vision active qui suggère d'intégrer le capteur dans le processus de perception.

La technologie d'imagerie CMOS paraît particulièrement adaptée pour implémenter ce type d'interaction. En effet certains imageurs CMOS offrent un mode d'accès au pixel qui permet un échantillonnage contrôlé de l'espace photosensible. Cette particularité autorise la mise place d'une rétroaction sur la commande d'adressage du capteur. Ainsi, l'échantillonnage spatio-temporel du signal image peut être intégré dans la stratégie de perception. Selon une approche de perception locale, les imageurs CMOS permettent d'accroître la réactivité du système en sélectionnant l'information utile directement sur l'imageur. Ces considérations préliminaires sont

à l'origine de cette thèse.

Afin d'unifier ces considérations, un système embarqué dédié à la vision active a été conçu. Il constitue une plate-forme de recherche versatile permettant l'implémentation de systèmes de vision bas niveaux basés sur le concept de vision active. La proximité du capteur et de l'architecture de traitement embarquée autorise un contrôle en temps réel du dispositif sensoriel pour définir des mécanismes de rétroaction efficaces.

Notre plate-forme de recherche constitue un outil d'expérimentation riche qui peut être exploité selon divers aspects de la problématique d'adéquation algorithmique et architecturale. Cet outil s'appuie sur la flexibilité de la technologie FPGA pour constituer une cible d'implantation configurable munie de toutes fonctionnalités nécessaires à l'implantation d'un système de vision. Les caractéristiques architecturales et sensorielles de la plate-forme ont été définies selon les contraintes d'une approche de la perception visuelle bas niveau. Cette approche se traduit par un squelette d'implantation sur la plate-forme qui constitue une structure générique d'interaction devant faciliter l'implantation de système de vision dédié.

Cette thèse débute par une présentation du concept de vision active. Les fondements de cette approche de la problématique vision sont rappelés et un état des systèmes existants est proposé. Dans le second chapitre, les technologies de la perceptions visuelles sont discutées. Les technologies d'imagerie et d'intégration sont étudiées afin de justifier les choix technologiques qui ont guider la conception de notre plate-forme de recherche. Le troisième chapitre est consacré à la description de la plate-forme de recherche proposée. Chaque élément est détaillés pour démontrer son adéquation avec l'approche de la vision active. Dans le quatrième chapitre une méthodologie d'implantation est proposée. Un squelette d'implantation générique est établis en adéquation avec l'architecture de la plate-forme de recherche. Enfin, le dernier chapitre propose deux utilisations de ce squelette pour exploiter les possibilités offertes par un imageur CMOS.

Chapitre 1

Vision Active

Peut-on raisonnablement parler de capteurs de vision sans aborder le cas particulier de la vision humaine? Bien que particulier, le capteur "œil" n'en est pas moins le modèle générique de nombreux systèmes de vision artificielle. Que ce soit les premières caméras cinématographiques des frères Lumière ou bien encore les récents appareils de photographie numérique, tous ces systèmes reposent sur un principe commun issu d'observations physiologiques.

Ainsi Léonard de Vinci est l'un des premiers à assimiler l'œil à une chambre noire (camera obscura) et à en procéder à des études poussées. Ce dernier ira même jusqu'à fabriquer un œil artificiel en verre et métal qu'il décrit dans le Codex D folio 3 verso de la manière suivante (fig.1.1):

"On fera une boule en verre d'un diamètre de cinq-huitième de brasse. Puis on la coupera d'un côté de manière à pouvoir y mettre le visage jusqu'aux oreilles. Ensuite on mettra au fond un fond de boîte d'un tiers de brasse ayant au milieu un trou quatre fois plus grand que la pupille de l'œil.

En outre, on établira une boule de verre mince de diamètre d'un sixième de brasse. Ceci fait qu'on remplisse le tout d'eau tiède et limpide. Mettez le visage dans cette eau et regardez dans la boule. Notez bien et vous verrez, cet instrument enverra les espèces à l'œil comme l'œil les transmet à la vertu visuelle."

Toutefois, bien que son système ne soit pas correct (l'ensemble de ce modèle d'œil constitue un système optique divergent et l'image reste droite), Léonard de

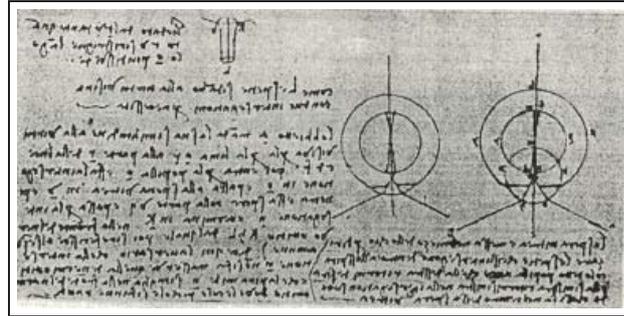


FIG. 1.1 – *Description de l'œil artificiel de Léonard de Vinci et du cheminement des rayons.*

Vinci fut le premier à décrire la "camera obscura". Cette dernière permit au 17^{ième} siècle à de nombreux peintres tels que Vermeer de tracer des paysages avant de les peindre.

Le principe de ces systèmes est aujourd'hui encore employé dans la plupart des capteurs de vision. En effet ces derniers se décomposent d'une partie optique (objectif réglant la focal, l'iris et la mise au point) et d'une partie sensible (pellicule, imageur, rétine artificielle,...). Cependant même si les technologies d'imagerie actuelle, sont capables de concevoir des cellules photosensibles bien plus performantes que celles utilisées par notre propre système de vision, la vision artificielle est bien loin d'égaliser la vision humaine.

En effet, les dispositifs artificiels sont capables de percevoir des nuances d'éclairément ou de couleurs extrêmement faibles. Ils sont aussi à même d'observer des phénomènes bien plus rapides que notre persistance rétinienne nous le permet. Malgré ces performances technologiques ne pouvant être égalées par l'œil humain, il ne nous faut que quelques centaines de millisecondes pour reconnaître un ami dans la rue alors qu'un système automatique... Voir n'est donc pas simplement recevoir des photons sur sa rétine!

1.1 Tâche visuelle et vision

En effet, le processus de vision se dessine de manière nettement plus complexe et fait à l'heure actuelle l'objet de nombreux travaux que ce soit en vision artifi-

cielle, cognition, biologie, médecine ou encore psychologie. Par conséquent, la vision ne peut pas être abordée comme un problème banal de traitement du signal. La difficulté majeure provient du fait que "voir" implique implicitement une part d'intelligence difficilement définissable. Cette forme d'intelligence apparaît comme une combinaison entre de nombreux processus de traitement d'information et des capacités à représenter l'information. Des tâches qui paraissent complexes telles que jouer aux échecs ou résoudre un problème de mathématique sont accessibles aux machines car les mécanismes de traitement et de représentation ont pu être identifiés. Par opposition, des tâches en apparence aisées telles que donner le contenu d'une boîte oblige en réalité à la mise en œuvre d'un grand nombre de processus complexes.

Toutefois sans rentrer dans les détails des différentes couches permettant à l'homme de percevoir de manière visuelle son environnement, il semble correcte de distinguer l'aspect visuel et l'aspect cognitif du processus de vision. Mais distinguer ces deux aspects ne signifie pas les séparer! Cela n'implique en aucun cas le fait de résumer la vision humaine à une caméra(œil) et un ordinateur (cerveau).

En effet, comme cela a été démontré dans de nombreux travaux, la vision humaine est dite **active** en ce sens qu'à partir d'un objectif, la chaîne œil/cerveau interagit de manière à optimiser sa perception. Ces interactions se font à tous les niveaux à commencer par le système oculaire.

Ainsi lorsque l'on passe d'un endroit très lumineux à l'obscurité, la pupille se dilate de manière à adapter le flux lumineux. Cependant le flux lumineux n'est pas le seul facteur commandant la pupille. En effet, des signaux du mécanisme de mise au point de l'œil affectent aussi la taille de la pupille. Pour exemple, la réalisation d'un travail à courte distance entraîne une légère fermeture de la pupille de manière à former une image plus nette. De plus, de récents travaux ont montré qu'il existait un facteur encore plus subtil agissant sur la pupille. En effet, l'émotion suscitée par ce que l'on regarde semble être un élément pris en compte pour le contrôle de la pupille. Ce paramètre était connu depuis plusieurs siècles des marchands de jade chinoise. Ces derniers présentaient aux acheteurs potentiels plusieurs pièces en observant leurs pupilles. Lorsque le marchand remarquait une légère dilatation de la pupille, il savait que le client avait envie de la pièce. Il en tenait alors compte pour fixer son prix!

A travers cet exemple, on perçoit aisément la notion de rétroaction dans le processus de la vision humaine. De nombreux travaux en sciences cognitives ont tenté de formaliser ces relations entre la perception et la tâche à réaliser. On notera les avancées significatives des universités de Reading (UK), Pasadena et Rochester (USA) qui sont parmi les plus avancées dans ces domaines et en particulier les récents travaux de Wilkie et Wann étudiant la relation entre le contrôle du mouvement des yeux et la tâche "locomotion" [1]. Dans leur article, ils abordent le problème du mouvement des yeux pendant une tâche de locomotion. En particulier l'article montre l'intérêt de fonctionner par tâche de fixation ponctuelle (gaze control) plutôt que par un "tracking" oculaire. Ces études concernant le rôle des mouvements oculaires en fonction des tâches visuelles ont aussi été largement abordées par Dana Ballard et Mary Hayboe. Ces dernières ont entre autres étudié la problématique de l'information visuelle dont nous avons besoin pour une tâche donnée. En d'autres termes, avons nous réellement besoin de tout ce que nous regardons pour réaliser une tâche donnée [2]? Leur expérimentation suggèrent que la vision humaine est fortement "dédiée à la tâche en cours". Ainsi, notre vision fonctionne par points de fixation sur lesquels l'analyse cérébrale est faite "just in time" de manière à résoudre la tâche visuelle. Cette dernière étude corrobore les travaux de Wilkie selon lesquels la vision fonctionne de manière ponctuelle de manière saccadée. La notion de vision locale, de focalisation et d'attention visuelle semble ainsi être au cœur du processus de vision. Cela a aussi été montré par M. Warren [3] qui s'est attaché à montrer les relations entre les mouvements de l'œil et le flot optique. Ce dernier pose le problème du découplage du mouvement perçu par le déplacement de l'œil et celui venant de la scène observée. Dans ses travaux, il montre que la perception du mouvement est seulement exécutée lors des phases de fixation.

De nombreux autres études portant sur ces problématiques liant la cognition, la psychologie et la biologie ont été proposées ces dernières années. Elles cautionnent naturellement le fait que la vision humaine n'est pas une simple chaîne perception-analyse-action. Elle montre aussi l'intérêt d'impliquer la tâche à réaliser au cœur du problème de perception.

Ces quelques remarques sur la physiologie de la vision, permettent d'aider à l'élaboration d'un système de vision active. En effet, comme cela a été dit plus haut, les tâches de focalisation d'intérêt sont au centre de la problématique. Toutefois

réaliser un système artificiel par "simple" copie du système humain est un challenge qui semble bien dérisoire pour une thèse... Pour cela, dans un second paragraphe, un bref état de l'art sur les différents modèles de représentations en vision est proposé. Il s'attache notamment à dégager les grands courants qui ont tenté de formaliser la vision artificielle ses dernières décennies et nous permettra d'introduire la notion de vision active artificielle.

1.2 Systèmes de vision et niveaux de représentation

Un système de vision par ordinateur comprend communément plusieurs niveaux de représentation de l'information visuelle allant classiquement du niveau le plus bas où l'image est décrite par le niveau de gris (ou de couleur) de chacun des pixels, au niveau plus élevé de représentation où le contenu de l'image est décrit par des symboles.

Durant ces trente dernières années, de très nombreuses tentatives ont été faites pour réaliser un système complet de vision. Aussi parmi tous les différents travaux proposés, nous pouvons mettre en évidence deux grandes tendances:

- Les premières tendances consistent à copier la vision humaine en utilisant par exemple des techniques neuromimétiques [4] ou encore en y incluant des simulations de "saccades oculaires" [5].

De manière plus traditionnelle, une des études les plus classiques en théorie de la vision a été proposée par David Marr à travers son désormais célèbre paradigme[6]. L'originalité de cette approche consiste à construire une description 3D de la scène sans incorporer d'informations sémantiques. Sans entrer en détails dans ce formalisme, l'approche proposée par Marr consiste à décomposer le problème de vision en différentes étapes conduisant à une décomposition en 3 structures de données:

- *Le croquis élémentaire* explicite les changements d'intensité lumineuse dans l'image, leur distribution et leur organisation géométriques,
- *Le croquis 2D 1/2* décrit la géométrie des surfaces visibles en termes d'orientations et d'informations de profondeur,
- *La description 3D* décrit les formes en termes de volumes d'occupation

Marr[6]	Demazeau [8]	Crowley[9]	Neumann[10]
Non défini	Scène	Interprétation	Interprétation de haut niveau
Description 3D	\simeq Objet	\simeq Groupement des caractéristiques	\simeq Objet reconnu
Croquis 2D1/2	\simeq Indices de scène	\simeq Description géométrique de la scène	\simeq Eléments de la scène
Croquis élémentaire	\simeq Indices d'image	\simeq Description géométrique de l'image	\simeq Eléments d'image
Images	\simeq Images	\simeq Images	\simeq Images

TAB. 1.1 – *Récapitulation par Demazeau de différents niveaux de représentation en vision par ordinateur.*

spatiales ainsi que leur organisation spatiale.

Depuis cette première ébauche de nombreux travaux ont été proposés. Demazeau en propose d'ailleurs une récapitulation dans le tableau suivant [7] :

Toutes ces approches conduisent plus ou moins aux mêmes type de représentation mettant en œuvre :

- une série de traitements extrayant des propriétés de l'image ou indices visuels,
- des modèles explicites de phénomènes visuels, du domaine d'application (vitesse, couleur, formes d'objet,...)
- des méthodes permettant une mise en correspondance entre ces modèles et indices.

Ainsi dans cette première catégorie d'approches, les connaissances *a priori* sur la scène n'existent pas et la détection d'indices visuelles de plus en plus riches se fait sans information sémantique. Ce sont des approches guidées par les données.

- Par opposition à ces premières approches, de nombreux autres travaux basés sur les architectures multi-agent ont vu le jour. Ces travaux ne sont pas inspirés de la vision humaine et ne visent pas une structure hiérarchique. Dans ce type de travaux, on cherche plutôt à faire collaborer des processus indépendants

par l'intermédiaire de zone d'échange. On citera par exemple un des systèmes les plus classiques VISIONS de Hanson et Riseman [11], qui utilise la notion de "tableau noir" pour réaliser l'échange d'information entre les processus. Ce système est capable de reconnaître des routes, des arbres et des maisons dans un environnement fortement contraint. Parmi ces approches peuvent aussi être citées les travaux de Clement et thonnat sur le système OCAPI[12] ou bien encore l'approche de baujard et Garbay avec une architecture multi-experte nommée MAPS [13]. Une des caractéristiques des ces approches est de pouvoir exploiter les connaissances de haut niveau pour guider la détection des indices visuelles, ce qui n'existent pas dans les approches traditionnelles proposées par Marr et al.

A travers ces deux approches, il est facile de déceler les avantages et les inconvénients de chacune d'entre elles. Un des inconvénients majeurs des premières approches est la hiérarchisation de l'information sans possibilité de retour du haut niveau sur le bas niveau. Dans le cas des méthodes multi-agent, dans la plupart des cas de fortes contraintes sur la scènes doivent être proposées mais le retour d'information permet de lever certaines ambiguïtés. Pour cela, depuis une quinzaine d'année, de nouvelles approches de la problématique "vision" sont apparues où l'on utilise la structure hiérarchique tout en autorisant le retour d'information du haut vers le bas niveau. Cette approche qui consiste à introduire l'action dans le mécanisme de perception est qualifié de vision active.

1.3 Vision active

1.3.1 De la vision passive à la vision active

A l'origine, était la vision dite passive. Cette dernière considérait une scène souvent statique observé par un capteur statique. Aux milieux des années 80, Aloimonos, puis Bajcsy et Ballard ont proposé de simplifier les problèmes de vision "classique" en y introduisant des contraintes. Pour cela, le concept de "vision active" a pris naissance par une tentative de simulation du système de vision biologique. Dans le cas de la vision humaine, l'adaptation de la pupille à l'illumination, le mouvement des yeux et de la tête font partie intégrante du processus de vision. De la même

manière, le but de la vision active est d'introduire ces facultés d'adaptation dans les mécanismes de vision artificielle.

La vision active souligne le fait que la perception est intimement liée à la physiologie de l'observateur et aux tâches qu'il doit réaliser. L'apparition de la vision active est due, d'une part à l'échec des approches traditionnelles pour résoudre un certains nombres de problèmes de vision difficiles, et d'autre part aux limites technologiques qu'elles ont rencontré. En effet, l'analyse visuelle statique est très rapidement restreinte par la puissance de calcul nécessaire à leur implémentation expérimentale. Ce constat d'échec à mener à ré-évaluer les buts de la vision par ordinateur elle-même afin de donner naissance à la vision active. Il est en effet désormais établi que même les systèmes de reconnaissance massivement parallèles ne peuvent assurer une analyse complète d'une image complexe du fait de l'explosion combinatoire. Il est alors vital de disposer d'un mécanisme de focalisation permettant de choisir quelle zone d'une image doit être analysée avant les autres parce qu'elle serait potentiellement plus intéressante.

Les théories traditionnelles de vision suggèrent que le rôle d'un système de vision consiste à reconstruire la carte de la surface visible, soit littéralement inverser le processus de création de l'image. Par opposition, un système de vision active est bien moins exhaustif. Il sélectionne les propriétés de l'image qui lui sont nécessaires pour réaliser la tâche qu'il lui est confié. La vision active est une **stratégie d'observation**. Plutôt que de traiter des images ou des séquences d'images individuellement, l'observateur et le capteur interagissent continuellement. Les données visuelles issues du capteur sont analysées intentionnellement pour dévoiler une information qui est immédiatement exploitée. L'approche active considère la vision comme processus actif à la fois dans l'espace et le temps. Cette théorie considère qu'il n'existe pas de système de vision universel. L'idée fondamentale réside dans la capacité du système de vision à adapter son état et son comportement d'une part à l'environnement observé et d'autre part à la tâche qu'il doit réaliser. Ainsi, le rôle d'un tel système ne se contente pas d'analyser son environnement visuel de manière statique. Il doit être capable de déduire un comportement à adopter, donc de définir une action suivant ses propres objectifs, afin améliorer sa perception. Ainsi, la vision active consiste à doter le système de vision d'un contrôle intelligent de l'ensemble des ressources matérielles et logicielles. Les objectifs de cette approche peuvent être d'améliorer

la qualité des résultats obtenus ou bien diminuer la complexité algorithmique et les quantités d'informations manipulées.

1.3.2 Fondements de la vision active

Le principe général du concept de vision active a abouti à des travaux de recherche hétéroclites suivant les centres d'intérêt des chercheurs. Différentes facettes du concept apparaissent dans les premiers travaux qui ont révélé l'intérêt d'une approche active pour aborder les problèmes de la vision artificielle. Les quatre grandes tendances sont ici récapitulées à titre de bibliographie.

1.3.2.1 La vision active de Y. Aloimonos

Les premières publications reconnues sur ce thème ont été proposées par Yannis Aloimonos [14]. Son objectif était d'aborder des problèmes fondamentaux de vision en considérant un observateur actif. L'idée consiste à contraindre le mouvement du capteur afin d'extraire des informations complémentaires qui permettent de résoudre plus facilement une tâche donnée. Il a démontré théoriquement l'intérêt d'un observateur actif pour divers problèmes tels que :

- La reconstruction de surfaces par "*shape form shading*", "*shape form texture*", "*shape form contour*",
- l'estimation de profondeur ("*depth computation*"),
- l'estimation de la structure d'objets 3D et du mouvement de la caméra à partir de primitives issues de leurs projections ("*structure from motion*").

Toutefois, ces travaux se focalisent avant tout sur l'aspect mathématique de la vision active et sur la résolution des problèmes de stabilité, linéarité et unicité des solutions proposées. Malheureusement peu de réalisations concrètes et de mises en œuvre ont été proposées par l'auteur.

1.3.2.2 La vision Animée de D. H. Ballard

Une seconde facette du concept de vision active a été proposée par Dana Ballard [15, 16] sous le nom d'*Animate vision* traduit par vision animée. Cette approche

consiste à concevoir le système de vision selon un modèle anthropomorphique. Ainsi, la vision animée considère un système binoculaire et fovéale. La rétine de l'œil humain présente une résolution décroissante entre son axe optique et sa périphérie. Aussi, afin de maximiser l'exploitation de la zone "fine" de la rétine, comme nous l'avons vu précédemment, le système de vision humain utilise un mécanisme rapide pour déplacer cette "région précise" par saccades sur les objets. C'est pour cette raison que la vision animée accorde une grande importance au contrôle rapide du regard. La vision animée exploite le contrôle du regard pour influencer la représentation du monde observé afin de faciliter l'extraction d'informations. Un objectif primordial de cette approche est d'utiliser les mouvements du capteur pour éviter le problème de calibration et rendre la représentation du monde invariant aux positions de l'objet et de l'observateur en abandonnant le repère centré caméra pour l'utilisation d'un repère de référence sur le point de fixation.

L'objectif de la vision animée est de modéliser la vision comme un comportement intelligent, en utilisant les théories de la robotique. La vision animée propose de lier dans le comportement du système les traitements cognitifs et le contrôle du regard, plutôt que de les considérer comme deux processus indépendants. Elle accorde une grande importance aux capacités d'adaptation de la stratégie de perception. Ainsi, les systèmes de vision animée ont pour but d'intégrer des mécanismes permettant de modifier la stratégie de perception suivant l'évolution de l'environnement. Cette recherche d'interactivité, avec le monde observé, nécessite un fonctionnement en temps réel du système et donc l'élaboration de stratégies visuelles rapides.

1.3.2.3 Le perception active de R. Bajcsy

Une autre facette du concept de vision active a été proposée par R. Bajcsy [17, 18]. Elle définit son approche, la perception active, comme l'étude de stratégies de modélisation et de contrôle de la perception visuelle. Cette approche considère la manière d'exploiter un capteur passif en changeant délibérément l'état de ces paramètres suivant une stratégie de perception. Le problème consiste à définir des stratégies de perception à appliquer sur le processus d'acquisition qui vont dépendre de l'état courant du capteur et de la tâche du dispositif de vision. La perception active intègre dans le dispositif de vision des fonctions intelligentes de contrôle. Ces fonctions se matérialisent par des mécanismes de raisonnement et de prise de décision

qui permettent d'établir un contrôle du système de vision. Ce contrôle se base sur un retour de l'information visuelle qui n'est pas uniquement effectué sur les données sensorielles mais aussi sur le processus d'interprétation de ces données. Le retour de l'information visuelle est une combinaison d'informations présentées à différents niveaux du système et qui dépend d'un modèle de connaissance *a priori*. Donc, la perception active nécessite la modélisation du capteur, des dispositifs de traitements et de leurs interactions pour mettre en œuvre différentes boucles d'asservissement au sein du système.

De part sa conception, la perception active est la première approche à intégrer directement le capteur dans la boucle de rétroaction et à ne pas séparer l'aspect capteur "hardware" de l'aspect algorithmique. De plus, l'aspect modulaire de cette approche interactive est bien adaptée à l'implémentation sur machine.

1.3.2.4 Le vision intentionnelle de Y. Aloimonos

Une autre notion nommée "la vision intentionnelle", a été présentée par Y. Aloimonos en 1990 [19, 20, 21]. Y. Aloimonos considère que rechercher une représentation complète d'un espace visuel réel nécessiterait une puissance de calcul extraordinaire car il y a tout simplement trop de choses et détails qui composent notre environnement. En effet, la quantité d'information contenue dans le signal image est bien plus importante que la quantité d'information susceptible d'être traitée par les systèmes actuels.

Pour cette raison, l'objectif de la vision intentionnelle est de déterminer quelles sont les relations entre le monde et le système qui doivent être implémentées. En d'autres termes, quelles sont parmi toutes les informations contenues dans l'image celles qui doivent être exploitées par le système, et comment ces informations doivent être représentées et traitées. Manifestement, pour réaliser un objectif spécifique donné, un système de vision n'a besoin que d'une connaissance partielle de son environnement. De plus, si le système visuel connaît quel type d'informations lui sont nécessaires, il est susceptible d'interagir dynamiquement avec son espace visuel afin de faciliter la détection de ces informations pertinentes.

La vision intentionnelle ne considère pas la vision comme un problème isolé mais comme une partie d'un système complexe qui interagit d'une manière spécifique avec

le monde. En d'autres termes, la vision intentionnelle défend une approche modulaire de la vision afin d'isoler les problèmes et de faciliter leurs résolutions. La modularité est un principe général qui est exploité pour concevoir et analyser des systèmes à la fois complexes et étendus. La vision intentionnelle suggère que ses modules soient des comportements qui sont définis par des séquences de raisonnements, d'événements et d'actions cognitives qui coopèrent et interagissent avec l'environnement. Cette notion de coopération entre modules cognitifs est comparable au concept de routines visuelles proposé par Shimon Ullman.

1.4 Routines visuelles

1.4.1 Aspect conceptuel

En 1984, S. Ullman propose un système de vision basé sur la décomposition des tâches visuelles en différents modules. Il propose la distinction entre les opérations visuelles précoces effectuées en parallèle ("bottom-up, and without focused attention") et les opérations plus tardives qui ne peuvent opérer qu'avec le bénéfice de l'attention focalisée sur un unique objet ou groupe. Il appelle ces dernières des "routines visuelles".

Ces routines visuelles peuvent être considérées comme des procédures séquentielles basées sur des opérateurs basiques de vision. L'objectif des routines visuelles consiste à extraire les propriétés abstraites des formes présentes dans l'image ainsi que leurs relations spatiales.

En effet, la perception de propriétés abstraites des formes est un problème difficile pour un dispositif artificiel. Tout d'abord, cette perception, apparemment simple pour notre système de vision, nécessite pour un dispositif artificiel des calculs complexes. Ensuite, le nombre et la variété de ces propriétés impliquent une capacité de calcul élevée. Pour répondre à ce problème, Ullman propose une approche du problème de vision à la fois ascendante (Bottom/up) et descendante (Top/down). Il divise le système de vision en deux processus distincts. Le premier est un processus ascendant dont le rôle est de créer une base de représentation (segment de droite, splines,...) de l'image. Ensuite, un second dispositif descendant exploite cette base de représentation en lui appliquant des processus cognitifs nommés routines vi-

suelles(fig. 1.2).

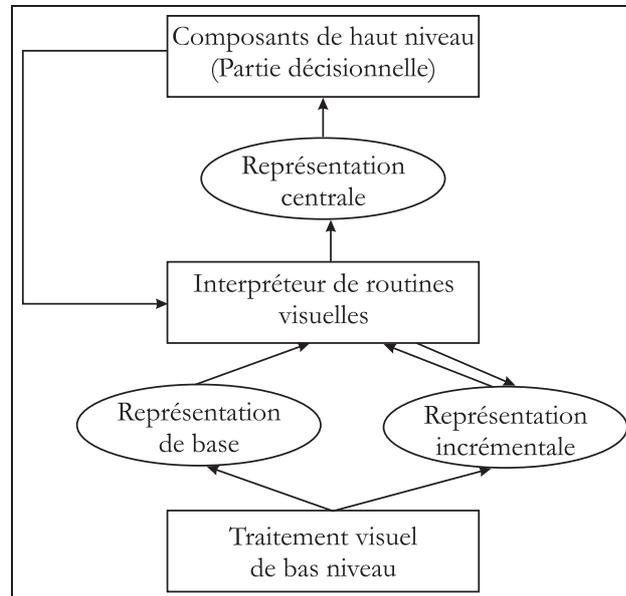


FIG. 1.2 – *Modèles des routines visuelles. Le traitement est décomposé en deux étapes principales : une ascendante consistant en la création de la représentation de base de bas niveau, la seconde, descendante, appliquant les procédures visuelles sur cette représentation pour extraire les propriétés spatiales nécessaires pour satisfaire les buts.*

Le processus ascendant fournit une représentation non construite ou non liée de l'information visuelle. Le rôle des routines visuelles est de déterminer les propriétés et les relations spatiales qui permettent de lier les informations et ainsi définir des objets ou des parties d'objets. Une routine visuelle est définie comme une séquence d'opérations élémentaires permettant d'identifier une information spécifique.

L'intérêt de cette approche est d'assembler divers routines à partir d'un nombre fixe d'opérations élémentaires dans le but de réduire la puissance de calcul nécessaire pour identifier une grande variété d'objets. La nature de ces opérations élémentaires et de la base de représentation sont liées et sont déterminées ou sélectionnées suivant les objectifs du système.

Il faut noter deux aspects sous-jacents de cette approche. Tout d'abord, certaines opérations élémentaires employées par les routines visuelles peuvent être exercées sur

une partie réduite du champ visuel à la quête d'une information spécifique. Ainsi, cette approche allie la notion d'étude locale et de focalisation afin d'optimiser les ressources de calcul du système. Un second aspect notable des routines visuelles est son aspect incrémental et empirique. En conservant, les résultats obtenues par l'application de routines visuelles, ils peuvent être utilisées comme stimuli d'autres routines afin de cibler itérativement la reconstruction de l'information.

1.4.2 Élaboration des routines visuelles

Comme expliqué précédemment, Ullman postule le fait que la perception de la forme d'un objet, et des relations spatiales entre ses différentes parties, peuvent être accomplies par des routines visuelles sophistiquées. Ces routines de "haut niveau" sont construites à partir cinq opérations élémentaires, appelées des **routines primitives** qui sont:

- La première opération consiste à déplacer le centre focal de traitement (shifting the focus of processing), c'est-à-dire à changer la cible de l'attention. Cette opération permet à toutes les routines visuelles d'être appliquées à n'importe quelle région du champ visuel, simplement par le déplacement du "centre focal" de traitement cognitif. Il correspond à un **changement de la "position" de l'attention focale**.
- La seconde est l'indexation (indexing). L'indexation implique la **sélection d'une localisation où quelque chose apparaît comme étant "différent"** ou saillant dans son contexte d'occurrence, comme dans les phénomènes de "pop-out visuel"¹ ainsi que les divers signaux qui capturent l'attention.
- Une autre opération intitulée "coloriage" (coloring) est utilisée pour faire une recherche à l'intérieur d'une région en diffusant un certain type d'activation à l'intérieur de ses frontières.
- La quatrième est le tracé de frontières (boundary tracing). Cette opération permet de déterminer si deux localisations sont situées sur les frontières du

1. Ce terme utilisé en psychologie de la vision, désigne le phénomène de focalisation visuel lié à la partie d'une scène qui "saute aux yeux" (pop-out) par opposition à une exploration sérielle de l'image. Typiquement, un observateur regardant une boîte contenant 10 balles bleues et une balle rouge, aura son regard attiré vers la cible "balle rouge". C'est un phénomène de "Pop-Out".

même objet.

- Enfin, la cinquième est le marquage (marking). Le marquage d'une localisation est une opération qui signale cette localisation comme étant celle qui doit être remémorée de telle sorte que son accès soit facilité, ou rendu plus rapide, à un moment ultérieur.

Cependant comme l'explique Dana Ballard dans [22], il est indéniable que l'analyse de scène visuelles nécessite plus d'opérations visuelles élémentaires que celles qui sont décrites ici par Ullman. De plus, malgré son aspect modulaire, la proposition de Ullman, ne s'appuie que sur des opérations passives. En effet, même si l'architecture globale de son approche permet une certaine rétroaction à travers les flux ascendants et descendants (fig. 1.2), il n'en demeure pas moins que les routines primitives agissent de manière passive. Aussi, sans avoir la prétention de "rivaliser" avec les prestigieux travaux de Ullman, nous proposons à travers ce manuscrit d'étudier comment il serait possible d'introduire un aspect "rétroactif" dans cette approche modulaire. Cela nous a conduit à mettre en œuvre un système de vision précoce dédiée que nous développerons dans les chapitres suivants. Aussi, afin d'argumenter et de motiver nos travaux, un bref état de l'art sur les différents systèmes dédiés à la vision active est proposé.

1.5 Systèmes de vision active

De manière générale, la composante "active" vient dans la plupart des cas se superposer au fonctionnement et aux problèmes de la vision passive. Bien que contribuant à l'augmentation de la fiabilité du système, la vision active introduit des tâches supplémentaires consistant à déterminer à tout moment les valeurs des paramètres du capteur puis à en réaliser les réglages [23]. Un autre problème de taille consiste en la gestion de l'accès des différents traitements à la ressource limitée que constitue la capteur!

Dans ce contexte, les capteur d'images sont alors considérés comme des modules de traitement fournissant une description de la scène aux traitements de niveaux supérieurs. L'objectif d'un système de vision active est donc d'interagir et de s'adapter à son l'environnement visuel en introduisant le contrôle du capteur dans le proces-

sus de perception. Cette particularité nécessite l'implémentation de mécanisme de retour de l'information visuelle pour faire évoluer le système de perception. Toutefois comme le note K. Pahlavan dans [24], un système de vision active peut être décomposé en 4 parties qui sont:

- Les paramètres mécaniques permettant le positionnement et le mouvements de ou des des caméras,
- Les paramètres optiques prenant en compte tous les problèmes d'objectif, zoom, mise au point,...
- Les paramètres électroniques liés aux capteurs tels que la conversion A/D, l'échantillonnage spatial et temporel, le rapport signal/bruit,...
- Les paramètres algorithmiques correspondant aux traitements et pré-traitements d'image.

Dans la littérature, de nombreuses tentatives de systèmes plus ou moins complets de vision active ont été proposés. Dans de nombreux cas ceux-ci ont surtout axés leurs recherches sur l'élaboration structure robotique mono ou multi-oculaire permettant de "plagier" le système oculaire humain. Cependant comme le souligne Thierry Viéville dans [25], *"la vision active ne se résume pas au contrôle d'une tête robotique mais plus généralement au contrôle de l'ensemble des paramètres du capteur. Ainsi, il est essentiel de noter le rôle majeur que joue le capteur au sein d'un système de vision active."*

Dans la suite de ce paragraphe, quelques travaux significatifs sont présentés en guise d'état de l'art. Il est cependant relativement difficile d'établir un quelconque classement puisque chaque entreprise a cherché à répondre dans la plupart des cas à un besoin qui lui était propre. On se contentera donc d'étudier dans un premier temps les travaux ayant plutôt axés leurs recherches sur la construction de têtes robotiques et la partie mécanique; il est toutefois à noter que, dans la plupart des cas, ces travaux ont aussi amené quelques contributions sur des aspects plus algorithmiques mais ce n'était pas là leurs principaux objectifs. Dans une seconde partie, un bref état de l'art est brossé sur les systèmes intégrés de vision active. En effet, quelques travaux ont investi le champ de la vision active à travers les yeux de la microélectronique et des technologie VLSI². Ces réalisations s'attachent dans la plupart des cas, à étudier les

2. Very Large Scale Intégration: Ce terme définit des circuits intégrés à haut niveau d'intégration, c'est à dire contenant plusieurs centaines de milliers de transistors.

possibilités d'intégration sur silicium d'algorithmes ou de techniques de perception propres à la vision active. Enfin dans un dernier paragraphe, une brève revue de système complet de vision active est proposée. Contrairement aux précédents, ces systèmes considèrent la globalité de la chaîne de perception active en allant de la mécanique à l'algorithmie en passant par les différents paramètres de contrôle.

1.5.1 Têtes robotiques pour la vision active

La manière la plus naturelle pour effectuer une rétroaction sur un système de vision est d'établir un contrôle mécanique de la caméra. Aussi la majorité des travaux réalisés en vision active sont basés sur le contrôle des paramètres géométriques de la caméra tels que la position, le zoom, l'orientation, le focus ou encore la vergence pour les systèmes binoculaires. Afin d'expérimenter cette approche, divers dispositifs robotiques de contrôle du regard ou têtes robotiques ont été développés par les laboratoires de recherche.

L'ensemble de ces dispositifs dédiés à l'implémentation de systèmes de vision active sont de nature diverse. Évidemment, les plate-formes expérimentales ont été contraintes par les solutions technologiques disponibles lors de leur conception. Mais, cette hétérogénéité provient aussi du fait que généralement chacune d'elle a été conçue pour une application spécifique. Les systèmes sont inspirés des mécanismes biologiques ou au contraire basés sur des principes mathématiques spécifiques. Malheureusement, aucune propriété universelle n'a pu être réellement établie pour caractériser les aspects mécaniques, algorithmiques et architecturaux de ces dispositifs de vision. L'absence de principes généraux qui permettrait de réduire les investissements et les temps de conception de ces systèmes limite leur exploitation.

Citons à titre d'exemple les quelques réalisations proposées sur le figure 1.3 comme la tête robotique développée par l'INRIA (fig. 1.3-d) qui est un système stéréoscopique permettant le contrôle de la direction du regard et l'adaptation des paramètres des objectifs. Ce système utilise un minimum de degrés de liberté (panorama, inclinaison et vergence d'une seule caméra). Les deux axes optiques sont contraints pour être toujours alignés quelque soit le mouvement du dispositif. Une des applications privilégiées de cette tête active est la reconstruction tridimensionnelle.

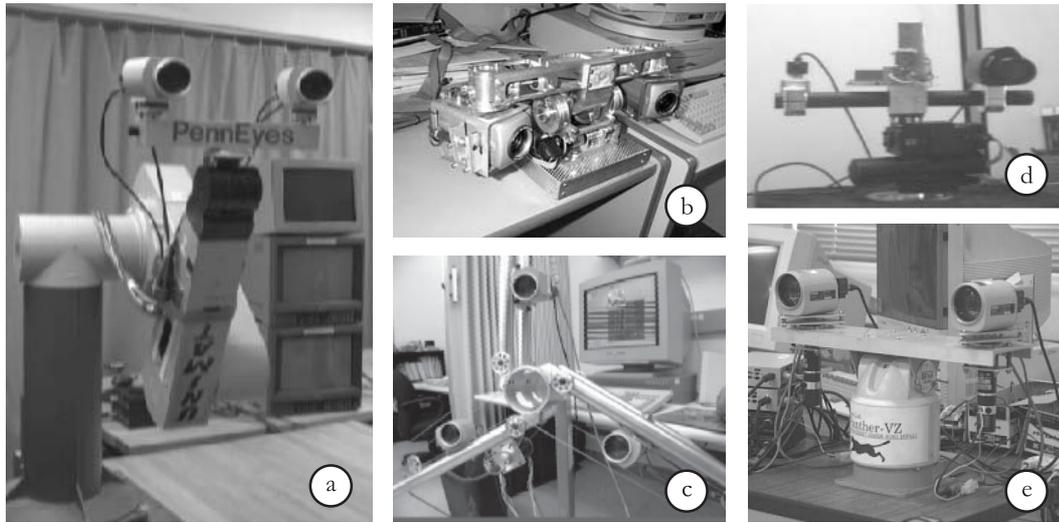


FIG. 1.3 – Exemples de dispositifs robotiques de contrôle du regard: **a** - la tête robotique *PennEyes* du laboratoire GRASP de Philadelphie, **b** - la tête robotique *CeDAR* du laboratoire RSL en Australie, **c** - la tête robotique trinoculaire *CardEye* du laboratoire CVIP de Louisville, **d** - la tête robotique de l'INRIA de Sophia-Antipolis, **e** - la tête robotique *PANTHER-VZ* du laboratoire japonais Shirai.

La dispositif *PennEyes* du laboratoire GRASP (fig. 1.3-a) de Philadelphie est une tête robotique binoculaire montée sur un robot de type PUMA[26]. Ce dispositif permettant de suivre des objets est architecturé autour de 2 caméras CCD et d'un réseau de DSP TMS320C40 permettant un calcul sur le flot à 60Hz. Les deux caméras sont équipées de zoom motorisé et peuvent se déplacer à $1000^\circ/s$. Dans ce projet, le suivi 3D d'objets est fait en jouant sur les mouvement du bras et les paramètres optiques (zoom).

Le système australien *CeDAR* admet la particularité d'un fonctionnement par câble. La plupart des actionneurs sont mus par des enroulements et déroulements de câbles autour des différents axes. La particularité de ce système se situe avant tout au niveau mécanique plus qu'au niveau du processus de vision active.

Le projet *CardEye* est basé sur un système de vision active trinoculaire. Le système trinoculaire est lui même monté sur un bras robotique à 3 degrés de liberté. On notera l'utilisation de lumières structurées permettant de lever les ambiguïtés lors de tâches de reconstruction. Une des principales difficultés dans ce systèmes

concerne le placement optimal des trois caméras.

De nombreuses autres plate-formes basées sur des structures de type "tête robotique" ont été réalisées et le lecteur intéressé trouvera de nombreuses références dans les ouvrages de Thierry Viéville [25] et de H. Christensen [27] ainsi que dans la thèse de Claus Andersen [28]. On notera toutefois une réalisation relativement originale de W. Mayol à l'université d'Oxford. Ce dernier a réalisé la première plate-forme de vision active directement embarquable sur un vêtement (Wearable Active Vision Platforms)[29]! Cette plate-forme (Fig. 1.4) située au niveau du cou de la personne



FIG. 1.4 – Plate-forme robotique "embarquable" de l'université d'Oxford.

est équipée de capteurs inertiels et permet une stabilisation d'images en temps réel. Cette stabilisation est réalisée en couplant les données image et inertielle. Les auteurs restent relativement vagues sur les applications potentielles d'un tel système qui tire avant tout son originalité des problématiques d'intégration.

1.5.2 Systèmes intégrés de vision active

Un autre aspect moins conventionnel des systèmes de vision active concerne la microélectronique. En effet, quelques équipes de recherche se sont posées la question de savoir s'il était possible d'intégrer sur silicium ou dans des composants programmables des "comportements" basés sur la vision active.

Dans ces approches originales, Vladimir Brajovic a développé un imageur [30]

dont la particularité est de s'adapter aux intensités lumineuses en adaptant la réponse des pixels suivant l'histogramme cumulé de l'image perçue. Il met en avant l'aspect "adaptatif" de son système, ce qui lui confère un intérêt non négligeable pour des problématiques de vision active. L'une des particularités électronique de l'imageur proposé est la conversion lumière/temps. En effet, dans son système la lumière est convertie en une impulsion électrique dont la durée est directement proportionnelle à l'éclairement. Cette conversion Analogique/Digital un peu particulière permet alors d'acheminer de manière aisée les valeurs à un processeur intégré auprès de la matrice. Malheureusement cette réalisation comme beaucoup de systèmes intégrés aborde des problématiques de microélectronique intégrés avant celle de la vision active qui se manifeste plus comme un contexte expérimental.

D'autres travaux traitant de la focalisation d'attention ont été proposés et sont quasiment toujours fondés sur l'architecture dite "Winner Take All" (WTA). L'idée générale dans ces recherches est d'élaborer une carte de primitives remarquables (saliency map) dont le concept originel a été proposé par Koch et Ullman [31]. Ce modèle, biologiquement plausible, est issu d'observations faites en neurobiologie et psychophysiologie. Dans le modèle des "saliency map", le mécanisme WTA permet de sélectionner la primitive la plus remarquable en ignorant purement et simplement les autres. Cette approche est purement bottom/up et ne permet pas de rétroaction du haut vers le bas.

Giacomo Indiveri de l'université de Zurich aborde le problème en utilisant une caméra CCD auquel il adjoint un circuit ASIC analogique dédié à ces opérations [32, 33]. Bien que de philosophie différente, cette approche s'apparente aux techniques de réseaux de neurones. Dans ces expérimentations, Indiveri utilise une zone 16x16 de l'image dans laquelle il sélectionne la primitive la plus remarquable. D'après ces résultats, la sélection semble se faire sur le plus fort contraste dans une région donnée.

Un travail relativement similaire a été proposé par l'équipe de De Weerth à l'université de Géorgie (USA). La différence avec les travaux précédents réside dans la possibilité de grouper dynamiquement des pixels répondant aux mêmes caractéristiques. Ce groupement se fait de manière analogique à l'aide de transistors montés en switches.

Au niveau de la conception même des imageurs, il est toutefois important de

souligner les travaux faits dans le cadre de la vision fovéale. En effet, en considérant la vision active comme une tentative de copie du système visuel humain, il est alors important de considérer les systèmes basés sur des imageurs à distribution fovéale. Un des avantages majeurs de la vision fovéale hormis la transformation que lui confère une invariance aux rotations et simplifie grandement le calcul des changements d'échelle, est le nombre restreints de pixels à traiter. De plus, la transformation fovéale permet avec un petit nombre de pixels d'être relativement représentative dès lors que l'on soit centré sur la zone à étudier. Différentes études relativement diverses ont été menées sur la question. On citera à titre indicatif les travaux de Bederson qui a réalisé une mini tête robotique à vision log/polaire [34]. Toutefois on remarquera dans cette réalisation, le fait que le capteur est une caméra conventionnelle CCD après laquelle un processeur applique la transformation cartésien/ log-polaire (Log/polar Warping). Au niveau conception même de l'imageur, on notera les études menées par Park [35], l'équipe de Sandini [36] ou encore Rojer [37]. Naturellement cette liste est loin d'être exhaustive.

Cet aperçu des recherches alliant intégration électronique ou microélectronique et vision active montre l'étendue des travaux dans ce domaine de la vision. Il prouve aussi, si cela est encore nécessaire, la place prépondérante du capteur et des traitements qui lui sont associés. Dans le paragraphe suivant, on se propose de présenter succinctement les systèmes prenant en compte tous les aspects de la vision active.

1.5.3 Systèmes complets de vision active

Le premier élément qui fait que chaque dispositif de vision active est unique est le lien étroit entre la nature de la tête robotique et le contrôle de l'attention. Aussi parmi toutes les recherches menées à ce jour, trois systèmes de vision dite "active" peuvent être soulignés:

- Le système MEDUSA de Y. Aloimonos
- L'architecture VAP: Vision As Process
- Le système TEA-1 proposé par Bacsy et Ballard.

1.5.3.1 Système MEDUSA

MEDUSA [19] est un système de vision active construit avec des capteurs inertiels, une caméra mobile, un outil de préhension visible de la caméra, le tout monté sur une plate-forme mobile. Le système traite les images tout en se déplaçant et en observant l'environnement. Deux structures de données principales sont utilisées par tous les composants du système: les images et les flots optiques normaux.

MEDUSA est composé de différents modules assurant les traitement de base suivant:

- Calcul de la composante normale du flot optique en chaque point,
- Trouver les zones en mouvement indépendamment de l'egomotion³
- Trouver les objets s'approchant du système.
- Suivre l'objet en déplacement en le gardant au centre de l'image
- Si l'objet s'approche, est-ce-que le point d'impact sera dans l'image? Savoir si l'objet heurtera MEDUSA?
- Commander les axes de MEDUSA pour intercepter l'objet en mouvement.

Malheureusement, le peu d'informations disponibles sur le système MEDUSA ne permet pas de se faire une opinion claire sur la probité du système. On notera en conclusion que ce système admet une architecture décomposé en tâches.

1.5.3.2 Architecture du système VAP(*Vision As Process*)

Le projet VAP, qui a débuté en juillet 1989, est le fruit d'un consortium de sept laboratoires européens. Les partenaires de ce projet sont l'université d'Aalborg, l'université de Surrey, KTH, Royal Institute of Technology, l'université de Linköping, LTIRF-INPG, LIFIA-INPG et l'université de Genoa. L'objectif de ce projet était de démontrer qu'un système de vision doit être implémenté comme un processus continu.

L'architecture proposée [38, 39] est constituée d'un ensemble de modules indépendants qui communiquent à travers un serveur de messages (fig. 1.5). Ce système considère le problème d'intégration des processus qui modélisent l'environnement

3. egomotion: Mouvement propre de la caméra

en termes de descriptions qualitativement différentes (descriptions 2D, 3D, étiquetage des objets) dans le but d'exploiter leur complémentarité. D'autre part, cette architecture a pour but d'autoriser le contrôle des mécanismes d'attention et de traitement qui constitue le dispositif de perception. Ce second aspect est relatif aux contraintes temps réel imposées par un processus de vision continu.

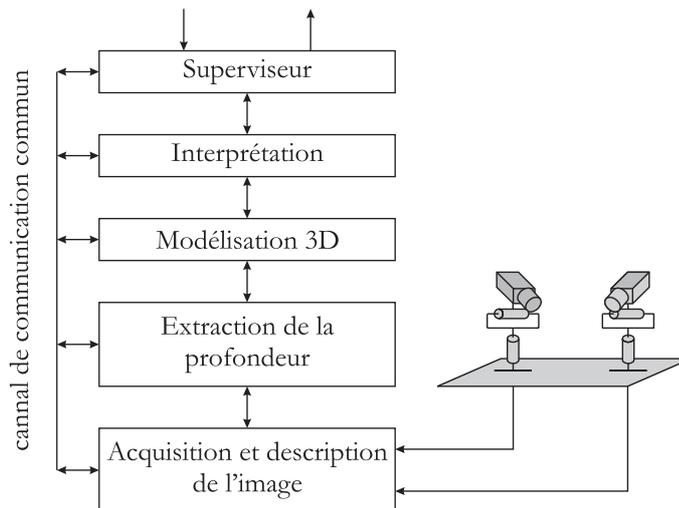


FIG. 1.5 – Architecture du système VAP.

Ce système possède une structure flot de données "*bottom-up*" allié à une structure "*top-down*". La structure "*top-down*" est exploitée pour permettre un contrôle direct des modules bas niveaux du systèmes

1.5.3.3 Système TEA-1

Les premières idées sur lesquelles reposent le système TEA-1 on été proposées dans [40] et reformulées dans [15] au travers du paradigme de la vision animée. Ce dernier a été développé en 1992 par [41, 42]. Il a pour objectif de répondre à des questions par une exploration contrôlée d'une scène en modifiant la position de la caméra. Le cœur de TEA-1 est basé sur des réseaux bayesiens. Le système est dirigé par un fonctionnement descendant: le contrôle de la caméra est directement en relation avec le but du système et avec la construction de la représentation. Tout comme le système MEDUSA, l'architecture TEA-1 est un système qui privilégie la décomposition en tâches.

1.6 Motivations et objectifs

Le problème de vision est abordé dans cette thèse suivant une approche active qui consiste à exploiter le capteur selon une stratégie d'observation intentionnelle. Jusqu'à présent la plupart des dispositifs développés utilisent un contrôle mécanique du capteur afin d'exploiter dans le processus de perception les paramètres optiques et les mouvements de la caméra. Dans cette thèse, il est proposé d'instaurer une conduite active de la chaîne de perception au niveau de l'imageur et des procédés de traitement. Pour introduire l'imageur dans une boucle de vision active, il est nécessaire d'appliquer un contrôle sur celui-ci à partir de l'information visuelle. Parmi les technologies d'imagerie actuelle, seule l'imagerie CMOS permet d'exercer un contrôle sur l'imageur. Plus précisément, l'architecture fonctionnelle des imageurs CMOS est basée sur un mode de lecture aléatoire des pixels qui influence l'échantillonnage spatio-temporel du signal visuel. La sélection contrôlée des pixels sur un imageur CMOS permet d'introduire l'imageur dans une stratégie de perception active et locale. L'objectif est de sélectionner l'information directement sur l'imageur afin d'accroître la cadence de rafraîchissement de cette information et de focaliser les ressources de calcul sur son traitement.

L'efficacité d'un système de vision active repose sur l'interaction entre le capteur et le système. La mise en œuvre de cette interaction fait apparaître les contraintes d'un système temps réel qui sont nécessaires à la réactivité du dispositif. Afin d'instaurer une rétroaction temps réel sur l'imageur, il faut disposer d'un lien de communication privilégié entre le capteur et le système actif. Pour satisfaire ces contraintes, il est proposé de délocaliser un système de vision active bas niveau au sein d'une architecture matérielle implantée auprès du capteur. Un premier objectif est d'effectuer une présélection de l'information au début de la chaîne de perception suivant les objectifs visuels considérés. D'autre part, le système doit gérer de manière autonome les tâches de perception bas niveaux afin de décharger le processus de vision haut niveau.

Les besoins d'un système de vision active étant fortement dépendant de la tâche considérée, cette thèse propose une plate-forme de recherche versatile permettant d'élaborer et d'évaluer des dispositifs de vision embarqués. Cette plate-forme exploite la technologie FPGA pour fournir une cible d'implantation matérielle riche

et adaptée aux besoins des systèmes de vision. Suivant les caractéristiques de cette plate-forme un schéma d'implantation est proposé pour élaborer un système de vision active exploitant un imageur CMOS.

Chapitre 2

Les technologies de perception visuelle

L'implantation standard d'un système de vision est composée d'un dispositif d'imagerie (une simple caméra) et d'un ordinateur hôte qui communiquent à travers un lien physique. Le rôle du dispositif d'imagerie consiste uniquement à convertir le signal d'éclairage en une information électrique, puis à transmettre via un bus de communication l'information visuelle vers une unité de traitement située sur un ordinateur hôte. Le développement d'un système de vision à partir d'une telle organisation matérielle présente de nombreux avantages. Tout d'abord, l'environnement de développement du système est purement logiciel. Cet environnement logiciel fournit un niveau d'abstraction entre l'architecture matérielle et la définition du système qui permet un développement indépendant des considérations matérielles. Cette solution est sans aucun doute la plus flexible car elle permet d'utiliser des outils de développement haut niveau, de construire et d'exploiter des bibliothèques qui réduisent les temps de développement. De plus, cette approche permet de bénéficier de la puissance de calcul toujours croissante des ordinateurs actuels. Malheureusement malgré l'évolution prodigieuse des ordinateurs, les dispositifs de vision ont beaucoup de difficultés pour travailler à la cadence vidéo standard de 25 images par secondes. La quantité d'information contenue dans le signal vidéo et la complexité des traitements requièrent des ressources de calcul colossales. Cette exigence a conduit les chercheurs et les ingénieurs à développer des machines dédiées à la perception visuelle. Le but est d'accroître la capacité de calcul des systèmes de vision

à partir d'une adéquation entre la problématique de vision et la cible d'implantation matérielle. Ces machines dédiées se basent sur une organisation matérielle qui permet d'améliorer l'efficacité des unités de traitements constituant le système de vision. Pour mettre en œuvre cette optimisation, divers unités de traitements sont délocalisées le long de la chaîne de perception au sein de dispositifs dédiés. Cette délocalisation permet d'exploiter des systèmes de traitement conçus pour implémenter efficacement un algorithme spécifique. Ce type de préoccupations, qui dépasse le cadre de la vision artificielle, est celui d'un domaine de recherche particulier qui est motivé par les problèmes d'adéquation algorithmique, architecturale et matérielle.

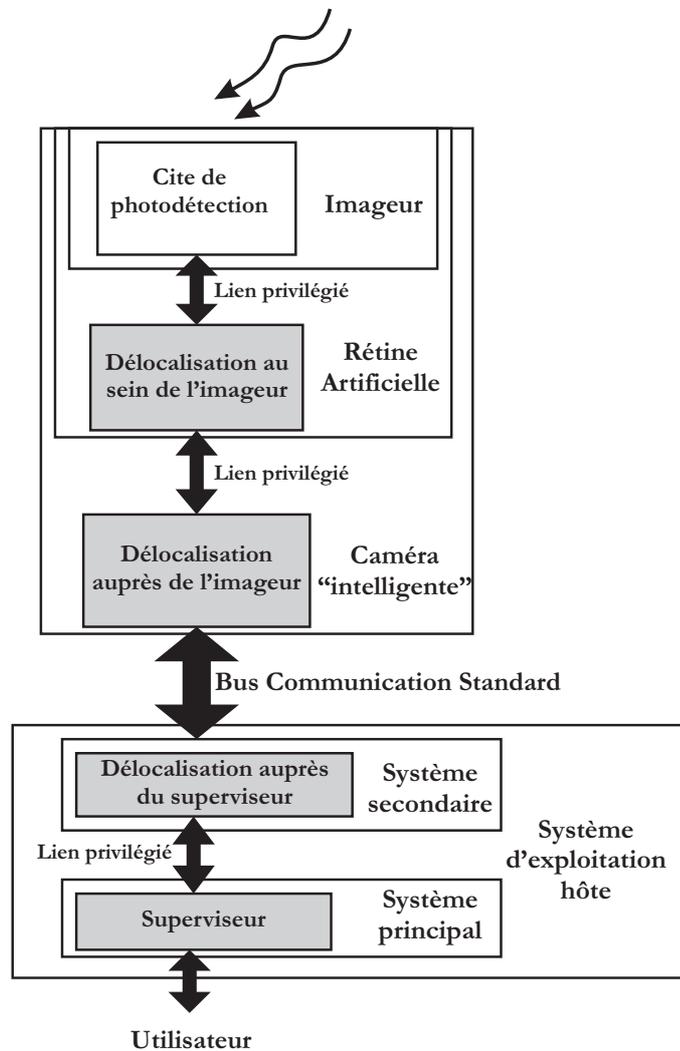


FIG. 2.1 – La délocalisation des traitements le long de la chaîne de perception.

L'objectif de tout système de vision est de fournir de l'information interprétée à un utilisateur. Son rôle consiste à réaliser une interface entre l'utilisateur et l'information visuelle brute décelée au niveau du site photosensible. Mis à part certains systèmes complètement autonomes, la première couche d'abstraction entre l'utilisateur et le photodétecteur est implémentée au sein d'un système d'exploitation informatique. Le long de cette chaîne de perception visuelle, la délocalisation des traitements s'articule suivant quatre niveaux principaux d'abstraction entre l'information visuelle brute et l'utilisateur (fig. 2.1). Ces niveaux sont hiérarchisés suivant leurs niveaux d'interprétation et la proximité du site de photodétection :

- Tout d'abord, des traitements bas niveaux peuvent être intégrés directement au sein de l'imageur pour constituer une rétine de vision artificielle. Ce niveau de délocalisation permet d'intégrer au plus près du signal détecté un haut degré de parallélisme et d'adéquation architecturale.
- Ensuite, afin d'exploiter la proximité du capteur et de bénéficier d'un lien privilégié avec l'information à traiter, la délocalisation des traitements peut être effectuée au sein de la caméra. Un tel système embarqué constitue une caméra dite "intelligente".

L'information visuelle est ensuite acheminée vers le système d'exploitation hôte à travers un bus de communication standard. Au sein de l'ordinateur hôte deux systèmes matériels peuvent être différenciés. Le système d'exploitation informatique est installé sur l'architecture de traitement principale de l'ordinateur. Cette couche logicielle supporte le niveau le plus élevé de représentation de l'information et supervise les couches de traitement inférieures. Il accède à l'information issue de la caméra à travers un dispositif secondaire qui dans le cas élémentaire se limite à la gestion de la communication avec la caméra. Cependant, on peut envisager de délocaliser des traitements au sein de ce système pour décharger le système principal en exploitant un co-processeur ou un calculateur dédié. Ce type de délocalisation est analogue à celle réalisée par la carte graphique d'ordinateur standard suivant un flot d'information inverse. L'objectif est de suppléer le processeur principal pour le décharger d'une partie des calculs les plus conséquents.

Ce chapitre est consacré à l'analyse des solutions technologiques envisageables pour concevoir une caméra "intelligente". Dans une première partie, les technologies d'imagerie basées sur la photodétection dans le silicium sont présentées. Ensuite,

les deux niveaux de délocalisation des traitements au sein d'une caméra sont discutés. Une section est dédiée aux rétines artificielles qui sont des composants qui intègrent des traitements au sein du composant photosensible. Les types d'architectures utilisées pour concevoir ces dispositifs de vision entièrement intégrés sont détaillés. Enfin, la dernière partie est consacrée aux technologies des systèmes vision embarqués. Les diverses cibles d'implantation matérielle permettant la conception de caméras "intelligentes" sont exposées. Les avantages et les inconvénients des solutions technologiques actuelles sont discutés suivant les préoccupations de cette thèse.

2.1 Les technologies d'imagerie

Le composant photosensible qui constitue la base des caméras (électroniques) actuelles est un capteur d'image basé sur la technologie des semiconducteurs et plus particulièrement la technologie du silicium. En effet le silicium présente une bonne caractéristique d'absorption dans le visible qui permet à ce composant de convertir l'éclairement en informations électriques. Dans le spectre visible, il existe aujourd'hui deux technologies d'imagerie majeures : la technologie CCD (*Charge Coupled Devices*) et la technologie CMOS ou CIS (*CMOS Image Sensor*). Ces deux technologies d'imagerie sont en concurrence directe [43, 44]. La technologie CCD a dominé le marché de l'imagerie durant de nombreuses années. Un historique retraçant l'évolution des technologies CCD et CIS peut être consulté dans l'article de E.R. Fossum [45], ou la thèse de C.S. Hong [46]. Lors de son apparition en 1970, le capteur CCD [47] présentait un faible bruit spatial (*Fixed Pattern Noise*) par rapport à ces concurrents. Cet avantage lui permis de devenir le capteur d'image électronique par excellence jusqu'en 1980. Il fut largement utilisé dans toutes sortes d'applications à vocation commerciale scientifique. Malgré le succès des rétines CCDs, les recherches sur les rétines CMOS ont perduré car ces capteurs d'image présentent un gros intérêt économique. En effet, sa compatibilité avec les technologies de l'électronique VLSI permet d'intégrer auprès de l'imageur une grande quantité d'électronique et de réduire le coût de fabrication du composant. Ainsi, lorsque les premiers capteurs CMOS présentant des performances comparables aux capteurs CCDs apparurent, cette technologie connut un engouement spectaculaire. Actuellement, les imageurs

CMOS sont utilisés pour de multiples applications. Dans le domaine grand public, ils sont utilisés pour les caméscopes, les appareils photo, les webcams ou encore les téléphones portables ... En ce qui concerne les domaines scientifiques, les capteurs CMOS sont exploités pour développer des caméras rapides [48, 49, 50] ou pour des applications qui nécessitent une grande dynamique de mesure ou la caractéristique logarithmique du pixel photodiode en mode courant est intéressante [51, 52].

Dans cette partie, le principe photodétection dans le silicium ainsi que les cellules de transduction électrique du signal sont présentés. Ensuite, les particularités des architectures d'imagerie CMOS et CCD sont détaillées et une comparaison technologique est établie pour faire opposés les avantages et inconvénients de chacun. Finalement, l'intérêt des imageurs CMOS pour implémenter système de vision et principalement un système de vision active est argumenté.

2.1.1 La photodétection

Les technologies CCD et CMOS sont basées sur la détection photonique dans le silicium. Lorsque le silicium est soumis à un éclairage, l'effet photoélectrique [53, 54] fait apparaître une paire électron/trou dans le semi-conducteur lorsqu'un photon d'énergie $h\nu$ est absorbé par un atome de silicium (h est la constante de Plank et ν la fréquence électromagnétique du photon). L'apparition d'une paire électron/trou est due à une transition de bande provoquée par l'énergie photonique de la lumière (fig. 2.2). Lorsque l'énergie d'un photon est supérieure à la largeur de la bande interdite ($E_g = E_c - E_v = 1.12\text{eV}$) un atome de silicium libère un de ces électrons de valence par une transition de bande. Cette transition de bande est à l'origine de la photodétection dans le silicium.

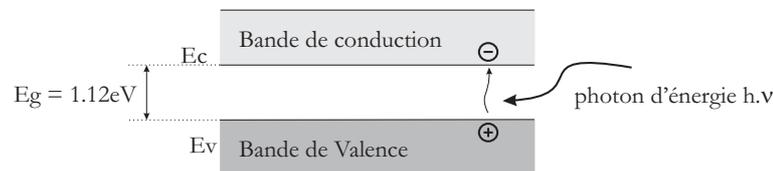


FIG. 2.2 – *Effet photoélectrique dans le Silicium.*

Afin d'exploiter ce phénomène physique pour traduire le signal lumineux en un signal électrique, deux types de cellules de photodétection sont utilisés.

Cette transduction peut être effectuée soit par une structure photoMOS, soit par une structure photodiode. Le principe de ces deux cellules de transduction est développé dans les sections suivantes.

2.1.1.1 La photodiode

La photodiode est une jonction P/N polarisée en inverse, qui génère un courant lorsqu'elle est éclairée par un flux photonique. La technologie CMOS permet de concevoir des photodiodes simples basées sur une structure N+/P. La région P correspond au substrat et la région N+ est fabriquée de la même manière qu'un drain ou une source de transistor NMOS.

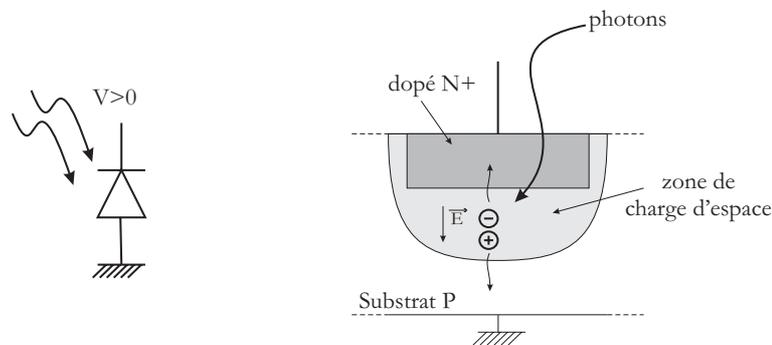


FIG. 2.3 – Structure d'un pixel photodiode.

La détection photonique est réalisée en polarisant la diode en inverse (fig. 2.3). La polarisation crée une zone de charge d'espace autour de la jonction. Le processus d'absorption conduit à la génération de charges électriques sous la forme de paires électrons/trous. Le champ électrique présent dans la zone de charge d'espace permet de séparer les charges positives et négatives qui font apparaître un courant électrique proportionnel à l'éclairement incident de la diode.

2.1.1.2 Le détecteur photoMOS

Le détecteur photoMOS est basé sur une capacité MOS (Métal Oxyde Semiconducteur). La structure de ce détecteur (fig. 2.4) consiste en un substrat de silicium dopé P, une couche d'oxyde mince (SiO_2) et une grille en polysilicium fortement conducteur. L'application d'une tension sur la grille de la cellule fait apparaître une

zone de dépeuplement sous celle-ci. Cette zone, appelée la zone de charge d'espace (ZCE), constitue un puits de potentiel qui permet de collecter et de retenir des charges photoniques, la lumière qui traverse la grille de la capacité MOS génère la création de paire électron/trou suivant le processus d'absorption. Sous l'effet du champ électrique présent dans la ZCE, les photoélectrons sont attirés sous la grille du détecteur photoMOS. Ainsi, le nombre de charges accumulées est proportionnel à l'éclairement.

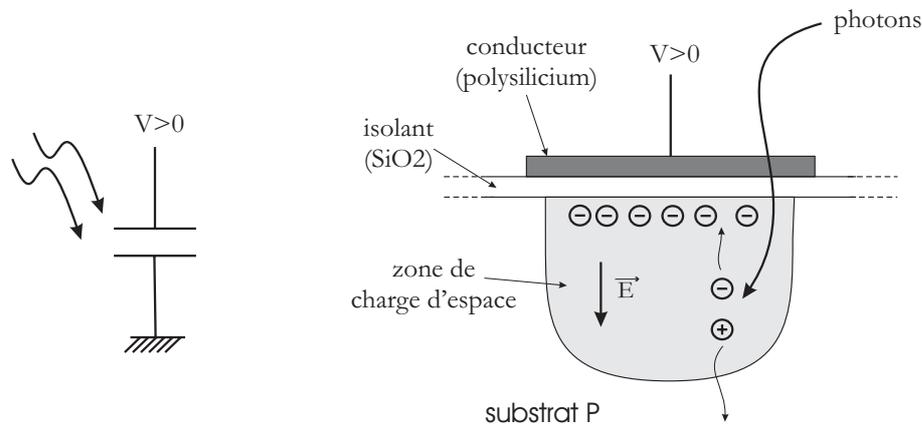


FIG. 2.4 – *Structure de détection photoMOS.*

Cette structure permet de réaliser la photodétection, mais aussi de stocker des charges. Cette propriété est la base de la structure des capteurs CCDs (fig. 2.5). En appliquant une séquence de polarisation appropriée sur les grilles de détecteurs photoMOS voisins, cette structure permet de transférer l'information lumineuse de proche en proche vers un circuit de lecture. Cependant, elle présente un rendement quantique inférieur à la photodiode.

Ces deux cellules de phototransduction sont à la base des technologies d'imagerie CMOS et CCD. Elles sont exploitées au sein d'architectures matricielles qui permettent un échantillonnage spatial et temporel du flux lumineux. Munies d'un circuit de lecture de l'information (séquentiel) ces composants constituent un imageur. Les technologies d'imagerie CMOS et CCD diffèrent à la fois par la nature des cellules de photodétection utilisées et par le mode d'accès à l'information visuelle utilisé. Leurs caractéristiques sont développées dans les sections suivantes.

2.1.2 L'imagerie CCD

Au sein des imageurs CCDs ("*Charge Coupled Device*"), chaque pixel est composé d'une zone photosensible, de type photoMOS ou photodiode connectée à une capacité MOS. Après un temps d'accumulation des photocharges, elles sont transférées de capacité en capacité par registre à transfert de charges analogiques (fig. 2.5). Ce registre permet d'évacuer l'information d'éclairement sous forme de charges qui sont ensuite converties en tension à partir d'une capacité de lecture externe à la matrice. Ensuite, cette tension est amplifiée et enfin numérisée (fig. 2.6). Dans

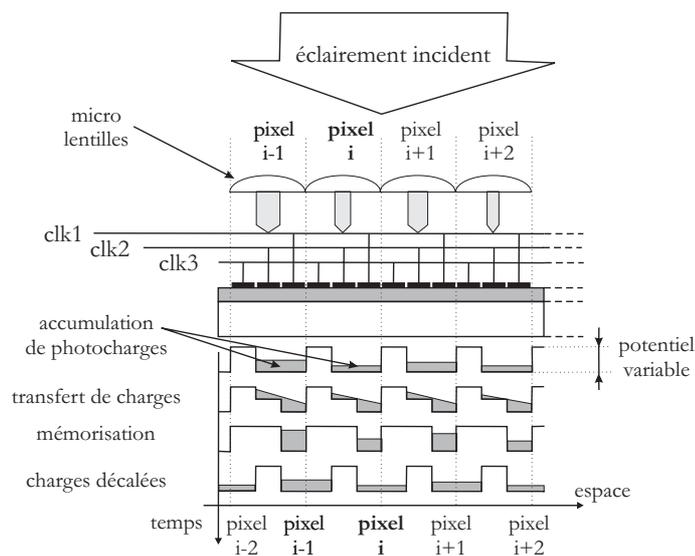


FIG. 2.5 – *Transfert de charges à partir d'un registre à décalage analogique (CCD).*

le cas d'une structure CCD standard (dite "pleine trame") constituée de détecteurs photoMOS, les éléments photosensibles sont exploités par le dispositif de lecture. Durant l'exposition, le pixel constitue le lieu de génération et de collection des électrons. Puis durant la phase de transfert, il constitue l'élément mémoire qui permet le transport des charges. Ce transport des charges de proche en proche est réalisé grâce à la propriété qu'ont les électrons à s'accumuler dans le puits de potentiel le plus profond (fig. 2.5). Ainsi dans une telle structure, l'intégralité de la surface du pixel est photosensible. Le facteur de remplissage (ou "*fill factor*") qui est le rapport de la surface photosensible sur la surface totale du pixel, est proche de 100%.

Le circuit qui réalise la lecture comporte une diode polarisée en inverse et un

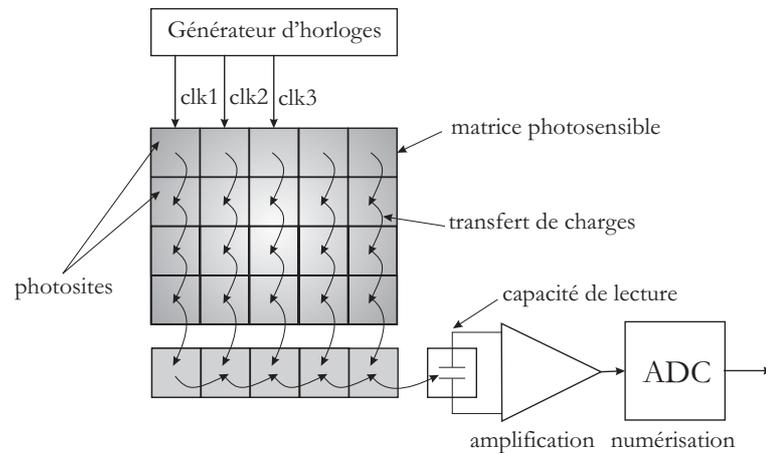


FIG. 2.6 – *Transfert de charges vers la capacité de lecture d'une matrice CCD.*

transistor fonctionnant en interrupteur commandé par un signal de *reset* (fig. 2.7). Un transistor suiveur assure l'adaptation d'impédance entre le bus de lecture et le photodétecteur 2.7. La lecture nécessite trois phases pour mesurer la quantité de charge présente dans la diode : une phase de reset, une phase d'intégration et une phase de mesure.

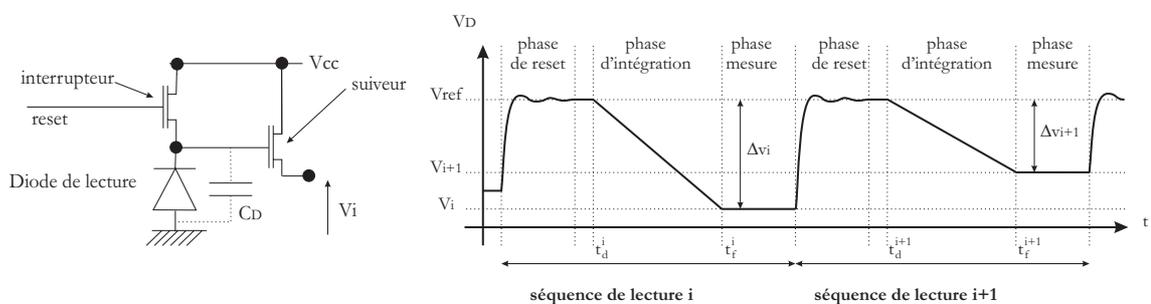


FIG. 2.7 – *Circuit de lecture par intégration.*

Lors de la *phase de reset*, le transistor interrupteur permet d'initialiser la tension aux bornes de la capacité interne C_D de la diode. Lorsque la tension au borne de la capacité atteint une certaine valeur de référence V_{ref} l'interrupteur isole la diode de la tension V_{cc} , la diode est dite flottante. A cette instant débute la *phase d'intégration*. Durant l'intégration, la capacité C_D de la diode se décharge et la tension à ces bornes V s'exprime en fonction du courant provenant de l'accumulation des charges

photoniques :

$$V = V_{ref} - \frac{1}{C_D} \int_{t_d^i}^{t_f^i} I_{photon}(t) dt$$

En considérant que l'éclairement reste constant durant la phase d'intégration, le photocourant I_{photon} peut être assimilé à une constante. Ainsi la variation de tension durant la phase d'intégration est directement proportionnel au photocourant pour un temps d'intégration δt donné :

$$\Delta V_i = V_{ref} - V_i = \frac{\delta t}{C_D} \cdot I_{photon}$$

Après le temps d'intégration δt , la variation de tension ΔV_i est mesurée durant la *phase de mesure*. Cette technique de mesure porte le nom de double échantillonnage corrélé (*Correlated Double Sampling*). L'avantage majeur de cette méthode est l'élimination du bruit en "*KTC*" de la capacité à partir d'une mesure différentielle. Une description détaillée de cette technique de lecture classique est proposée dans la thèse C. Cavadore [53] ou encore de M. Tabet [55].

La structure CCD élémentaire nécessite généralement un obturateur électronique pour pouvoir exécuter la lecture hors éclairage. Dans le cas contraire, si le temps de lecture de la matrice n'est pas négligeable par rapport au temps décalage des lignes de pixels la mesure de l'éclairement sera faussée. Car, durant la phase d'intégration(2.7) les pixels lus en dernier auront été plus exposés que ceux lus en premier, ainsi les quantités de charges accumulées ne sont plus significatives. Ce problème impose la séparation des dispositifs photodétection et de transport des charges. Une première solution consiste à rajouter un registre de stockage de même taille que la matrice photosensible, ce qui double la surface de silicium utilisée, ce sont les matrices CCDs dites "à transfert de trame". Une autre solution est l'ajout au sein du pixel d'un registre de transport protégé de la lumière. Ce type de structure est dite "CCD interligne" (fig. 2.8) permet de remplacer la cellule photoMOS par une photodiode au détriment du facteur de remplissage ($\sim 25\%$).

Par nature, les pixels de ces imageurs ne sont pas isolés électriquement les un par rapport aux autres. Lorsque la capacité maximale de stockage du pixel est atteinte, les charges accumulées s'échappent aisément vers les pixels voisins et les mesures d'éclairement effectuées sont faussées dans tout le voisinage. Ce phénomène est appelé le "*blooming*". Cette structure d'imageur présente un autre inconvénient lié au

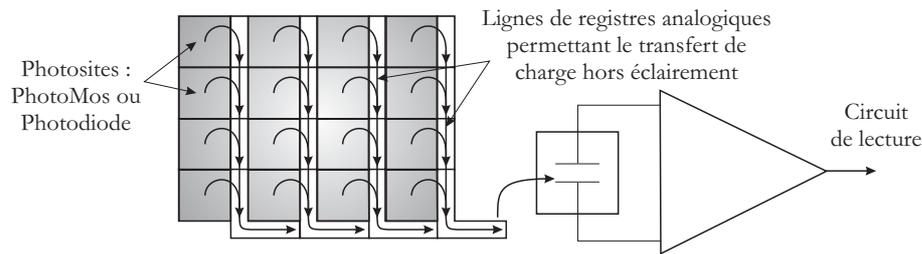


FIG. 2.8 – Structure d'une matrice CCD "intelligente".

mécanisme permettant le transfert des charges. Des signaux d'horloge complexes doivent être générés par ce dispositif. Ils nécessitent de hauts niveaux tensions pour assurer un transfert de charge efficace ce qui induit une consommation importante. Les niveaux tensions utilisés ne sont pas compatibles avec les technologies de conception standard. Cette technologie d'imagerie est complètement dédiée et de ce fait relativement coûteuse.

Cependant, la technologie CCD a acquis une grande maturité. Ces matrices offrent des performances optiques et électriques très élevées qui conduisent à une qualité d'image bien excellente. Cette technologie est incontournable pour les applications scientifiques qui nécessitent des performances radiométriques. Et, elle est majoritairement utilisée dans le secteur de la photographie numérique.

2.1.3 L'imagerie CMOS

Les premiers capteurs CMOS, appelés capteur d'images MOS-XY, ont été élaborés avant les dispositifs CCD. Ce type de matrice est appelée actuellement capteur PPS pour *Passive Pixels Sensor* par opposition aux nouvelles matrices CMOS dénommée APS pour *Active Pixels Sensor*. La structure des matrices CMOS sont présentées ci-dessous. Tout d'abord, la structure PPS est décrite pour aborder le mode de lecture particuliers de ces imageurs. Ensuite, la particularité de la structure APS est énoncée puis les diverses architectures de pixels utilisées par ces matrices sont développées.

2.1.3.1 La structure CMOS PPS

Chaque pixel des matrices PPS contient une photodiode connectée à la sortie du capteur par un bus d'accès colonne et un bus de lecture ligne (fig. 2.9). La conversion de l'éclairement en tension est réalisée au niveau du photosite. L'acheminement du signal de chaque pixel est réalisé dans le domaine des tensions. L'accès à l'information visuelle est analogue à l'accès d'une donnée dans une mémoire bidimensionnelle. A partir d'une adresse Y, le décodeur d'accès ligne connecte l'ensemble des photodiodes de la ligne sélectionnée aux bus de lecture colonne de la matrice. Ensuite, à partir d'une adresse X, le décodeur de lecture colonne connecte le bus colonne sélectionné à la sortie. Par nature, la structure de la matrice CMOS permet un accès aléatoire aux pixels. De ce fait, la lecture d'une image nécessite un dispositif d'adressage pour définir une séquence d'accès aux différents pixels constituant une image. Par conséquent, l'échantillonnage temporel du signal visuel dépend de l'ordre d'adressage des pixels. La phase de lecture des imageurs CCD présente un problème d'exposition variable des pixels si aucun dispositif permettant de figer l'image à un même instant (*shutter*) n'est inséré dans la structure. Dans la structure CMOS, ce problème due à l'accès séquentiel aux pixels à un effet différent. Le mode de lecture induit une indépendance temporelle des pixels. Ainsi, les mesures spatiales de l'éclairement sont réalisées à des instants différents. Ainsi, l'observation de mouvements rapides sans dispositif permettant de figer l'image conduit à des déformations du à l'incohérence temporelle des mesures. Les imageurs CMOS sont généralement équipés d'un "*shutter*" non uniforme, appelé "**Rolling Shutter**". Ce type de dispositif fige les valeurs de la rangée de pixels qui est consultée. Ce dispositif offre un compromis qui permet d'économiser le facteur de remplissage des pixels.

Les performances de ces premiers capteurs MOSXY ou PPS sont très faibles par rapport à celles des capteurs CCD. Le principal inconvénient de la structure MOSXY est la différence de capacité entre la photodiode de détection et le bus de lecture vertical. La photodiode de détection de faible capacité est directement connectée au bus de lecture vertical dont la dimension génère une forte capacité. Ainsi le faible signal utile généré par la photodiode est fortement dégradé. La résolution de ce problème d'adaptation d'impédance a donné naissance aux matrices CMOS APS.

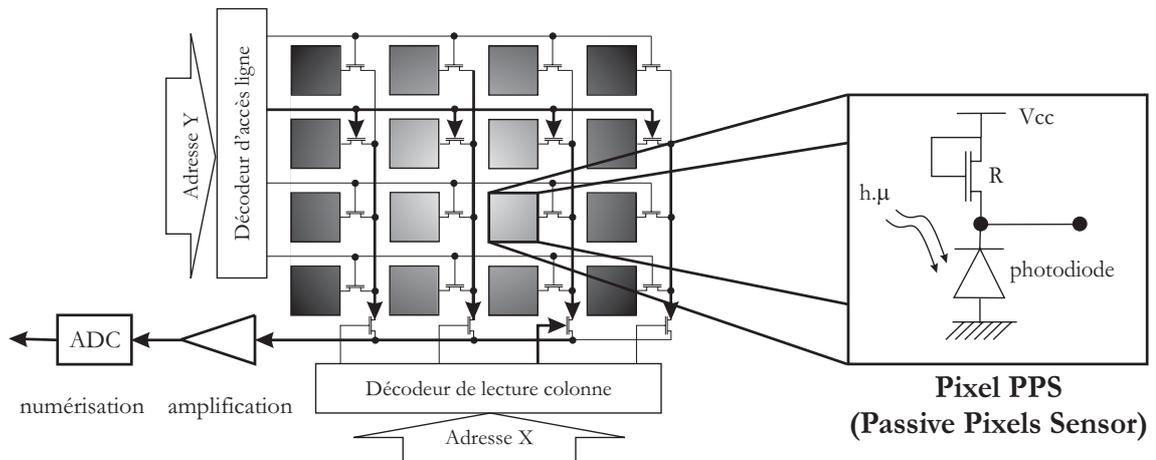


FIG. 2.9 – Structure d'un imageur CMOS PPS ou MOSXY.

2.1.3.2 La structure CMOS APS

Le concept APS (Active Pixels Sensor) consiste à intégrer au sein du pixel, auprès du photosite, un circuit d'amplification réalisant l'adaptation d'impédance entre le bus de lecture et la cellule photosensible (fig. 2.10). Le concept APS n'est pas réellement récent, mais la réalisation de capteurs CMOS APS performants n'est devenue possible que bien plus tard. L'intégration d'un élément actif dans le pixel dégrade le rendement du photodétecteur. La surface de photodétection est réduite par rapport à la surface totale du pixel. Ce facteur de remplissage diminue la sensibilité du pixel et donc la qualité de l'image. L'intégration du pixel APS, en conservant un facteur de remplissage raisonnable, est devenue possible grâce à la diminution des règles de dessin de la technologie CMOS.

Le concept APS a évolué, aujourd'hui vers une idée plus riche qu'une simple adaptation d'impédance. Cette idée consiste à réaliser la conversion des photocharges au niveau du photosite, ce qui revient à intégrer à la structure de chaque pixel de la matrice CMOS, un dispositif de lecture qui peut être semblable à celui présent dans les matrices CCDs. La conversion sur le lieu du pixel des photocharges en tension permet d'imaginer des structures de pixel variées pour concevoir une rétine CMOS APS. Ainsi, la structure du pixel APS peut être définie suivant les contraintes d'une application spécifique. Les principales structures de pixels utilisées actuellement sont le pixel photodiode en mode courant, le pixel photodiode en mode intégration ou

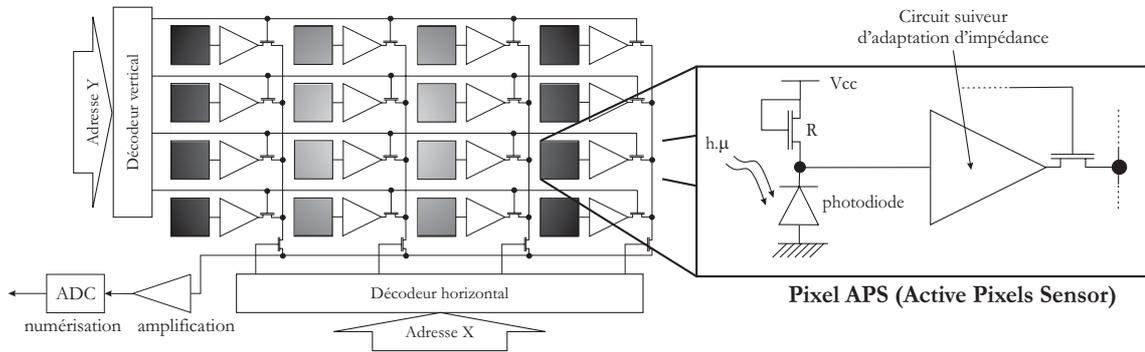


FIG. 2.10 – Structure d'un imageur CMOS APS.

encore le pixel photoMOS. Les circuits de ces pixels sont détaillés ci-dessous.

2.1.3.3 Le pixel photodiode en mode courant

Ce mode de lecture est basé sur la mesure du photocourant généré par le déplacement des photocharges. L'avantage majeur de ce mode lecture est sa large dynamique de mesure [52, 56, 51]. La réponse en courant du photodétecteur est linéaire sur une gamme d'intensité lumineuse étendue. Les capteurs, basé sur ce type de pixels, sont capable de mesurer aisément des variations d'intensité lumineuse de plus de 120 dB (10^6) soit à peu près la différence de luminosité entre le jour et la nuit. Pour exploiter cette dynamique de mesure, le photocourant circule dans la diode est converti en tension à partir de la structure de conversion de la figure 2.11.

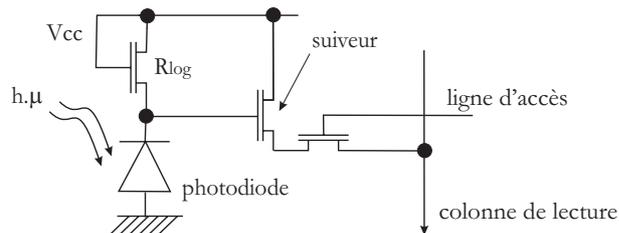


FIG. 2.11 – Pixel photodiode en mode courant.

Cette structure de lecture utilise un transistor de charge R_{log} (en faible inversion) qui admet une caractéristique courant/tension logarithmique. Et, le transistor suiveur réalise l'adaptation d'impédance entre le bus de lecture et le photodétecteur.

La réponse quantitative de ce pixel est de la forme :

$$V = V_{cc} - \frac{kT}{q} \ln \left(\frac{I_p}{I_0} \right)$$

où V_{cc} est la tension d'alimentation. I_p est le courant photonique proportionnel au flux lumineux. I_0 est le courant de forte inversion, $\frac{kT}{q}$ est une variable proportionnelle à la température. V est la tension de sortie du pixel.

Ce pixel présente un bon taux de remplissage (40 à 60 %) car il ne comporte que trois transistors. Par conséquent, la surface photosensible du pixel étant relativement élevée, il possède une bonne sensibilité. Le principal défaut de ce type de pixel est due à la disparité des transistors de charge R_{log} et du circuit suiveur. La précision de dessin du processus de fabrication du capteur induit une disparité des caractéristiques des pixels. Les différences de tensions de seuil V_{th} des transistors de charge R_{log} induisent une caractéristique de réponse distincte pour chaque pixel. Ainsi ce type de capteur présente un bruit spatial fixe (*Fixed Pattern Noise*) important qui nécessite une correction. De plus, le terme $\frac{kT}{q}$ engendre une réponse sensible aux variations de température. En conclusion, les imageurs, basée sur ce type de structure, sont relativement bruités et donc ne sont pas adaptés à des applications de radiométrie. Cependant, la caractéristique logarithmique de ce pixel lui permet de fonctionner dans des conditions d'éclairément très variées. Cette propriété reste très intéressante pour des applications de vision environnement réel. De plus, des travaux [57] ont été réalisé pour résoudre efficacement le problème du bruit spatial fixe.

2.1.3.4 Le pixel photodiode en mode intégration

La mesure de l'éclairément, à partir d'une photodiode en mode intégration ou précharge, est réalisée à l'aide de la capacité interne à la diode. La structure de conversion de détection de ce type de pixel est décrite par la figure 2.12.

Cette structure comporte une photodiode polarisée en inverse, un transistor fonctionnant en interrupteur commandé par un signal *reset* et d'un transistor suiveur qui réalise l'adaptation d'impédance entre le bus de lecture et le photodétecteur. La mesure est réalisée par la méthode d'échantillonnage corrélé décrite précédemment (fig. 2.7) en utilisant la photodiode comme diode de lecture.

La présence du bus *reset* dans la structure de ce pixel réduit son taux de remplis-

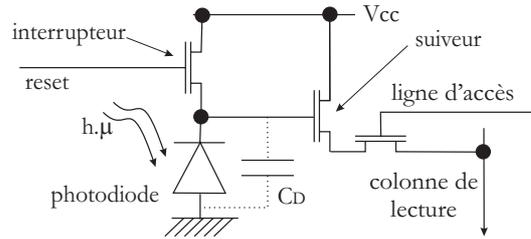


FIG. 2.12 – *Pixel photodiode en mode intégration.*

sage et sa conception nécessite une technologie plus fine. Sa sensibilité est inférieure au pixel photodiode en mode photocourant. De plus, pour ce type de pixel, le temps de mesure de la quantité de charges photoniques dépend du temps d'intégration alors qu'il est quasi instantané en mode photocourant. Les principaux avantages des imageurs, basés sur ce type de pixel, sont d'une part son rapport signal sur bruit élevé et d'autre part sa réponse linéaire qui rend ces capteurs intéressants pour des applications de radiographie.

2.1.3.5 Le pixel photoMOS

Le pixel photoMOS est identique à la structure de lecture des imageurs CCD, cette structure, illustrée figure 2.13, est du même type que celui étudié précédemment. Seule, la photodiode du schéma précédent est remplacée par une diode de lecture D_{lect} liée à un détecteur photoMOS par un transistor d'isolement. Le transistor *Inter* permet d'effectuer une mesure par intégration de la quantité de charge présente dans la diode de lecture D_{lect} . Comme pour les pixels précédents, le transistor *suiveur* assure l'adaptation d'impédance vis à vis du bus de lecture. Le détecteur photoMOS est conçu à partir d'une photogrille (ϕ) de grande taille qui définit la zone d'accumulation des photocharges. Le transistor d'isolement permet de créer une barrière de potentiel entre la zone de charges d'espace du détecteur photoMOS et celle de la diode de lecture D_{lect} . La particularité de ce pixel est le transfert des charges accumulées par le détecteur photoMOS vers la diode de lecture D_{lect} . Lors de la phase d'intégration, la diode de lecture est rendue flottante par le transistor *inter*. Le potentiel ϕ est porté à zéro ainsi les photocharges sont évacuées vers la diode de lecture D_{lect} à travers la zone de charges d'espace du transistor d'isolement. Ce transfert se traduit par la décharge de la capacité C_D de la diode lecture. La mé-

thode d'échantillonnage corrélé permet d'estimer la variation ΔV de tension aux bornes de la diode de lecture qui est proportionnelle à la quantité de photocharges accumulées N_e :

$$\Delta V = \frac{q}{C_D} \cdot N_e$$

ΔV est la variation de tension de la diode lecture, q est la charge de l'électron, C_D la capacité de la diode de lecture et N_e le nombre de photocharges accumulées par le détecteur photoMOS.

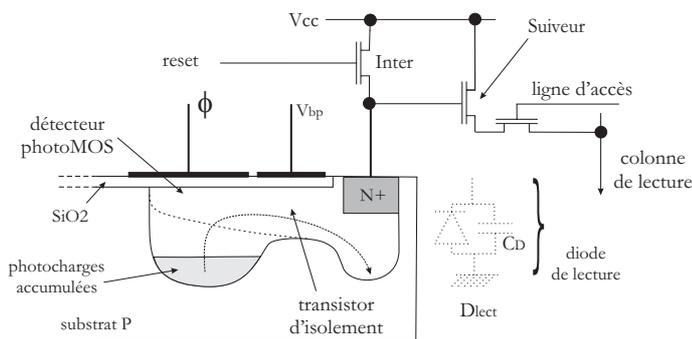


FIG. 2.13 – *Pixel photoMOS*.

Le nombre de lignes de commandes de ce pixel est important ce qui réduit son taux de remplissage et sa sensibilité. L'avantage de ce type de pixel est son niveau de bruit très faible, par rapport aux structures utilisant des photodiodes. La diode de lecture est conçue pour présenter une capacité plus faible que celle d'une photodiode. Cette faible valeur de capacité permet d'élever le facteur de conversion $\frac{q}{C_D}$ et ainsi améliorer le rapport signal sur bruit de la réponse du pixel.

2.1.4 Comparaison technologique

Une comparaison des technologies d'imagerie CMOS et CCD est proposée dans ce chapitre afin d'évaluer conjointement les avantages et les inconvénients de chacun [58, 43, 44]. La technologie CMOS présente plusieurs intérêts provenant d'une part de sa structure et d'autre part de sa compatibilité avec les procédés de fabrication CMOS. Les avantages liés à la structure des imageurs CMOS sont les suivants :

- La structure d'un imageur CMOS permet un **accès aléatoire** aux pixels qui autorise la mise oeuvre de techniques de **fenêtrage directement sur**

l'imageur. L'accès au pixel des matrices CCD, rend difficile la capacité d'accès rapide à des fenêtres d'intérêt dans l'image. Pour sélectionner des zones d'intérêt, cette technologie nécessite un transfert préalable de l'image entière dans une mémoire qui permet ensuite de sélectionner des régions d'intérêt. Cette technique de fenêtrage présente deux inconvénients majeurs par rapport à l'accès direct au pixel autorisé par la technologie CMOS. Premièrement, le transfert préalable de l'image inclut un retard à la source du système de perception. Ce retard est très pénalisant pour les algorithmes de suivi qui sont tenus de réagir promptement à l'évolution temporelle des données visuelles. De plus l'accès sériel à l'information visuelle des imageurs CCD engendre une cadence d'acquisition image fixe. Ainsi, la cadence d'acquisition d'une zone d'intérêt est identique à la cadence d'acquisition image. Par opposition, la cadence d'acquisition des imageurs CMOS dépend du nombre de pixels considérés et augmente grandement l'intérêt du fenêtrage. Pour cette raison, la vitesse d'acquisition des imageurs CCD s'exprime en images par seconde alors que celle des imageurs CMOS s'évalue en pixels par seconde.

- La structure d'un imageur CMOS permet d'utiliser une structure de pixel de type **photodiode en mode courant**. Ce pixel autorise une mesure non destructive de l'information visuelle qui permet une vitesse de lecture élevée. De plus, la caractéristique logarithmique de ce pixel lui permet de bénéficier d'une grande dynamique de mesure (~ 120 dB). Cet atout est intéressant pour travailler dans un environnement réel qui présente de grandes variations de luminosité (par exemple le secteur automobile).
- Contrairement au pixel CCD, les pixels CMOS sont isolés électriquement et spatialement les un par rapport aux autres. Les mouvements de charges entre les pixels sont fortement réduits et les capteurs CMOS sont bien moins sensibles au **phénomène de "blooming"**. D'autre part, la présence d'un pixel défectueux sur la matrice n'inhibe pas toute une colonne comme sur un imageur CCD.

En plus des avantages liés à leur structure interne, la compatibilité de cette technologie d'imagerie CMOS avec les procédés de fabrication standard de la micro-électronique lui offre d'autres atouts :

- La compatibilité de l'imagerie CMOS avec les procédés de fabrication VLSI

a permis d'envisager une technologie d'imagerie à **faible coût** qui présente un grand intérêt commercial. Par opposition, la technologie CCD est basée sur un dispositif capacitif qui assure le transport de charges au sein de la matrice photosensible. La réalisation de ce système a conduit à développer une technologie de conception spécifique aux capteurs CCDs qui n'est pas compatible avec les technologies standard de la microélectronique. Cette particularité fait de l'imagerie CCD une technologie bien plus coûteuse que l'imagerie CMOS.

- Le dispositif CCD requiert un mécanisme qui génère plusieurs horloges de niveaux de tensions élevés. Cette particularité indispensable à un transfert des charges efficace conduit à une forte consommation. Les imageurs CMOS réalisés selon le procédé de fabrication standard CMOS présentent une **faible consommation**.
- Les procédés de fabrication développés pour les circuits numériques permet à l'imagerie CMOS de bénéficier des l'augmentations régulières de la **finesse de cette technologie** ($0,12 \mu\text{ m}$ pour certaines rétines). La diminution constante des règles de lithographie permet la réalisation de matrices de grande taille et de pixels de plus en plus complexes.
- Enfin, la compatibilité technologique offre la possibilité d'intégrer au sein du même circuit une matrice photosensible et des modules de contrôle, de traitement et de communication. Ainsi, cette technologie permet de développer des composants dédiés (ASICs) qui possèdent des fonctionnalités supérieures à un simple imageurs. Les dispositifs de gestion de d'adressage et un moyen de communication standard peuvent être intégrés pour faciliter l'exploitation du composant. De plus, des dispositifs de perception visuelle originale, peuvent être développés en exploitant la proximité directe du pixel. Ces composants qui intègrent un dispositif de vision sont appelés des **rétines artificielles**. Les possibilités qu'offrent les rétines artificielles sont discutées dans le paragraphe 2.2.

Cependant, les imageurs CMOS présentent plusieurs inconvénients par rapport aux imageurs CCD. Les inconvénients majeurs sont les suivants :

- La capacité de **contrôler l'exposition** de la matrice photosensible est implémentée de façon standard dans les dispositifs CCDs sans affecter

considérablement le taux de remplissage. Ce contrôle de l'exposition est prépondérant pour les applications liées à la métrologie dynamique. Introduire un "shutter" électronique uniforme dans un imageur CMOS nécessite plusieurs transistors supplémentaires au sein de chaque pixel. Ces transistors supplémentaires réduisent le facteur de remplissage du capteur et augmente sa superficie, donc son prix. Généralement, les matrices CMOS sont équipées d'un "shutter" non uniforme ("Rolling Shutter") qui exige un nombre de transistors supplémentaires moins élevé. Ce type de dispositif contrôle l'exposition des lignes du capteur. Chaque ligne de la matrice est exposée à différent instant. Ce dispositif est suffisant pour une grande majorité d'application, mais il peut engendrer des distorsions de l'image lors de l'observation de mouvements rapides.

- Le procédé de fabrication CMOS fait que **le courant d'obscurité** est plus important dans les imageurs CMOS que dans imageurs CCD. Ce courant est dû aux générations de charges thermiques dans le pixel. Il provient des électrons qui se sont libérés des atomes de silicium grâce à leur énergie thermique. Ces électrons créent des charges qui s'ajoutent aux charges photoniques représentatives de l'éclairage. Le nombre de ces charges thermiques varie beaucoup d'un pixel à l'autre ce qui engendre une non uniformité du courant thermique de chaque pixel ("*Dark Signal Non Uniformity*"). Ce courant d'obscurité est très dépendant de la température et du procédé de fabrication.
- La structure CMOS est également à l'origine d'un inconvénient lié à la dispersion des caractéristiques individuelles des éléments actifs de la matrice. Le **bruit spatial fixe** exprime l'écart de réponse des éléments des circuits de lecture situés dans les différents pixels et les différentes colonnes d'une matrice APS (fig. 2.14). Il est dû principalement aux non uniformités des paramètres des transistors (tensions de seuil, transconductances) qui apparaissent lors de la fabrication.

Le développement de la technologie d'imagerie CMOS est lié à l'origine à des objectifs commerciaux. La compatibilité de cette technologie d'imagerie avec la technologie VLSI de l'électronique a permis d'envisager une technologie d'imagerie à faible coût qui suscite un grand intérêt commercial. Ces perspectives de profit ont permis une évolution rapide de cette technologie depuis l'apparition du concept APS. Ac-



FIG. 2.14 – (a) - Bruit spatial fixe colonne important du à la présence d’amplificateurs colonnes (Fuga 1000 - Fill Factory), (b) - Image sans correction du bruit spatial fixe (NC1802 - Neuricam), (c) - Image après correction du bruit spatial fixe (NC1802 - Neuricam).

tuellement, l'évolution de la technologie d'imagerie CMOS est motivée parallèlement par le développement d'une technologie à moindre coût et d'autre part à dessein de performance technologique. De nombreux efforts sont fournis pour améliorer la miniaturisation, la consommation et la qualité d'images de cette technologie. En effet, les imageurs CMOS sont de bons candidats pour les applications spatiales grâce à leur résistance aux radiations. De plus, leur faible consommation [59], par rapport à la technologie CCD, suscite de vifs intérêts pour toute application embarquée (applications spatiales, téléphones et ordinateurs portables). L'engouement due à de multiples perspectives d'applications ont suscité de nombreux travaux de recherche qui améliorent constamment la qualité d'image [60], la rapidité [48, 50, 49] et la résolution des imageurs CMOS. L'imagerie CCD et l'imagerie CMOS évoluent toutes deux vers des performances de plus en plus élevées, motivées par un nombre d'applications croissants. Cependant, le devenir de ces technologies concurrentes sera vraisemblablement dicté par les lois du marché.

2.1.5 Conclusion sur le choix d'un imageur

L'objet de cette thèse est la conception d'une plate-forme de recherche destinée à expérimenter des systèmes de vision bas niveau basés sur le concept de vision active. La technologie d'imagerie présente plusieurs atouts pour développer un système de vision. La structure à accès aléatoire des imageurs CMOS leur confèrent une pro-

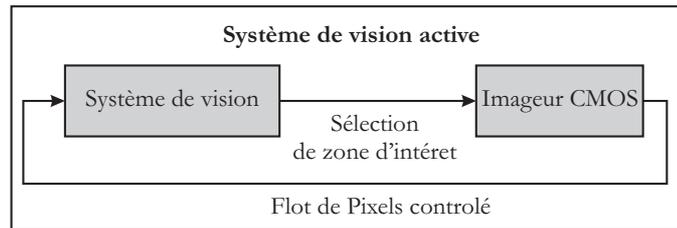


FIG. 2.15 – *Le contrôle CMOS permet de définir un système de vision active qui exploite l'échantillonnage du signal image.*

priété essentielle pour introduire l'imageur dans un système de vision active. Les imageurs CMOS fonctionnent comme une mémoire qui possède un bus d'adressage et de lecture. Ils se comportent donc comme un périphérique esclave vis à vis du système de vision. Suivant les commandes d'acquisition générées par celui-ci, l'imageur CMOS produit un flot de pixels contrôlé. L'accès aléatoire au pixel permet un échantillonnage contrôlé de l'espace photosensible. Cette différence devient déterminante lorsque l'on désire exploiter l'imageur dans une boucle de vision active. La possibilité de contrôle de l'acquisition offerte par les imageurs CMOS permet d'instaurer une rétroaction sur la commande d'adressage du capteur à partir du flot d'informations visuelles (fig. 2.15). Ce type de rétroaction constitue un système vision active originale qui représente un axe de recherche fondamental de cette thèse.

2.2 Les rétines artificielles

Le nom associé à cette catégorie de composants provient du fait que les premiers circuits réalisés s'inspiraient d'observations neurobiologiques. Une rétine est un photodétecteur dans lequel cohabite des dispositifs de phototransduction et de traitement d'images. L'intégration d'élément photodétecteur et de circuits de traitement sur la même puce (fig. 2.16), pour obtenir de meilleures performances ou pour rendre le détecteur et le système de traitement plus compact n'est pas une idée nouvelle. La compatibilité de la technologie d'imagerie CMOS fait de cette technologie, la technologie par excellence pour concevoir des rétines de vision. Les rétines artificielles sont par nature des composants analogiques.

Les rétines artificielles sont conçues suivant deux objectifs distincts. Dans un

premier cas, elles sont complètement dédiées à une application spécifique, et donc leur architecture est entièrement déterminée et optimisée par rapport aux contraintes d'une unique application [61]. Dans un deuxième cas, elles sont développées en considérant que l'ensemble des systèmes de vision partagent le même type de prétraitements qui ne nécessitent pas d'être reprogrammés. Ce type de rétines intègre, de manière figée, des prétraitements standards des systèmes de vision (filtrage, adaptation du contraste, ou encore détection de bord ou de mouvement) dans le but d'améliorer la vitesse de calcul et de décharger le système de vision hôte.

2.2.1 Architectures des rétines artificielles

Les architectures de traitements implantées dans les rétines sont contraintes par l'organisation élémentaire de la structure de phototransduction. Les rétines basées sur la technologie CMOS sont susceptibles d'intégrer quatre types d'architecture de traitements le long du trajet du flot d'informations visuelles:

- au sein du pixel,
- au sein des colonnes,
- auprès de la matrice.

La délocalisation au niveau des pixels (fig. 2.16 - a) consiste à insérer une unité de traitements (UT) entre l'élément photosensible (PD) et le transistor de sélection ligne. Cette approche est intéressante pour améliorer le rapport signal sur bruit, tout en assurant une faible consommation du composant et donne la possibilité d'adapter la capture de l'image en contrôlant le processus d'intégration du photodétecteur. Cependant, l'exploitation de cette structure est restreinte par la dimension des pixels ce qui se traduit par une baisse du facteur de remplissage. Ainsi, le nombre de transistors qui peuvent être intégrés au sein du pixel devient critique. Même, si les technologies de lithographie évoluent continuellement, la complexité des circuits de traitements intégrés dans ce type de structure est actuellement fortement limitée.

La délocalisation au niveau du flot image (fig. 2.16 - c) est la structure la plus intuitive. Elle offre une grande liberté de conception car le circuit de traitement est spatialement indépendant de la matrice photosensible. Elle n'affecte pas le taux de remplissage des pixels et donc ce type de rétine présente une photosensibilité optimale. Cependant, l'architecture de traitement ne présente plus un fort degré de

parallélisme. En d'autres termes, cette structure est limitée par un goulot d'étranglement entre le capteur et l'unité de traitements. La nécessité d'une cadence de fonctionnement élevée de l'unité de traitements entraîne une surconsommation. En conclusion, cette structure présente peu d'intérêt par rapport à une solution embarquée équivalente.

La délocalisation au niveau des colonnes de lecture (fig. 2.16 - b) est la solution intermédiaire aux deux précédentes. Ici, Le dispositif de traitement est inséré sur chaque colonne de l'imageur avant le transistor de lecture colonne. Dans ce type de structure, l'image est lue colonne par colonne. Des mécanismes de mémorisation analogiques sont exploités pour lire et stocker les colonnes entières. Ensuite, les colonnes sont envoyées en série vers l'unité de traitement pixel par pixel. Cette structure permet de bénéficier d'une architecture de traitement parallèle qui autorise des fréquences de travail basses et donc une faible consommation. De plus, le facteur de remplissage des pixels est moins affaibli que dans la première solution. Ainsi, la photosensibilité du capteur se trouve moins détériorée.

Une dernière structure consiste à exploiter une mémoire de travail. L'image est transférée vers une mémoire analogique située à coté de la matrice photosensible. Ensuite les traitements sont effectués sur cette matrice de mémorisation afin d'exploiter des cellules de calcul qui ne sont pas contraintes par l'espace photosensible. Cette structure introduit par principe un retard dans la chaîne d'acquisition. De plus, elles nécessitent une grande surface de silicium qui entraîne un coût de fabrication important et forte consommation d'énergie.

Le choix de l'une de ces architectures fonctionnelles définit le mode d'interaction entre les photodétecteurs et les unités de traitements. Le choix de la meilleure structure dépend principalement de la nature de l'algorithme implémenté. Les différents algorithmes sont différenciés suivant leurs contraintes spatiales, temporelles et les interconnexions qu'ils nécessitent.

2.2.2 Classification des rétines

La première classe de traitement, la plus simple, est celle qui ne présente pas de contraintes spatio-temporelles. Le résultat de l'opération effectuée sur un pixel à un instant donné ne dépend que de la valeur de ce pixel au même instant. L'exemple

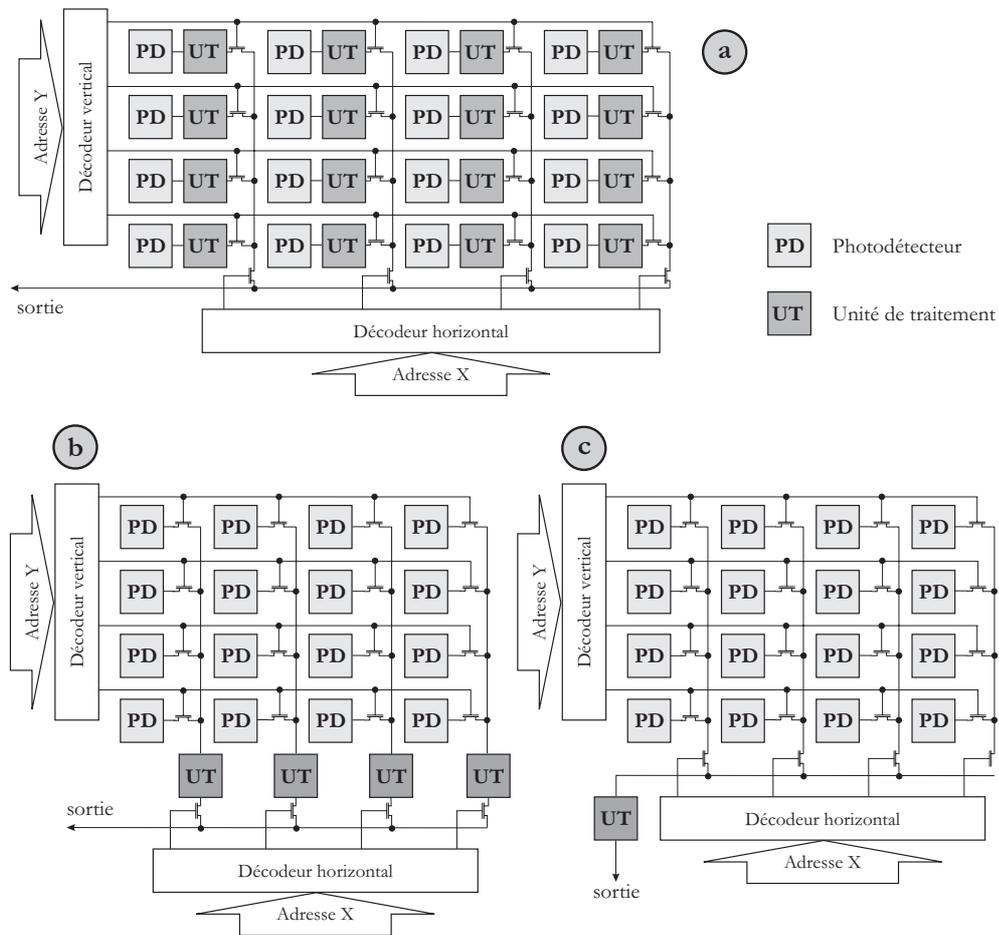


FIG. 2.16 – *Délocalisation des traitements au sein du capteur d'image: les rétines artificielles.*

classique de ce type de traitement est la conversion analogique digital du signal. Il peut être, en principe, implémenté sur l'ensemble des architectures fonctionnelles décrites précédemment. Cependant, la complexité des fonctions implantées, la préservation du facteur de remplissage et l'économie de l'espace de silicium conduit à utiliser une architecture traitement délocalisée au niveau du flot image (2.16 - c).

La seconde classe de traitement présente uniquement des contraintes spatiales. Le résultat de l'opération effectuée sur un pixel à un instant donné dépend de la valeur d'un ensemble de pixels à cet instant. Ce type de traitement nécessite des interconnexions entre les pixels généralement voisins. La délocalisation au niveau des pixels (fig. 2.16 - a) semble la plus indiquée. Mais, les limites technologiques actuelles

réduisent le nombre d'interconnexions envisageables à un peu de voisins. Ainsi, le traitement de plusieurs pixels impose une implantation au niveau des colonnes ou au niveau du flot image (fig. 2.16 - a) et nécessite l'emploi de cellules de mémorisation.

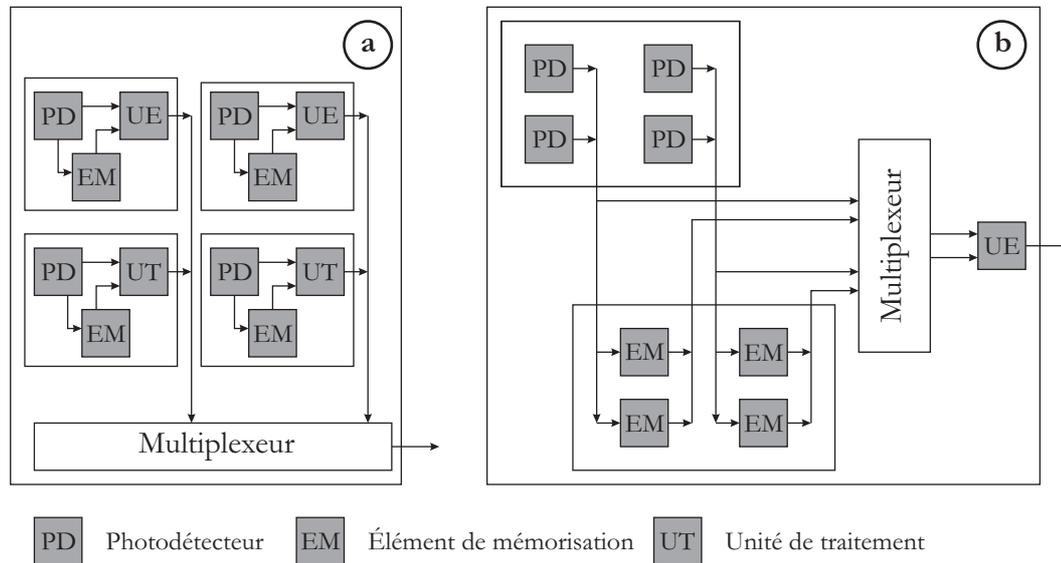


FIG. 2.17 – Introduction d'éléments de mémorisation pour effectuer un traitement spatio-temporel: a - mémorisation au sein du pixel, b - mémorisation au sein d'une matrice indépendante.

La dernière classe de traitement présente à la fois des contraintes spatiales et temporelles. Le résultat de l'opération effectuée sur un pixel à un instant donné dépend de la valeur d'un ensemble de pixels à des instants différents. Par définition, ce type de traitement nécessite des cellules de mémorisation. Ces cellules de mémorisation sont implantées soit au niveau du pixel (fig. 2.16 - a), soit au sein d'une matrice de mémorisation indépendante (fig. 2.16 - b). Les interconnexions, comme précédemment, et l'introduction de cellules de mémorisation sont limitées par la technologie d'intégration et ce type de rétines artificielles se limite pour l'instant à la mémorisation d'une seule image.

Ces composants s'illustrent dans toutes sortes de traitements bas niveau. L'ouvrage Moini [62] constitue un état de l'art relativement complet qui donne un aperçu de la richesse et la diversité des travaux de recherche sur les rétines artificielles. La segmentation [63, 64, 46], le calcul du flot optique [65, 66, 67], le rehaussement de

contraste [30] sont des applications largement traitées. Parmi les travaux les plus récents certaines rétines intègrent des systèmes de localisation et la reconnaissance de formes simples [68, 69, 70]. On note aussi l'apparition de composants de suivi rapide très performants [71, 72, 73]. Dans les travaux de [74] une rétine est utilisée dans un système de suivi binoculaire actif. Dans cette application la rétine n'est pas utilisée suivant un comportement actif. Par contre, Brajovic[75] propose une rétine de suivi qui adopte un comportement actif pour faire évoluer une zone d'acquisition sur la matrice photosensible de la rétine. Ce composant est capable de suivre l'image d'un objet qui se déplace sur l'espace photosensible à 7000 pixels/s. L'exploitation de l'accès aléatoire à l'information au sein d'un système de vision active lui permet de sélectionner dynamiquement une zone d'intérêt sur l'imageur.

2.2.3 Avantages et inconvénients des rétines artificielles

Le principal avantage de la délocalisation de traitements au sein d'une rétine de vision est bien évidemment l'accroissement de la puissance de calcul. La conception d'un circuit directement connecté à la source de d'information visuelle élimine le goulot d'étranglement toujours présent dans les dispositifs constitués d'un imageur relié à un processeur externe. Les rétines de vision peuvent être considérée comme une classe spécifique d'ASICs¹. De ce fait, les rétines présentent les mêmes avantages et inconvénients que ceux-ci. La conception de circuits spécifiques permet d'établir les structures de traitement de manière optimale et personnalisée (*custom*). De ce fait, les rétines permettent d'intégrer des traitements directement sur le signal analogique et d'exploiter des architectures parallèles dédiées. De plus, le système peut être spécifié complètement, jusqu'au dimensionnement individuel de chaque transistor. Ainsi ces composants permettent un accroissement considérable des vitesses de traitement. D'autre part, cette technologie permet de concevoir des systèmes qui consomment très peu d'énergie. Ainsi, l'implantation dans un composant unique d'algorithmes de traitement permet la réalisation d'un système compact. Ces propriétés sont essentielles pour toutes les applications industrielles et scientifiques embarquées où le gabarit du système et la puissance consommée détermine la faisabilité.

Bien que les rétines de vision soient susceptibles d'évoluer et de devenir une so-

1. ASIC - Application Specific Integrated Circuit

lution d'avenir pour concevoir une partie d'un système de vision, elles présentent actuellement des faiblesses importantes. La diminution du facteur de remplissage réduit les performances des photodétecteurs. Ce problème est aujourd'hui très contraignant en terme de qualité intrinsèque de l'information pixel. Même si l'évolution des technologies de lithographie permet d'espérer un amoindrissement de cet effet néfaste, il reste actuellement un problème considérable et la résolution des rétines artificielles est limitée par ce faible taux de remplissage. De plus, la précision de technologie VLSI analogique limite la précision des traitements implantés en introduisant une incertitude élevée sur les paramètres de calcul. D'autre part, la conception de rétine artificielle nécessite un temps développement important (*full custom designed*) tout en étant dédiée à une application spécifique. En effet, les rétines sont actuellement conçues pour une application ou une tâche donnée et n'autorise que peu de contrôle sur la tâche implantée. Ainsi, elles ne peuvent être adaptée aux besoins de divers dispositifs de vision.

En conclusion, la technologie des rétines artificielles n'est pas suffisamment mature pour être exploitée efficacement dans les systèmes de vision. De plus, le manque de flexibilité est un inconvénient rédhibitoire pour expérimenter l'approche de perception considérée dans cette thèse car elle s'appuie sur la flexibilité du dispositif de traitement. Cependant, cette technologie reste extrêmement prometteuse. La capacité d'intégrer un système auprès de l'imagerie devrait permettre le développement futur de systèmes d'imagerie multimédia performants et hautement miniaturisés.

2.3 Les caméras dites "intelligentes"

Les caméras intelligentes offrent une alternative intermédiaire pour concevoir un système de vision. Cette solution forme un compromis entre une implantation directe dans le silicium et une solution purement logicielle. Cette appellation ne fait pas forcément référence au domaine de l'intelligence artificielle. Ces systèmes embarqués sont appelés "caméra intelligente" dès lors qu'ils intègrent une centrale traitement.

2.3.1 Organisation fonctionnelle

Afin d'exploiter la proximité du photodétecteur et donc de bénéficier d'un lien privilégié avec l'information à traiter, une partie des traitements constituant le système de vision peuvent être délocalisés dans la caméra (fig. 2.1 - niveau 2). Ce type de dispositifs combine un imageur, une unité de traitement et une interface de communication pour définir un système embarqué. Contrairement aux rétines artificielles qui permettent d'intégrer des traitements directement sur les signaux analogiques représentatifs de la valeur de chaque pixel, la lecture de l'ensemble des pixels ne peut être parallélisé. La séparation de l'unité de traitement et du composant photosensible sur deux composants distincts limite le nombre de broches de connexions envisageables. Cette contrainte technologique impose généralement à ce type de dispositif un accès séquentiel à l'information visuelle. Ainsi, la délocalisation des traitements au niveau de la caméra induit un goulot d'étranglement du flot d'information à l'interface entre la matrice de photodétection et le calculateur (fig. 2.18). Cependant, ces systèmes de vision embarqués manipulent uniquement des signaux numériques qui permettent l'implantation d'algorithmes complexes et de calculs précis. De plus, ils s'appuient sur une méthodologie de conception modulaire qui permet une adaptation aisée du système à l'application envisagée.

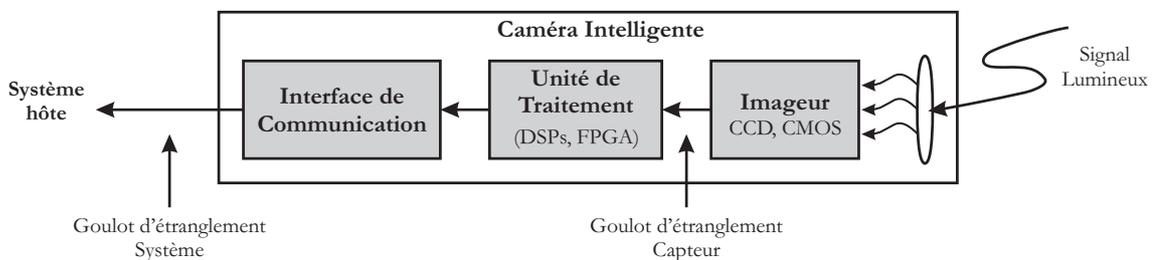


FIG. 2.18 – Structure générale d'une caméra "intelligente".

La plupart des systèmes de vision sont exploités à partir d'un système d'exploitation informatique hôte. Que le système de vision soit implémenté à partir d'une caméra "intelligente" ou d'une caméra classique, le dispositif nécessite un lien de communication entre le système informatique et le capteur. Ce lien représente la barrière de communication principale du dispositif. Les bus de communication standards tels que le IEEE1394, firewire, l'USB 2.0, ou encore le bus Camlink autorisent des bandes passantes tout fait satisfaisante pour transmettre le signal vidéo. Ce-

pendant, la plupart des systèmes d'exploitation utilisés n'intègre pas un mode de fonctionnement temps réel et ces moyens de communication standard nécessitent des protocoles contraignants. Cette interface entre le système informatique et le capteur limite la réactivité du système de vision sur le capteur et induit un goulot d'étranglement de l'information (fig. 2.18).

La délocalisation d'une partie du système, au sein de la caméra permet de réduire ce problème de communication. La proximité physique du capteur et de l'unité de traitement embarquée permet de bénéficier d'un lien de communication privilégié. En introduisant, en temps réel, une rétroaction de l'information visuelle sur le dispositif sensoriel, cette relation étroite entre le capteur et système embarqué permet de concevoir un capteur de vision active efficace. De plus, la délocalisation d'une unité permet d'effectuer des prétraitements avant la barrière de communication afin d'effectuer une étape de sélection d'informations à transmettre vers le système hôte. L'intégration d'un système "intelligent" au plus près du capteur permet donc d'optimiser l'utilisation de la bande passante du bus de communication.

La mise en œuvre d'une machine de perception visuelle impose le choix d'une architecture matérielle. L'éventail technologique actuel propose des solutions variées telles que les PCs standards, les stations de travail, les PCs embarqués, les réseaux de processeurs (machines parallèles), les processeurs dédiés au traitement du signal (DSP), la logique câblée programmable (FPGA) ou encore la logique câblée figée (ASIC). Ces solutions matérielles présentent chacune leurs propres avantages et limitations. Afin de bénéficier des avantages de chacun, des architectures mixtes peuvent être imaginées ce qui étend plus encore la gamme de solutions envisageables. Cependant, le choix technologique est fortement dépendant des spécificités d'une application. Par exemple, le coût qui est lié au volume de fabrication ou une exigence de confidentialité peut conduire à la solution ASIC. La nécessité d'un système embarqué qui impose des contraintes aux niveaux du dimensionnement et de la consommation s'oriente généralement vers les technologies FPGA, DSP ou les PC embarqués. La flexibilité ou le temps de développement oriente le choix vers des architectures préconçues standards telles que les PCs. Une exigence en puissance de calcul peut conduire à opter pour une machine parallèle ou une station de travail. Une fois l'éventail de solutions réduit par la spécificité de la tâche, le choix final est déterminé par l'adéquation entre l'algorithmie envisagée et la cible matérielle.

Les systèmes de vision traitent des volumes de données importants en imposant des contraintes temporelles élevées. L'adéquation avec la cible matérielle nécessite d'évaluer précisément la complexité du système, le parallélisme de tâches et les capacités de stockage nécessaires.

Les démarches de conception se distinguent par la méthodologie d'implantation. Une première approche consiste à adapter l'algorithme de manière logicielle à une architecture de traitement préconçue. Cette démarche aboutit à des solutions dites programmées. Par opposition, une seconde approche consiste à concevoir une architecture dédiée à un algorithme. Cette démarche inverse aboutit à des solutions dites câblées. La conception d'une caméra "intelligente" exige le choix d'un support physique afin d'implémenter les fonctionnalités du système. Cette décision initiale nécessite une étude approfondie car elle conditionne complètement l'adéquation avec les exigences de l'application et les méthodologies de développement. Dans cette partie, les divers cibles d'implantations envisageables sont énumérées suivant les une approche d'implantation "programmées" et "câblées".

2.3.2 Les solutions dites "programmées"

L'implantation de l'algorithme suivant une solution programmée consiste à développer un programme à l'aide d'un langage évolué ou en assembleur qui est exécuté sur un ordinateur standard, une station de travail. La solution programmée consiste à réaliser une adéquation algorithmique suivant les contraintes d'une architecture figée. Ces architectures de traitement préconçues exploitent des microprocesseurs CISC ou RISC. Les microprocesseurs CISC (*Complex Instruction Set Computer*) traitent des instructions de tailles, de longueurs variables et de structures complexes. Cette spécificité a pour but de minimiser l'occupation mémoire et les appels à cette mémoire. Le fonctionnement de la machine assez complexe et pousse les calculateurs (Pentium, PowerPC, Alpha) à extraire plusieurs instructions à l'avance de manière à anticiper sur le résultat du calcul en cours, tout cela pour ne pas ralentir la chaîne de traitement.

Par opposition, les ordinateurs munis de microprocesseurs RISC (*Reduced Instruction Set Computer*) utilisent un jeu d'instructions simple implanté sous forme de machine pipeline. Le concept de RISC (Sparc, MIPS) est basé sur une simplifica-

tion des architectures pour augmenter les cadences des traitements. Les opérations entre les registres s'effectuent en un seul cycle, ce qui permet un contrôle simple qui peut être définitivement câblé sans recourir au microcode. Les opérations qui nécessitent plusieurs cycles (virgule flottante par exemple) sont alors exécutées soit de façon logicielle, soit dans un coprocesseur de façon à ne pas ralentir le CPU. Parmi les processeurs RISC on trouve des processeurs dédiés au traitement du signal. Ces processeurs communément appelés DSP possèdent des structures monoprocesseurs ou multiprocesseurs, où les opérations standards de traitement du signal sont optimisées de façon matérielle. Ce type de solution est constituée d'une architecture matérielle qui peut être implantée sur une carte standard afin d'être intégré dans un environnement conventionnel (stations de travail, PC), soit sur une carte spécifique au système.

2.3.2.1 Les microprocesseurs dédiés au traitement du signal(DSP)

Les microprocesseurs actuels bénéficient des avancées continues de la technologie des semiconducteurs. Cette évolution a permis l'intégration d'un nombre toujours croissant de fonctionnalités aux sein de microprocesseur d'utilité générale. Malgré, l'augmentation régulière des performances de ces processeurs (station travail, PCs standard), ils restent souvent insuffisant pour implémenter efficacement des applications de traitements d'image toujours plus exigeantes. Le manque d'efficacité de ces processeurs standards provient de la grande diversité des opérations qu'ils sont susceptibles de réaliser. En conséquence, seulement une faible partie de leur architecture est utilisée pour réellement implémenter le système de vision. Par exemple, un processeur conçu pour réaliser une opération flottante utilise qu'une partie infime de ces circuits pour réaliser une opération logique élémentaire. L'architecture classique des microprocesseurs n'étant pas optimale pour le traitement d'images, les processeurs utilisés dans la conception de caméra intelligente [76] sont le plus souvent des processeurs possédant une architecture dédiée aux traitement du signal (DSP²). Ces processeurs possèdent une unité arithmétique et logique qui permet de réaliser des opérations spécifiques telles que la multiplication accumulation de manière très rapide grâce à une architecture dédiée. De plus, ces composants sont en constante évolution, et atteignent des fréquences de traitement voisines du Gigahertz. Ces

2. Digital Signal Processors

processeurs incorporent des dispositifs de prédiction de boucles, plusieurs niveaux mémoires caches, des noyaux temps réel qui leurs permettent d'accroître leur niveau de performance. Cependant, malgré l'effort d'adéquation architecturale que représente ces composants, ils demeurent une architecture figée de nature séquentielle. Cette architecture séquentielle est exploitable grâce à un nombre d'instructions fixes et prédéterminées mais ne permet pas une adéquation matérielle complète suivant un algorithme donné. Comme toute solution programmée, elle revient alors à réaliser une adéquation algorithmique suivant les contraintes d'une architecture figée.

2.3.2.2 Les machines parallèles

Les architectures classiques imposent un traitement séquentiel des données qui réduit leur efficacité pour implémenter des algorithmes comportant un fort degré de parallélisme. Aussi, les applications de vision artificielle exigent bien souvent une puissance de calcul importante qui nécessite une architecture matérielle parallèle pour satisfaire des contraintes temporelles élevées. Une solution consiste à utiliser plusieurs processeurs et diviser le problème à résoudre en plusieurs sous problèmes qui sont traités en parallèle. Les performances de ce type d'architectures sont tributaires de la vitesse des communications entre les processeurs, qui dépend, entre autres paramètres, de la topologie du réseau d'interconnexion choisie lors de la construction de la machine. Malgré, les nombreuses recherches et avancées réalisées sur la conception et la programmation de ces machines parallèles, cette solution d'implémentation pose de réelles difficultés de partitionnement et de communication entre les processeurs. De plus, malgré les recherches sur les langages de programmation [77, 78] qui permettent d'exploiter ces machines, une implémentation algorithmique efficace nécessite l'intervention de spécialistes. Cette approche est basée sur une architecture de traitement générique destinée à implémenter des systèmes de vision variés. Pour mettre en œuvre un système disposant d'une capacité de traitement élevée, une solution antagoniste consiste à réaliser une adéquation architecturale à partir des spécificités algorithmiques d'une application donnée. Le caractère générique des processeurs basé sur une séquence de lecture, de décodage et d'exécution d'instructions réduit l'efficacité réelle du calculateur. Ainsi, le développement d'une architecture dédiée, sans soucis de généricité, permet de bénéficier à la fois d'une implémentation parallèle et d'une efficacité optimale du calculateur.

2.3.3 Les solutions dites "câblées"

Par opposition aux solutions précédentes, les solutions câblées s'orientent vers la conception d'une architecture matérielle dédiée à un algorithme. Par rapport aux solutions programmées, les solutions câblées permettent des cadences de traitement élevées grâce à une adaptation spécifique de l'électronique et la mise en oeuvre d'architectures parallèles optimisées. Parmi, ces solutions câblées, on distingue les architectures matérielles figées (ASIC³) et les architectures matérielles programmables (FPGA⁴).

2.3.3.1 Les architectures matérielles complètement dédiées

Une architecture matérielle figée permet une adéquation de l'électronique jusqu'au dimensionnement spécifique de chaque transistor dans le silicium. Ainsi cette solution permet d'atteindre des performances bien supérieures à celles des architectures matérielles programmables ou celles des solutions programmées. Dans le cas la vision, les rétines artificielles peuvent constituer une solution câblée qui met en oeuvre un parallélisme massif en introduisant une cellule de traitement analogique auprès de chaque pixel. Les structures et les capacités de ce type de composant ont été abordées précédemment 2.2. L'aspect attractif de ces solutions complètement dédiées est modéré. La conception d'un tel système nécessite une spécification très fine qui nécessite un temps de développement important malgré des outils de CAO de plus en plus performants. De plus, ces solutions sont rigides et onéreuses (mis à part le cas d'une production massive). L'algorithme implémenté est figé dans le silicium et généralement il est peu ou pas paramétrable, ce qui nuit à la flexibilité du dispositif. Cependant, ces circuits semblent une solution d'avenir pour intégrer avec de hautes performances temporelles des traitements d'images bas niveaux tels que la détection de contours ou le calcul flot optique.

3. Application Specific Integrated Circuit

4. Field Programmable Gate Array

2.3.3.2 Les architectures matérielles programmables

Par opposition aux ASICs, les architectures matérielles programmables (FPGA) permettent une adéquation électronique moindre, mais autorisent une plus grande souplesse de développement. Grâce, à un dispositif de reconfiguration, équivalente à un bloc de transfert mémoire, il est possible de changer l'architecture fonctionnelle implantée sans modifier l'architecture matérielle. Les capacités d'intégration actuelles favorisent particulièrement le développement des circuits programmables. Ils représentent, aujourd'hui, une alternative aux circuits dédiés, dès lors que les volumes de production sont faibles, ou que la reconfigurabilité et le temps de conception font parti des critères de décision. L'émergence de nouvelles générations de circuits programmables, qui offrent des densités de plusieurs dizaines de millions de portes logiques, permettent même d'intégrer des coeurs de processeurs. Ces composants permettent de bénéficier de la rapidité des architectures câblées et d'une souplesse qui était initialement l'apanage des architectures programmées. Ils représentent une solution prometteuse pour concevoir un système de vision embarqué [79, 80].

2.3.3.3 Les solutions mixtes

L'ensemble de ces méthodes d'implantation représente différents niveaux de compromis entre la performance temporelle et la flexibilité. L'adéquation algorithmique et matérielle étant fortement dépendante des contraintes d'une application, des solutions mixtes existent afin de réaliser un compromis entre ses méthodes d'implantation. L'objectif est de bénéficier à la fois des avantages des solutions programmées et des solutions câblées. Généralement, ces dispositifs intègrent d'une part un processeur dédié au traitement du signal afin de bénéficier d'une puissance de calcul séquentiel élevée et de la flexibilité des solutions programmées et d'autre part, ils intègrent conjointement une architecture matérielle programmable qui permet une adéquation de l'architecture fonctionnelle du système. Ces architectures mixtes aboutissent à des structures hétérogènes qui nécessitent des méthodes et des outils de conception variés. Elle fait apparaître la question essentielle du partitionnement algorithmique. Lorsque cette tâche pourra être automatisée [81, 82], les solutions mixtes devraient rapidement s'imposer.

La conception d'une caméra "intelligente" exige le choix d'un support physique afin d'implémenter les fonctionnalités du système. Cette décision initiale nécessite une étude approfondie car elle conditionne complètement l'adéquation avec les exigences de l'application et les méthodologies de développement.

2.3.4 Les méthodologies de conception

Une grande variété de méthodologies de conception sont répertoriées. Pour ne citer que les plus courantes, la méthode descendante "top-down", la synthèse comportementale, les méthodes de conception modulaire et des méthodes d'adéquation algorithme architecture. La première tâche du concepteur est de choisir une cible d'implantation parmi celles citées précédemment. Suivant les spécificités de l'application ce choix se pose comme une évidence ou est réalisé de façon intuitive. Des besoins indépendants des solutions algorithmiques envisagées tels que la confidentialité, le temps de développement, les besoins de flexibilité, l'embarquabilité sont des critères qui permettent de s'orienter vers une solution particulière ou contraire d'éliminer une solution. Mis à part ces critères, le choix d'une cible d'implantation nécessite de confronter les contraintes de l'application avec les solutions algorithmiques envisagées. Le parallélisme de tâche, les spécifications temporelles doivent être identifiés pour faire ressortir l'adéquation matérielle d'une cible d'implantation avec l'algorithmie. Dans le cas d'applications génériques des cibles conjointes logiciel/matériel sont souvent envisagées en particulier pour implanter des systèmes complexes.

Afin de concevoir des circuits toujours plus complexes, les designers proposent d'utiliser des approches méthodiques pour maîtriser le flot de développement. Une méthodologie consiste à exploiter une "boite à outils" dans laquelle le concepteur trouve une variété d'outils : modèles, solutions, méthodes. Le rôle du concepteur est de trouver pour chaque situation l'outil approprié pour une résolution efficace de son problème. Cette approche est très utilisée parmi les concepteurs de circuits intégrés ASICs et FPGAs. Elle permet la simulation comportementale du système pour assurer la validité fonctionnelle du système tout au long du cycle de conception. Les méthodes de développement des circuits FPGAs et ASICs ont beaucoup profité des récentes avancées de la microélectronique. Ainsi lors de la phase de conception,

l'utilisation de langages de description HDL⁵ se généralise. Hormis le fait que ces langages présentent un aspect relativement convivial, ils aident au développement d'un ensemble de couches d'abstraction du circuit ainsi qu'à la division du flot de conception en sous problèmes. De tels niveaux d'abstraction existent en conception logicielle. Le niveau le plus bas est le langage microcode, puis le langage assembleur, le langage C ou C++. Cette méthode de conception hiérarchique descendante s'applique aux circuits ASICs et FPGAs. Généralement, une décomposition fonctionnelle du système est réalisée pour définir un ensemble de circuits qui sont eux composés de sous circuits. Finalement, cette décomposition s'étend jusqu'à un schéma électronique composé de transistors.

Par opposition, à cette méthode qui consiste à décorréliser les aspects architecturaux et algorithmiques, la méthode d'adéquation algorithme architecture consiste à considérer conjointement ces deux aspects de l'implantation pour aboutir à une solution optimale. Cette méthodologie prenant en compte leurs interactions, en vue d'effectuer une implantation optimisée de l'algorithme tout en réduisant les temps de développement et les coûts de l'application étudiée. L'approche consiste à adapter la structure de l'algorithme et de l'architecture pour satisfaire les contraintes d'une application.

2.4 Choix technologiques

Durant ces dernières années, le développement des circuits programmables à haute densité d'intégration a modifié considérablement les méthodologies de conception des architectures de traitement numérique. Ces composants représentent une solution d'intégration à moindre coût suivant un temps de développement réduit. Actuellement, cette technologie est suffisamment avancée pour permettre la conception de système complexe au sein d'un unique circuit programmable. Les performances et la flexibilité des circuits à haute densité d'intégration récents constituent des avantages essentiels pour évaluer et mettre en œuvre rapidement une architecture matérielle donnée. L'aspect programmable de ces composants est un atout majeur pour concevoir une plate-forme de recherche permettant d'implanter et d'expérimenter des systèmes embarqués basés sur les concepts de vision active. De plus,

5. Hardware Description Language

le parallélisme de tâche autorisé par ces circuits est adapté à l'implémentation de tâches concurrentes inhérentes à un système de vision active. Les capteurs embarqués qui incorporent un processeur directement connecté au flot d'image, utilisent une partie importante des ressources de calcul de leur CPU uniquement pour acquérir les images. Ce type d'organisations fonctionnelles ne semble pas adapté pour exploiter efficacement les facultés d'un processeur. L'utilisation d'une unique unité de traitement séquentielle pour réaliser à la fois l'acquisition et l'ensemble des traitements d'images nécessite un partage de la bande passante difficile. Les processeurs fournissent une unité de calcul séquentielle extrêmement performante et sont capables de manipuler des représentations de l'information variées. Mais, ils sont souvent limités par leur capacité de communication avec l'extérieur.

Pour ces raisons, nous proposons de constituer notre plate-forme suivant une solution d'implantation mixte constituée d'un DSP et d'un FPGA. Le nombre d'applications de vision développées à partir de processeurs DSP démontrent leur adéquation pour la conception de machine de vision. Ce type de solution est essentielle pour implémenter les dispositifs de haut niveau des systèmes de vision manipulant divers types de représentation de l'information. D'autre part, la solution câblée programmable fournit un espace d'implantation matérielle versatile. Le composant programmable permet au système d'adopter toutes sortes d'organisations fonctionnelles et de solutions architecturales. La plate-forme d'implantation basée sur cette solution d'implantation mixte est présentée dans le chapitre suivant.

Chapitre 3

Conception d'une plate-forme de recherche

La plate-forme de recherche proposée pour l'expérimentation de systèmes de vision active embarqués (fig. 3.3) est une caméra "intelligente" configurable. Le dispositif sensoriel est basé sur un imageur CMOS et une centrale inertielle. Un dispositif d'imagerie CMOS a été choisi afin d'utiliser le mode de lecture contrôlé de l'image au sein d'un système de vision active. Le dispositif de mesures inertielles est adjoint pour étendre les capacités de perception. La caméra embarquée est reliée avec un système d'exploitation hôte via un module de communication. L'espace d'implantation de cette caméra "intelligente" repose sur une architecture de traitement hétérogène composée d'un FPGA et d'un DSP.

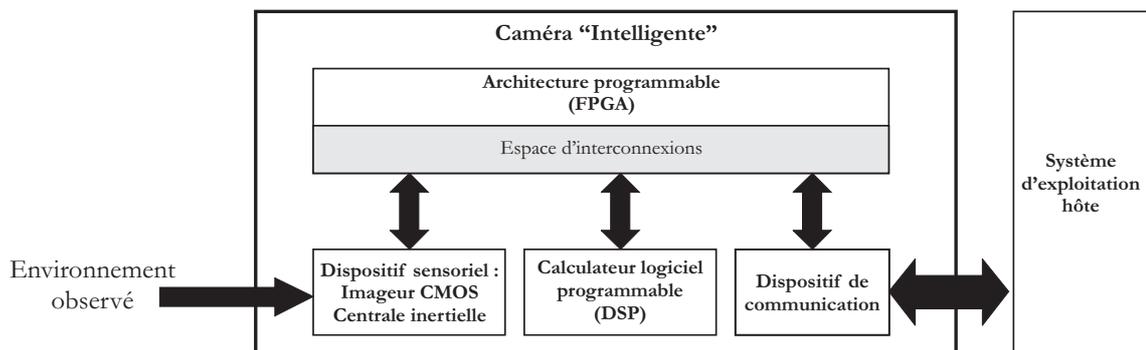


FIG. 3.1 – Organisation fonctionnelle de la plate-forme.

L'organisation fonctionnelle du capteur est centrée sur le circuit à architecture programmable (fig. 3.1). L'ensemble des modules fonctionnels de la caméra "intelligente" sont reliés à ce circuit qui permet de définir leurs interactions fonctionnelles. De cette manière, le circuit FPGA est utilisé à la fois comme calculateur dédié et comme dispositif d'interconnexion programmable. Suivant les desseins de l'utilisateur, l'espace d'implantation permet de définir un système de vision spécifique. Le procédé de traitement de la caméra peut utiliser comme calculateur le FPGA ou/et le DSP. Dans le cas d'une architecture mixte, une organisation hiérarchique peut être instaurée en considérant l'un des deux calculateurs comme superviseur. D'autre part, la possibilité d'implémenter toute sorte de flots de données au sein du FPGA permet la mise en œuvre "aisée" de systèmes de vision active. Grâce aux capacités d'interconnexion du circuit, des rétroactions mécaniques, optiques, sensorielles et algorithmiques sont envisageables.

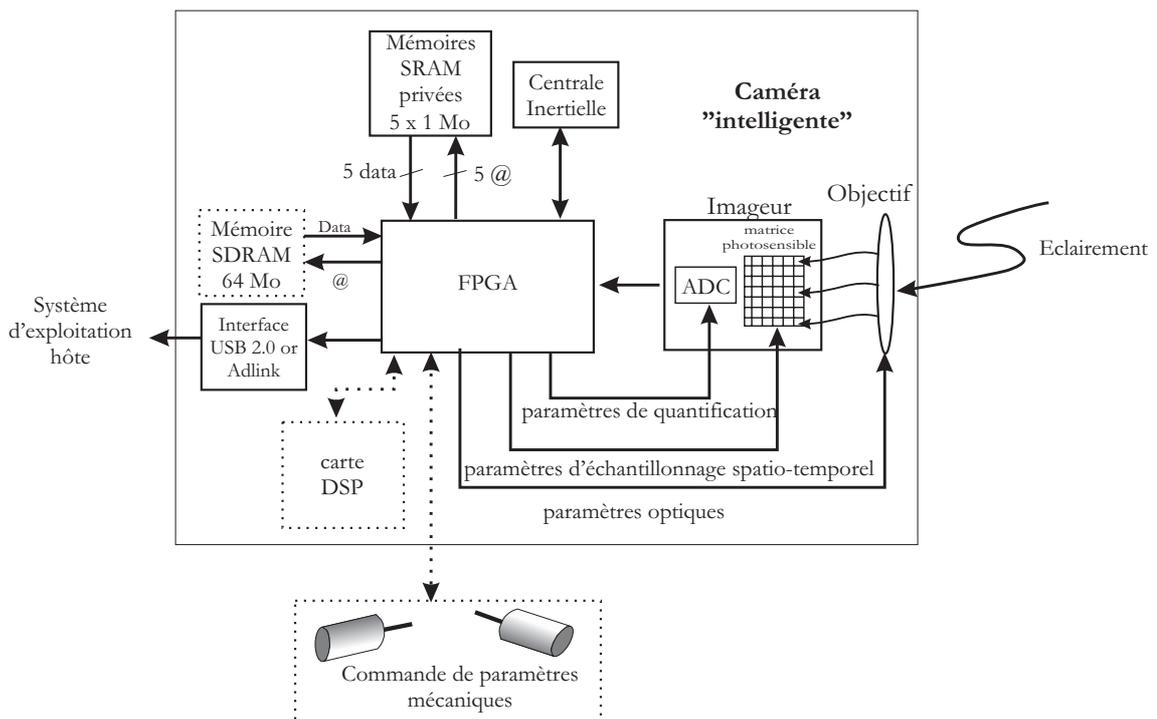


FIG. 3.2 – *Synoptique du système embarqué.*

Un schéma synoptique représentant les différentes fonctionnalités du système embarqué est proposé figure 3.2. Ce dispositif constitue un outil novateur pour éva-

luer et expérimenter l'adéquation algorithmique et architecturale des systèmes de vision embarqués. Bien que le système proposé présente un intérêt générique, cet outil d'implantation a été conçu pour développer des systèmes exploitant les propriétés des imageurs CMOS dans un système embarqué dédié à la vision active. Ce chapitre est consacré à la description de la structure matérielle de la plate-forme de recherche proposée. Dans la première partie la structure générale du capteur est présentée. Ensuite, les différentes cartes qui constituent la caméra sont détaillées afin d'étudier ses fonctionnalités.

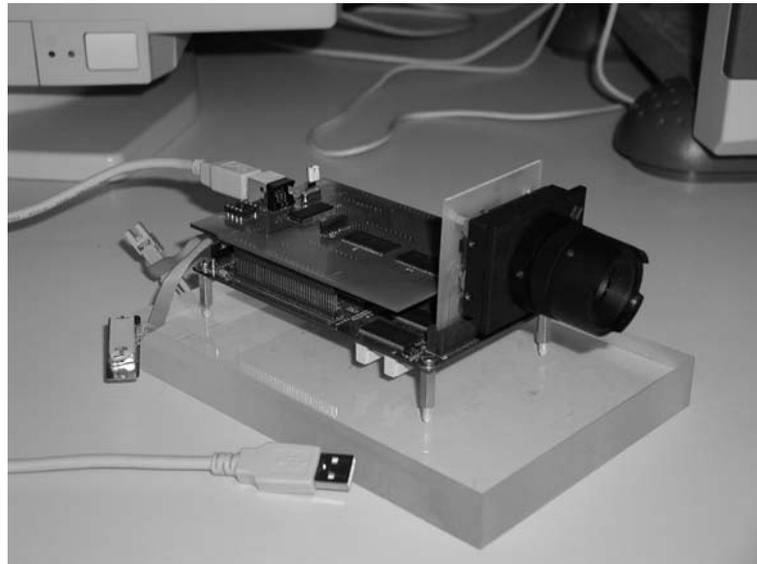


FIG. 3.3 – Plate-forme de recherche proposée pour l'expérimentation de système de vision active embarqué.

3.1 Structure matérielle

La structure matérielle de notre système est basée sur une architecture de type fagot (fig. 3.4). Les fonctionnalités majeures du système sont séparées sur un ensemble de cartes indépendantes. L'objectif de cette décomposition est de permettre une évolution rapide du système et un maximum de flexibilité. Le système s'articule autour de la carte de traitement principale qui accueille le composant à architecture programmable (SOPC). Ce composant présente un lien de communication vers l'en-

semble des cartes périphériques du système embarqué. Ces liens de communication sont contraints uniquement par un nombre de connexions maximum et l'emplacement des signaux d'horloge et d'alimentation.

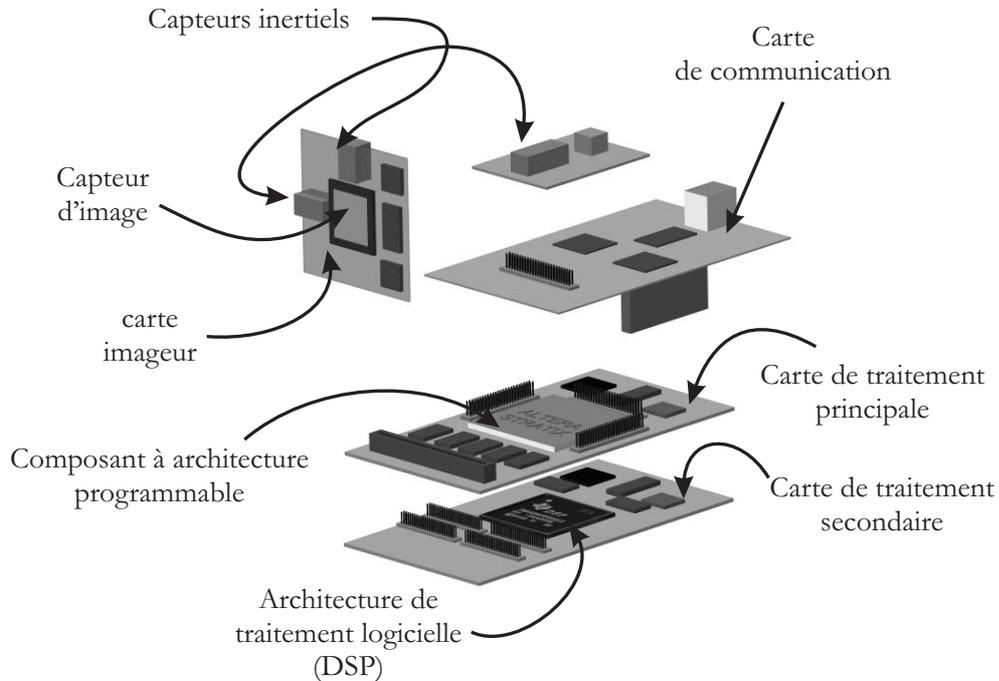


FIG. 3.4 – *Structure fagot évolutive constituant le système embarqué.*

Chaque carte peut être remplacée par une nouvelle présentant des fonctionnalités différentes afin de définir un nouveau système embarqué. Ainsi, la plate-forme peut évoluer au gré des progrès technologiques de ces composants sans nécessiter une reconstruction totale du système.

Cet outil d'implantation s'appuie sur la flexibilité de la technologie FPGA. En effet, les composants à architecture programmable (SOPC) permettent de concevoir des architectures fonctionnelles suivant les spécificités d'une application en disposant d'un certain nombre d'éléments architecturaux figés. Dans notre cas, différents blocs mémoires (SDRAM et SRAMs externes), une carte munie d'un microprocesseur dédié au traitement du signal (DSPs), un ensemble de modules sensoriels (imageur et capteurs inertiels) et divers ports d'entrées/sorties sont disponibles pour faciliter la spécification système embarqué par un utilisateur.

La plate-forme comporte trois cartes périphériques principales interchangeables (fig. 3.4):

- Une carte dédiée à la communication avec le système hôte utilisant le capteur vision embarqué.
- Une carte supportant le dispositif sensoriel de la caméra "intelligente".
- Une carte de traitement secondaire dont le rôle est de décharger la carte de traitement principale lors de la mise oeuvre de systèmes complexes.

Les fonctionnalités des cartes constituant la caméra "intelligente" sont détaillés dans la suite du chapitre.

3.2 La carte de traitement principale

La carte de traitement principale accueille le circuit programmable à haute densité d'intégration (Stratix EP1S60 - fig. 3.5). Ce composant permet à l'utilisateur de redéfinir indéfiniment le système implémenté sur la plate-forme et d'exploiter les diverses fonctionnalités matérielles proposées. La plate-forme de recherche bénéficie ainsi de la richesse des outils de conception logiciels de la technologie FPGA qui permettent une implémentation à la fois simple, rapide et peu onéreuse.

Dans les parties suivantes, les fonctionnalités internes au FPGA choisi sont présentées, ainsi que les modules de mémorisation externes implantés sur la carte.

3.2.1 Fonctionnalités du circuit programmable

Le circuit choisi lors de la conception de la plate-forme est un Stratix EP1S60 (fig. 3.5) développé par la société Altera en 2002. Ce circuit est basé sur une technologie de conception 1,5V, 0,13 μ m. Les interconnexions et la configuration des blocs logiques sont définis par un module SRAM qui permet une reconfiguration rapide du circuit. Il s'appuie sur une architecture bidimensionnelle qui permet d'interconnecter des tableaux de d'éléments logiques (LABs) comportant chacun 10 éléments logiques configurables (LEs). Cette architecture comporte 57 120 LEs, différents blocs mémoires internes et un ensemble de modules arithmétiques optimisés ainsi que 1022 entrées/sorties. Les caractéristiques de ce circuit sont détaillées par la suite. Tout

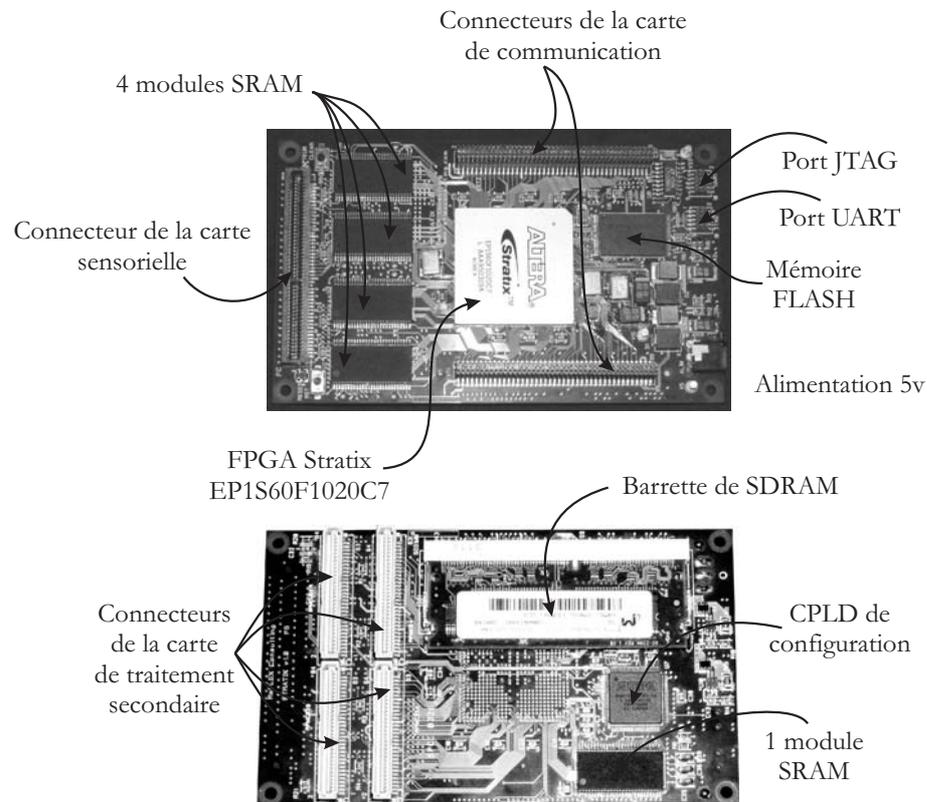


FIG. 3.5 – Composants de la carte de traitement principale.

d'abord, la structure matérielle programmable du composant est présentée. Ensuite, les fonctionnalités offertes par cette technologie de SOPC sont énumérées.

3.2.1.1 Structure Matérielle Configurable

Ces tableaux d'éléments logiques sont disposés dans une matrice d'interconnexion de 90 colonnes par 73 lignes et chacun d'eux comportent 10 éléments logiques configurables (LEs). Ces éléments constituent les composants basiques permettant l'implémentation d'une architecture câblée quelconque dans le circuit. Un élément logique configurable est constitué d'une LUT¹ suivi d'une bascule paramétrable. La LUT permet d'implémenter toute fonction logique combinatoire à quatre entrées et permet de propager un ensemble de retenues afin d'implémenter efficacement

1. LUT - Look Up Table

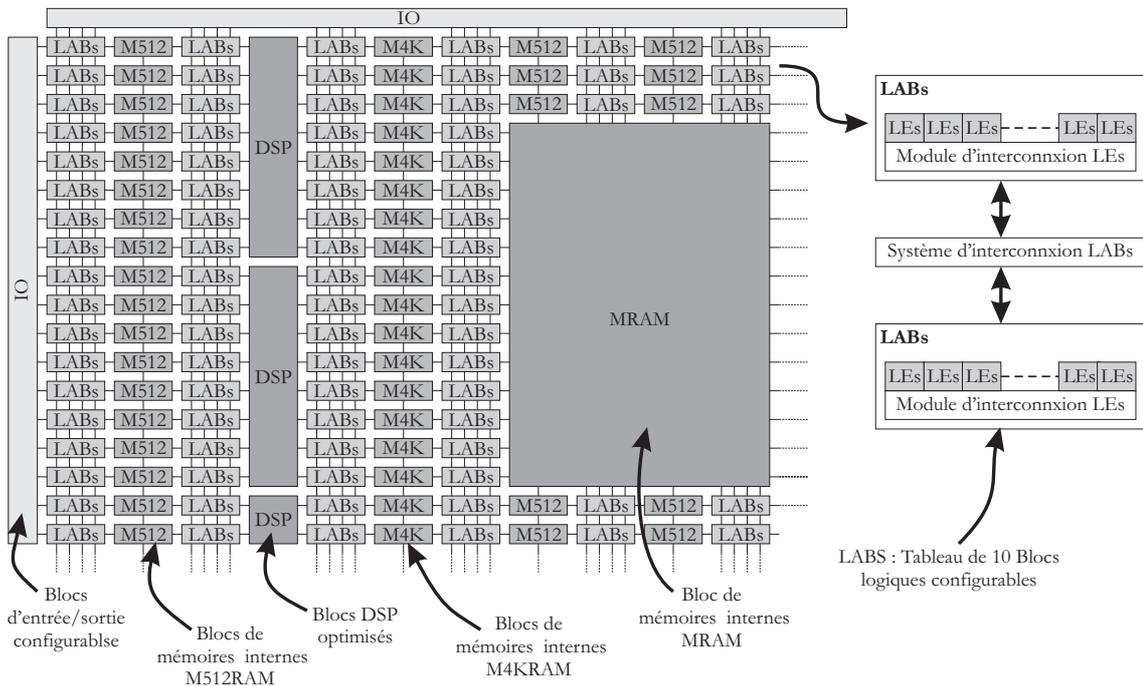


FIG. 3.6 – Organisation structurelle des FPGAs Stratix.

des fonctions arithmétiques. La bascule peut adopter différents comportements synchrones et asynchrones suivant la configuration de l'ensemble de ces fonctions combinatoires périphériques (fig. 3.7).

L'interface du système implantée avec l'extérieur est réalisée dans une structure dédiée d'entrées/sorties. Chaque broche du composant est accessible via un circuit d'adaptation d'impédance bidirectionnel. Cette structure présente une grande variété de mode de fonctionnement. Cette gamme de mode de communication permet d'accorder le circuit FPGA avec les divers standards exigés par ses composants périphériques externes. Notre plate-forme de développement exploite environ 700 broches d'entée/sortie sur les 1020 disponibles. Ce nombre important de connexions permet au système configurable de disposer d'un nombre de périphériques importants. Deux horloges cadencées à 60 Mhz sont connectées à deux entrées réservées reliées à des PLLs internes au FPGA. Ces PLLs hautement paramétrables permettent de générer de multiples horloges internes pour synchroniser les architecture implémentées.

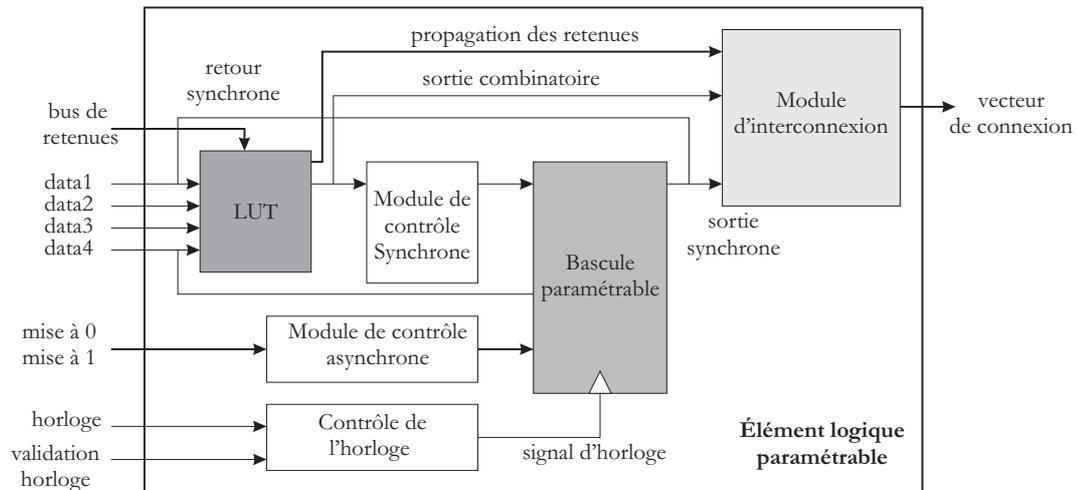


FIG. 3.7 – Structure fonctionnelle d'une cellule élémentaire configurable (LE).

Les circuits Stratix intègrent dans leur architecture un ensemble de blocs de mémorisation et de fonctions arithmétiques élémentaires câblées (fig. 3.6). La synthèse standard d'éléments de mémorisation et de multiplieurs rapides nécessite l'emploi d'un grand nombre de cellules élémentaires. Aussi, ces structures précâblées permettent une implémentation efficace de ces fonctions tout en économisant l'espace d'implémentation.

3.2.1.2 Les blocs de mémorisation internes

Le circuit intègre un ensemble de modules de mémorisation dénommé mémoires TriMatrix ©. Ces blocs de mémorisation sont de trois types (fig. 3.6) : M512, M4K et M-RAM. Ils permettent d'implémenter différents types de mémoires synchrones simple ou double ports qui sont utilisables sous la forme de RAMs, de ROMs ou encore de FIFOs. Les cycles d'écriture et de lecture peuvent être effectués suivant différentes cadences d'horloge et les blocs M4K et M-RAM supportent des configurations munies de deux ports d'écriture et deux ports de lecture. De plus, ces blocs mémoires embarqués peuvent être exploités comme des registres à décalage qui sont des cellules essentielles pour synthétiser efficacement des dispositifs de traitement du signal. Chaque type de cellules de mémorisation est optimisé pour permettre l'implémentation de diverses fonctions. Les blocs M512 sont les plus adaptés pour

Propriétés de cellules	cellules M512 (32×18 bits)	cellule M4K (128×36 bits)	cellule M-RAM (4K×144 bits)
Fréquence maximum de fonctionnement	318 MHz	291 MHz	269 MHz
Mode 2 ports d'écriture et 2 ports de lecture		✓	✓
Mode 1 port d'écriture et 1 port de lecture.	✓	✓	✓
Mode registre à décalage	✓	✓	
Mode bi-horloges	✓	✓	✓
Configurations	512 × 1	4K × 1	64K × 9
	256 × 2	2K × 2	32K × 18
	128 × 4	1K × 4	16K × 36
	64 × 9	512 × 9	8K × 72
	32 × 18	256 × 18	4K × 144
		128 × 36	

TAB. 3.1 – Propriétés des blocs mémoires internes

implémenter des cellules FIFOs. Ils supportent une largeur de mots maximum de 18 bits et une fréquence d'horloge de 318 MHz. Par opposition, les blocs M-RAM permettent un volume de stockage important suivant une largeur de mots élevé. Chacun peut mémoriser jusqu'à 4000 mots de 144 bits et fonctionner à une fréquence de 269 Mhz. Le type de blocs M4K présente un compromis entre les deux types précédemment cités. Les propriétés de ces divers blocs mémoires sont résumées dans le tableau (tab. 3.1).

Le circuit Stratix EP1S60 choisi dispose de 574 blocs M512 32×18 bits (330 624 bits), 292 blocs M4K 128×36 bits (1 345 536 bits) et 6 blocs M-RAM 4K×144 bits (3 456 000 bits) ce qui représente une capacité totale de plus de 5 Megabits.

3.2.1.3 Les fonctions arithmétiques câblées

Les circuits Stratix intègrent des fonctions arithmétiques élémentaires câblées (fig. 3.6 - DSP) destinées à l'implémentation de modules de traitement du signal (FIR,

IIR, FFT ...). Ces composants fournissent un outil pour intégrer des fonctions de multiplication/addition et multiplication/accumulation largement utilisées dans les algorithmes de traitement du signal et de traitement d'images. Chacun de ces blocs est configurable en 8 multiplieurs 9 bits, 4 multiplieurs 18 bits ou 1 multiplieur 36 bits et peut être utilisé suivant quatre configurations : multiplieur simple, multiplieur accumulateur, 2 multiplieurs additionneurs, 4 multiplieurs additionneurs. Le circuit Stratix EP1S60 possède 144 blocs arithmétiques configurables fonctionnant à une fréquence maximum de 250 MHz.

L'implémentation de mécanismes de vision artificielle exige à la fois une grande capacité de stockage et des cadences de fonctionnement élevées. Les performances et l'accès double ports de ces cellules précâblées font d'elles des outils d'adéquation algorithmique et architecturale majeurs. Cependant, la capacité de stockage interne reste limitée pour l'implémentation de systèmes de vision et pour permettre des volumes de stockage importants, plusieurs mémoires externes sont alors nécessaires. Le rôle de la mémoire interne au composant est analogue à celui des mémoires caches utilisées dans les microprocesseurs. Les cellules de mémorisation et de multiplication câblées doivent être utilisées avec parcimonie pour leur rapidité afin de satisfaire des contraintes temporelles critiques.

3.2.1.4 Les processeurs paramétrables

Les circuits Stratix sont compatibles avec la technologie ©Nios proposée par la société Altera. Nios est une famille de processeurs embarqués qui sont synthétisés à partir du réseau de cellules programmables des FPGAs Altera. A partir d'un outil logiciel (SOPC Builder), l'utilisateur définit son propre processeur suivant les spécificités de son application. Ensuite, le logiciel génère une spécification synthétisable de l'architecture du microcontrôleur en langage de description matérielle. Le processeur Nios est un processeur RISC d'usage générale. Sa version la plus récente Nios II possèdent 32 bits de données, 32 bits d'adresse. Il atteint classiquement une cadence de traitement de 150 MIPS², dispose de 32 registres internes et 32 sources externes d'interruption. Un système basé sur un processeur Nios est équivalent à un microcontrôleur qui inclut une CPU³ paramétrable et une combinaison de périphériques

2. MIPS - Million d'instruction par seconde

3. CPU - Central Processing Unit

câblées interfacées sur le bus Avalon (fig. 3.8).

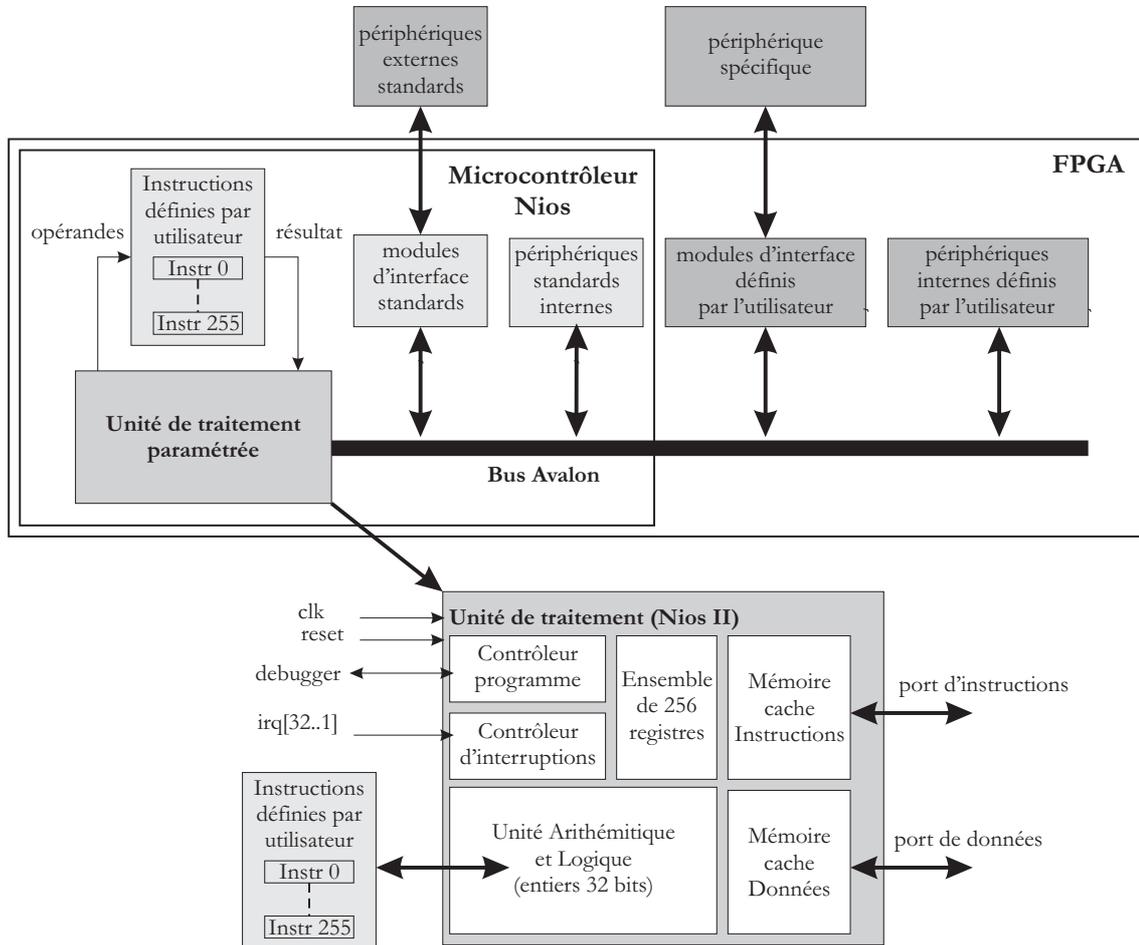


FIG. 3.8 – Description fonctionnelle d'un microcontrôleur NIOS.

La CPU d'un système Nios comporte 5 niveaux de pipeline et est munie d'un port de données et d'un port d'instructions distincts. Les versions les plus récentes du processeur Nios permettent l'insertion d'un mécanisme de mémoire cache pour les données et/ou les instructions dimensionné par l'utilisateur. Le logiciel de développement prend complètement en charge l'arbitrage des périphériques et la configuration de la CPU. De plus, il offre la possibilité de surcharger l'ALU avec un maximum de 256 instructions définies par l'utilisateur(fig. 3.8).

Cette surcharge d'instruction permet à l'utilisateur de définir une architecture câblée qui sera inséré comme une instruction supplémentaire de l'ALU standard.

De plus, Altera propose un ensemble de modules d'interface permettant la communication avec des circuits externes standard (SRAM, SDRAM, ...) afin de définir le système microcontrôleur Nios. D'autre part, l'utilisateur peut définir ses propres périphériques et les interfacier sur le bus de communication du processeur (Bus Avalon). Ainsi, les composants du processeur Nios peuvent être spécifiés précisément pour définir un système optimisé suivant les contraintes d'une application donnée.

Cependant, la capacité de calcul de ce processeur est insuffisante pour implémenter efficacement des algorithmes de traitements d'images. Dans le cadre de nos travaux, l'intérêt du processeur Nios réside dans sa capacité d'interfaçage avec un ensemble de périphériques spécifiques. En considérant ces périphériques comme des modules indépendants composant un système de vision embarqué, le processeur Nios constitue un mécanisme de contrôle logiciel du dispositif. Ainsi, il joue le rôle de superviseur programmable chargé de synchroniser les tâches et de gérer l'échange d'informations au sein du système.

3.2.2 Les modules de mémorisation externes

La quantité d'informations contenue dans le signal image induit un volume élevé de données à manipuler. Les algorithmes de vision sont basés sur des traitements spatio-temporels de ces données et l'implémentation d'un système de vision nécessite une capacité de mémorisation importante. Bien que les éléments de mémorisation interne (TriMatrix©) au Stratix soient performants, ils ne représentent qu'une faible capacité de stockage vis à vis d'un système de vision. Dans notre cas de figure, différents modules mémoires ont été câblés autour du FPGA de manière à faciliter l'implantation d'un système de vision. L'utilisateur dispose de trois types de circuits de mémorisation externes (fig. 3.5):

- **Cinq modules de SRAM indépendants** sont interfacés sur le FPGA. Chaque module possède ses propres bus d'adresse et de données afin de permettre des accès mémoires distincts. L'objectif est de permettre à divers modules du système implanté d'exploiter chacun son module mémoire simultanément et indépendamment. Ces blocs mémoire autorisent un temps d'accès minimum de 10 nanosecondes et une capacité de stockage d'un million de mots de 16 bits.

- La carte peut accueillir **une barrette de SDRAM** qui permet au système embarqué de disposer d'une capacité de stockage élevée. Ce type de circuit mémoire qui est utilisé dans les ordinateurs standards profite d'une technologie avancée. Suivant un protocole d'accès fortement contraint, ces mémoires autorisent à la fois une cadence d'accès rapide et une capacité de stockage importante. Malgré l'effort de mise en œuvre que requiert son utilisation, ce module de mémorisation paraît incontournable pour implémenter les algorithmes exigeant le stockage d'une grosse quantité d'information tels que des procédés de reconnaissance. La plate-forme de recherche permet d'exploiter une barrette SDRAM SODIM de 144 broches cadencée à 133 MHz ou 100 MHz. Contraint par le nombre de broches du connecteur SODIM, la capacité de stockage maximum sur la barrette SDRAM est de 64 Mo.
- **Une mémoire FLASH** est également implantée sur la carte de traitement principale de la caméra "intelligente". La technologie du Stratix est basée sur un procédé de configuration volatile. Afin d'éviter une reprogrammation logicielle systématique du FPGA, cette mémoire est utilisée pour stocker une configuration prédéterminée du composant. Lors de chaque mise sous tension du système embarqué, un circuit programmable non volatile (CPLD) configure le FPGA suivant les données stockées dans la mémoire FLASH. Une partie de cette mémoire peut être utilisée pour stocker des données de l'utilisateur. Dans notre cas, elle peut être utilisée pour conserver, hors tension, l'image de référence servant à corriger le flot de pixels fournit par l'imageur CMOS (§ 2). Cette mémoire représente une capacité de stockage de 8 Mo avec un temps d'accès minimum de 90 ns.

L'objectif de cet ensemble de dispositifs de mémorisation externes est de fournir des moyens de stockage suffisamment importants et variés pour satisfaire l'implantation de la plupart des algorithmes dédiés à la vision active.

3.2.3 L'environnement d'implantation matérielle

L'ensemble des fonctionnalités internes et externes présentées précédemment constitue la boîte à outils qui va permettre la mise d'un système vision embarqué. Le composant programmable Stratix constitue le coeur de la plate-forme de

développement réalisée. Il représente la partie modelable de l'architecture du système embarquée. L'utilisateur dispose de l'environnement logiciel et des outils de développement lié à ce circuit programmable pour implémenter son propre système. La fonction première du circuit programmable est de permettre l'implantation rapide et interchangeable d'architectures liant les divers fonctionnalités proposées par la plate-forme. L'ensemble des fonctionnalités externes, la capacité d'intégration du circuit programmable et la variété des outils développements qui lui sont associés forment une combinaison qui constitue un outil de recherche versatile pour étudier les problèmes d'adéquation algorithmique et architecturale des systèmes de vision active embarqués. Le nombre important entrées/sorties disponibles sur le circuit Stratix (fig. 3.5) a permis de doter le système d'une variété de fonctionnalités matérielles. Ces fonctionnalités dispersées sur les diverses cartes constituant la caméra sont détaillées dans les sections suivantes.

3.3 La carte sensorielle

Cette carte supporte les éléments sensoriels du système embarqué. Actuellement, la partie sensorielle développée pour la plate forme de recherche allie un dispositif d'imagerie et un dispositif de mesures inertielles (fig.3.9).

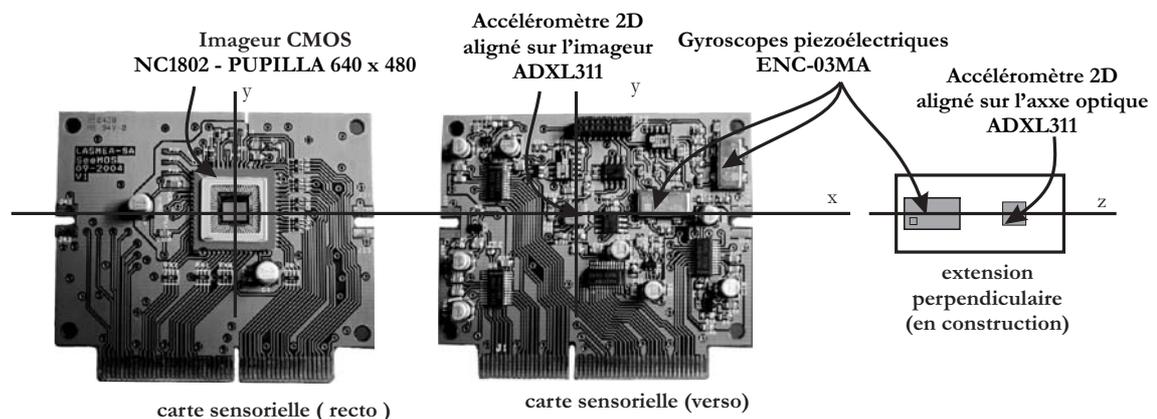


FIG. 3.9 – Les capteurs implantés sur la carte Sensorielle.

L'association de données inertielles et visuelles est motivée par des méthodes de reconstruction tridimensionnelle intéressantes [83, 84]. Le système inertiel permet

d'identifier les vitesses de rotation et les accélérations linéaires subies par la caméra. Ses mesures fusionnées avec les données visuelles devraient permettre d'estimer la position relative d'un objet. Bien que des solutions basées uniquement sur la vision soient déjà proposées, cette approche originale permet d'envisager une simplification du système de vision et une augmentation de la fiabilité de l'estimation ce qui conforte les postulats de la vision active.

En accord avec les conclusions établies dans le paragraphe 2, le dispositif d'imagerie choisie est basé sur un imageur CMOS. Ce choix est motivé par les possibilités offertes par cette technologie d'imagerie pour mettre en œuvre des systèmes de vision active novateurs. Tout d'abord, le mode d'accès aléatoire aux pixels autorise un contrôle sur l'échantillonnage de la matrice photosensible. Cette propriété permet d'élaborer des techniques de fenêtrage afin d'augmenter la cadence d'acquisition image du capteur en ne sélectionnant uniquement que les pixels utiles. Les ressources de calcul disponibles ne sont ainsi focalisées que sur des zones d'intérêt de l'espace visuel. De plus, le contrôle de l'échantillonnage spatial peut être exploité dans un système de vision active afin d'augmenter ses facultés de perception. D'autre part, la grande dynamique offerte par les pixels de type photodiode en mode courant permet une adaptation du capteur aux conditions variables de luminosité d'un environnement réel. Afin d'exploiter cette dynamique, une rétroaction visuelle peut être effectuée sur le capteur pour sélectionner la partie utile de cette caractéristique suivant les conditions lumineuses instantanée et profiter d'un maximum de dynamique.

Selon ces motivations, le système embarqué possède un dispositif d'imagerie et un dispositif de mesures inertielles implantés sur la carte sensorielle (fig.3.9) qui sont détaillées dans les paragraphes suivants.

3.3.1 Le dispositif de mesures inertielles

Le dispositif de mesures inertielles est composé de deux accéléromètres linéaires 2D et trois gyroscopes piézoélectriques (mesures des vitesses angulaires). Deux gyroscopes et un accéléromètre 2D sont placés auprès de l'imageur afin d'effectuer les mesures inertielles γ_x , γ_y , ω_x et ω_y relatives au centre C de la matrice photosensible (fig. 3.10). Les gyroscopes sont disposés suivant deux axes perpendiculaires (fig. 3.9)

qui sont alignés avec le repère \mathcal{R}_c (fig. 3.10). Ces gyroscopes permettent de mesurer les vitesses angulaires ω_x et ω_y par rapport aux axes x et y du repère \mathcal{R}_c . Afin d'effectuer les mesures γ_x, γ_y représentatives des accélérations linéaires subit par le centre C de la matrice photosensible, l'accéléromètre 2D est positionné au plus près du point C . Ensuite, pour évaluer les mesures γ_z et ω_z suivant l'axe optique, un autre gyroscope et un autre accéléromètre sont placés sur une carte d'extension perpendiculaire à la carte sensorielle (fig. 3.9). Cette carte d'extension est positionnée suivant l'axe z du repère \mathcal{R}_c de manière à effectuer des mesures significatives de γ_z et ω_z . Les gyroscopes utilisés (ENC-03MA) sont proposés par la société muRata.

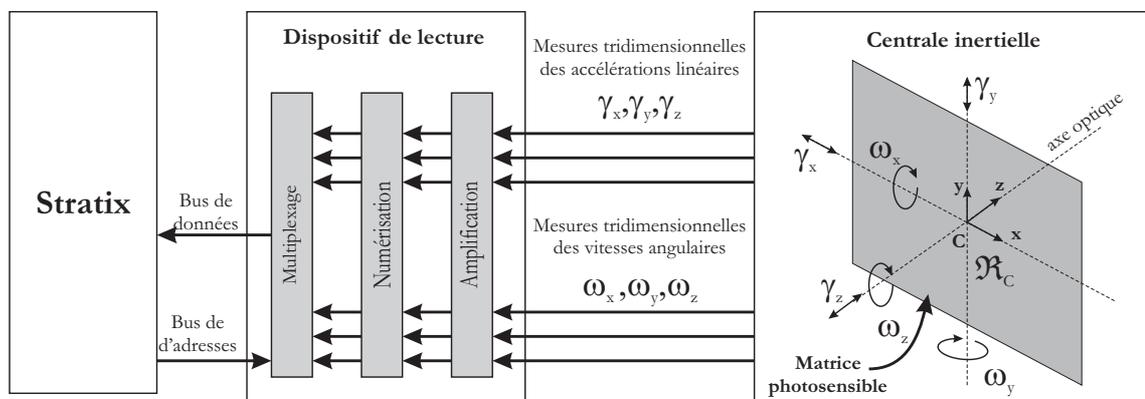


FIG. 3.10 – *Dispositif de mesures inertielles.*

Le capteur ENC-03MA quantifie la vitesse de rotation angulaire à partir de la force de Coriolis subit par un vibreur piézoélectrique. Il fournit une réponse linéaire ($\pm 5\%$) mesurant une vitesse angulaire maximum de ± 300 degrés/seconde suivant un temps de réponse de 20 ms. Les accéléromètres 2D, développés par AnalogDevices (ADXL311) sont basés sur une structure de mesure capacitive. La structure de l'accéléromètre est constituée d'une microsurface de polysilicium usinée. L'élasticité de cette surface permet aux forces d'accélération de la déformer. Une capacité différentielle constituée d'une armature mobile positionnée entre des armatures fixes permet d'évaluer les accélérations subies par la structure. Attachée à une masse sismique la position de l'armature centrale varie et modifie les valeurs de deux capacités formées par ces trois armatures. Le capteur ADXL311 est capable de mesurer des accélérations de ± 2 g selon une précision de 2 mg avec un temps de réponse similaire aux gyroscopes. Ces temps de réponse étant faibles, le circuit de lecture de cette centrale

inertielle est multiplexé afin de réduire le nombre de broches de connexions à la carte de traitement principale.

3.3.2 Le dispositif d'imagerie

Le dispositif d'imagerie est basé sur un imageur CMOS NC1802 - PUPILLA de la société Neuricam. Fabriqué en technologie $0.35 \mu\text{m}$, ce composant intègre une matrice photosensible de 640 par 480 pixels, un amplificateur analogique et un convertisseur numérique 10 bits. La matrice photosensible est constituée de pixels photodiode en mode courant qui autorisent une grande dynamique de mesure (120 dB). Ces pixels sont de géométrie carrée et présentent un taux de remplissage de 40 %.

La plupart des imageurs CMOS du commerce multiplexent les bus d'adressage ligne et colonne de la matrice CMOS afin d'économiser les broches de connexions du circuit. Les deux premiers imageurs qui ont été évalués, les composants Fuga 15 et Fuga 1000 de la société Fill Factory présentent ce type de circuit de multiplexage. Ces dispositifs introduisent des temps de commutation importants qui limitent l'exploitation de l'échantillonnage contrôlé sur la matrice. Ainsi, l'imageur choisi possède un bus d'adresse colonne et un bus d'adresse ligne distincts. Cette architecture d'adressage basique permet d'accéder à chaque pixel de la matrice avec un temps de lecture similaire. Cette caractéristique est essentielle pour développer des dispositifs de fenêtrage efficaces et non contraints par le capteur.

La cadence de fonctionnement standard de cet imageur est 8 Mhz (25 images/sec). Cependant, en réalisant un réglage précis de la tension précharge des pixels (fig. 3.12 - V_{prec}) et en occasionnant une surconsommation du circuit, la cadence d'acquisition peut être élevée à 16 MHz (52 images/sec). Le rôle de cette tension V_{prec} est de charger les capacités internes des lignes de connexions de la matrice lorsqu'elles ne sont pas connectées à un pixel. Cette précharge permet de réduire le temps de réponse du circuit de lecture et ainsi d'accélérer la cadence d'acquisition des pixels. Mais, un mauvais réglage de ce système de précharge dégrade le contraste de l'image.

Le circuit dispose d'un ensemble de tensions analogiques qui permettent de contrôler la conversion numérique du signal image. Une tension référence V_{ref} (fig.

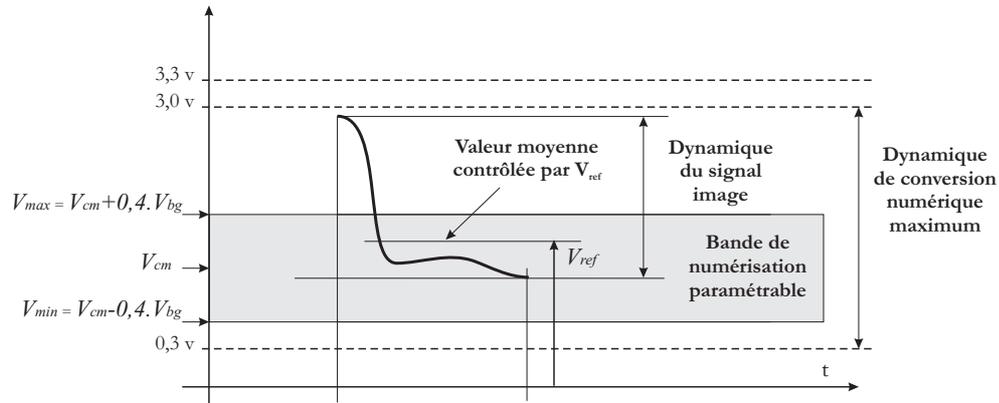


FIG. 3.11 – *Dynamique de conversion numérique paramétrable.*

3.11) permet de centrer la forme d'onde du signal analogique sur la dynamique de convertisseur analogique numérique (10 bits). Cette dynamique de conversion est paramétrée par les valeurs des tensions V_{cm} et V_{bg} . La tension V_{cm} fixe la position du centre de la dynamique et la tension V_{bg} détermine la largeur de la bande de conversion. Afin de permettre au système de faire évoluer ses tensions, 4 convertisseurs numérique/analogique rapides sont implantés sur la carte sensorielle. Ce dispositif de commande permet de contrôler les quatre tensions V_{prec} , V_{ref} , V_{cm} et V_{bg} en temps réel. Le temps réponse du dispositif de conversion autorise une cadence de commande supérieure à la cadence maximum d'acquisition des pixels (16 MHz).

Ce dispositif d'imagerie a été étudié pour permettre au système de vision de contrôler l'échantillonnage spatial de l'information visuelle et le procédé de numérisation du signal image. Ainsi, Il autorise deux types de rétroactions visuelles pour mettre en œuvre un système de vision active (fig. 3.12). Tout d'abord, le procédé d'adressage aléatoire de l'imageur CMOS, permet de définir des fenêtres d'intérêt de géométrie et de résolution variées. Ensuite, la caractéristique logarithmique des pixels permet une photodétection efficace malgré des conditions changeantes d'éclairément. Le contrôle du processus de numérisation autorise la sélection instantanée d'un domaine conversion utile afin d'optimiser le contraste dynamiquement. Ainsi, des méthodes d'optimisation du contraste peuvent être envisagée pour régler indépendamment l'histogramme de chaque fenêtre d'intérêt.

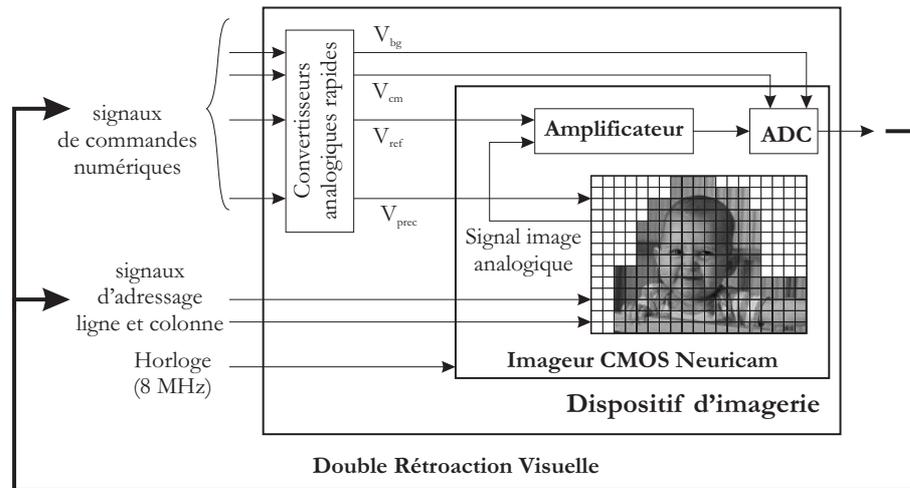


FIG. 3.12 – *Système d'imagerie CMOS permettant une double rétroaction visuelle.*

3.4 La carte de communication

L'approche active est basée sur l'implémentation d'une rétroaction de l'information visuelle sur le système. Cette rétroaction visuelle est fortement dépendante de l'efficacité des moyens de communication disponibles. Dans cette thèse, il est proposé de déporter les boucles de rétroaction critiques auprès du capteur afin de bénéficier de liens de communication privilégiés et dédiés. La proximité du capteur et du système embarqué autorise l'implémentation de système réactif temps réel efficace.

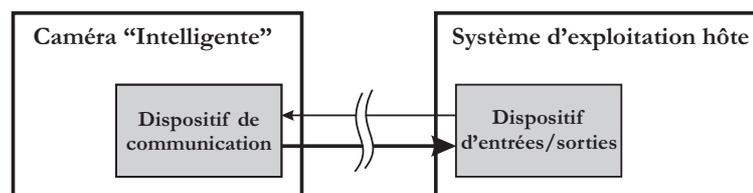


FIG. 3.13 – *Lien de communication de la caméra "intelligente" avec un système d'exploitation hôte.*

Cependant, le système embarqué nécessite un moyen de communication avec un système d'exploitation hôte (fig. 3.13). Le système hôte qui va exploiter les in-

formations délivrées par le capteur peut être de nature diverse (système robotique ou simple dispositif d'affichage). Dans notre cas, le système hôte est un ordinateur standard et l'interlocuteur de notre capteur embarqué est un système d'exploitation informatique classique. Le lien de communication entre le système informatique et le capteur représente une barrière de communication importante. Dans le cas des systèmes de vision active, ce problème de communication est d'autant plus important qu'il conditionne sa réactivité globale. L'efficacité d'un système vision haut niveau qui tire profit du système embarqué proposé est fortement dépendante de ce lien de communication. Pour satisfaire ce besoin d'échange d'information inter-systèmes, deux cartes de communications sont proposées. La première utilise une carte d'acquisition PCI et l'autre est basé sur un protocole de communication USB2.0 standard (fig. 3.14).

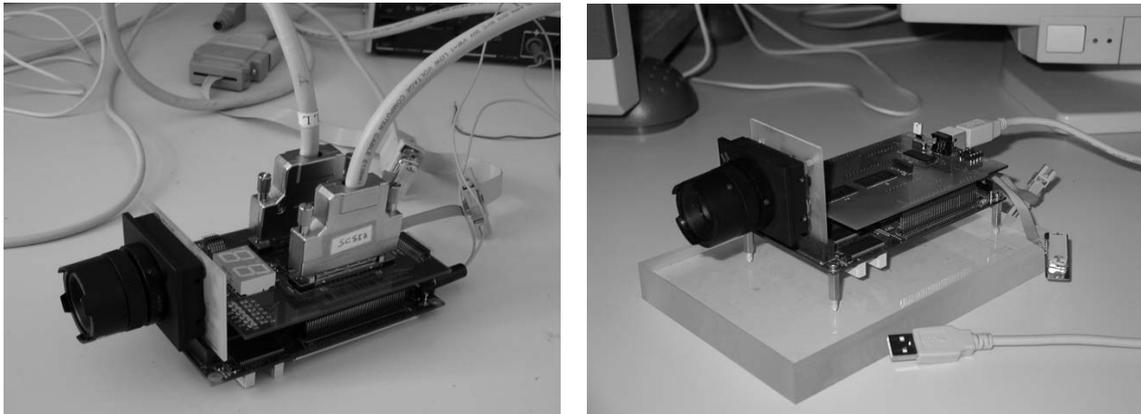


FIG. 3.14 – *Système embarqué muni à gauche de sa carte de communication PCI et à droite de sa carte de communication USB2.0.*

3.4.1 La carte de communication PCI

Cette carte de communication établit un lien vers une carte d'acquisition numérique rapide PCI-7300B développée par la société ADLINK Technology Inc. Cette carte d'acquisition possède 32 voies bidirectionnelles et permet une cadence de transfert maximum de 80Moctets/seconde. Elle constitue un moyen de communication rapide via le bus PCI⁴ du système informatique. L'ensemble de ces voies sont confi-

4. PCI - Peripheral Component Interconnect

gurable en lecture et en écriture à partir de bus 8, 16 ou 32 bits et elle peut être utilisée suivant quatre modes opératoires:

- Un mode de synchronisation interne où les transferts sont synchronisés par la carte d'acquisition à 10 ou 20 MHz.
- Un mode de synchronisation externe où les transferts sont synchronisés par une horloge provenant du système embarqué (maximum 20 MHz).
- Un mode "*Handshaking*" où les transferts sont effectués selon un protocole de validation par données ou par paquets.
- Un second mode de synchronisation interne où les transferts sont cadencés selon la configuration d'un programmeur interne (PLL⁵) qui permet un réglage spécifique de la cadence de communication.

Parmi ces protocoles, le mode de synchronisation interne à 20 MHz est utilisé pour transférer des données du système hôte vers la caméra. Dans l'autre sens, le mode de synchronisation externe a été choisi. La communication des pixels est cadencée à la fréquence d'acquisition de l'imageur. Ainsi, le stockage de l'image dans la caméra n'est pas obligatoire.

Sur cette carte un ensemble de ports de communication additionnels ont été prévus. Ces ports permettent d'exercer une palette de commandes numériques afin d'insérer le dispositif embarqué d'un système de vision active robotique. Ainsi, la plate-forme proposée peut exercer un contrôle sur sa vitesse et sa position suivant un dispositif mécanique approprié mais aussi exploiter un zoom motorisé pour contrôler les paramètres optiques du système implanté.

Cette solution de communication est extrêmement performante et constitue un très bon outil d'expérimentation. Cependant, elle ne satisfait pas réellement des contraintes d'embarquabilité et nécessite une carte d'acquisition spécifique. Afin de proposer une solution de communication générique et standard une autre carte de communication a été développée.

5. PLL - Phase-Locked Loop

3.4.2 La carte de communication USB2.0

Cette seconde carte de communication a été développée pour permettre une utilisation générique de notre système embarqué. Parmi les moyens de communication standard disponibles sur les ordinateurs actuels, deux d'entre eux présentent une bande passante suffisamment importante pour effectuer des transferts d'images efficaces : le bus USB2.0⁶ et le bus IEEE1394 ou FireWire.

Lors de la conception de cette carte, la nouvelle norme IEEE1394b présentant une bande passante théorique de 800 Mb/s n'était pas commercialement disponible. Aux vues des bandes passantes théoriques de USB2 (480 Mb/s) et de l'IEEE1394a (400Mb/s), le protocole USB 2.0 a été choisi. Le bus USB 2.0 se présente sous la forme d'une architecture étoilée et pyramidale centrée sur un serveur hôte qui relie l'ensemble des périphériques. Ce bus permet de connecter jusqu'à 127 périphériques au même hôte et permet le branchement à "chaud". Les transactions USB sont réalisées par paquets suivant des formats de transfert standard. Dans notre cas, la gestion de ce protocole est effectuée par le circuit dédié EZ-USB FX2 CY7C68013 proposé par la société Cypress. Ce circuit intègre un microcontrôleur intel 8051, une interface série, un émetteur/récepteur USB, de la RAM, des FIFOs et une interface programmable. L'interface série manipule toutes les fonctions de base de l'USB. Le microcontrôleur est utilisé pour définir un protocole spécifique utilisé pour échanger des données avec notre système embarqué. Le protocole de communication est cadencé par le microcontrôleur à 48 MHz. Les données échangées transitent par un bus 16 bits bidirectionnel.

Actuellement, le firmware et les drivers qui ont été développés pour ce composant permettent un débit maximal de 10Mw/s lors de transferts de grandes trames (une image entière). Suivant le nombre de données transférés la cadence décroît et lors du transfert de petites image (<1000 pixels), la cadence peut descendre en dessous de 1Mw/s. Cette interface entre le système informatique et le capteur limite la réactivité du système de vision sur le capteur et induit un goulot d'étranglement de l'information. L'état actuel de ce lien communication ne permet pas d'exploiter l'acquisition rapide de petite zone d'intérêt. Cependant, il permet d'interfacer rapidement et aisément notre plate-forme de développement sur un ordinateur standard.

6. USB - Universal Serial Bus

Les futurs développements du firmware et des drivers de cette carte USB devraient augmenter ses performances, mais elle ne pourront en aucun cas égaler l'efficacité de la carte d'acquisition PCI dédiée.

Pour satisfaire des besoins de portabilité et de performances brutes, deux cartes de communication sont proposées. Via de tels moyens de communication, l'exploitation efficace de notre système embarqué par un système haut niveau implanté sur un ordinateur peut bénéficier d'un système d'exploitation temps réel (RTOS). Ce type de système d'exploitation permettra de satisfaire des contraintes élevées de réactivité entre le système embarqué et système informatique.

3.5 La carte de traitement secondaire

3.5.1 Propriétés du calculateur

La carte de traitement secondaire supporte un microprocesseur dédié au traitement signal (DSP⁷). Ce calculateur est relié au composant programmable principale pour être exploiter dans l'architecture implantée. La plate-forme de développement constitue ainsi une architecture mixte permettant d'aborder diverses problématiques d'adéquation architecturale et algorithmique. Le calculateur implanté sur cette seconde carte fournit une alternative efficace à la solution entièrement "câblée".

Le processeur DSP choisi est le TMS320C6416 proposé par Texas Instruments. Il est cadencé à 720 MHz et présente des performances supérieures à 5760 millions d'instructions par seconde. Son noyau 3.15 possède 64 registres de 32 bits et 8 unités fonctionnelles indépendantes. Parmi ces unités fonctionnelles, on trouve deux multiplieurs qui fournissent des résultats sur 32 bits. Les six autres sont des unités arithmétiques et logiques. Cette de traitement est capable de réaliser 4 multiplications accumulations par cycle sur 16 bits soit 2880 millions d'opérations par seconde. Ce processeur utilise deux niveaux de mémoire cache. Le premier de niveau est composé de deux mémoires de 16 KOctets qui stockent distinctement les instructions (L1P) et les données (L1D). La mémoire cache de niveau 1 réservée aux données dispose de deux chemins d'accès vers le noyau processeur. Les instructions et les

7. DSP - Digital Signal Processor

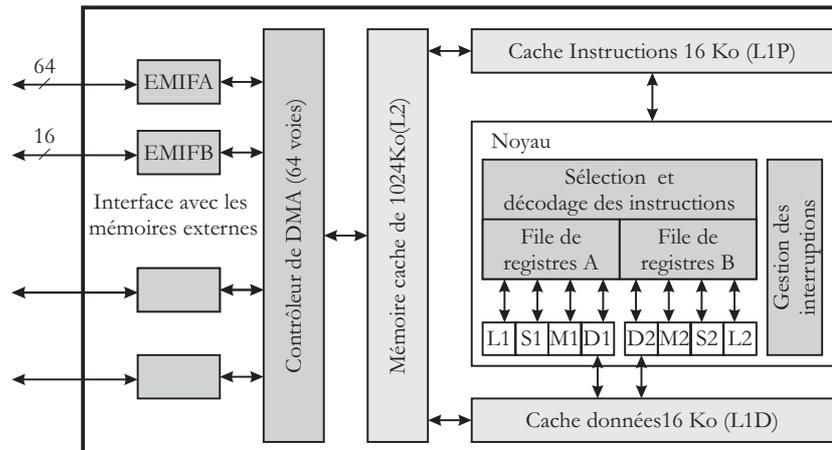


FIG. 3.15 – Architecture synoptique du microprocesseur de traitement du signal.

données transitent par une mémoire cache de niveau 2 de 1024 KOctets. Un contrôleur de DMA⁸ gère les flots d'informations provenant de l'extérieur via des modules d'interface. Principalement, le DSP possède des bus d'interface 64 bits (EMIFA) et 16 bits (EMIFB) utilisés pour interfacer des mémoires externes.

3.5.2 Cohabitation des calculateurs

L'approche classique des architectures mixtes composées d'un composant à architecture programmable et du processeur de traitement du signal consiste à définir le processeur comme superviseur du système. Nous proposons d'instaurer la relation inverse entre les deux calculateurs. L'objectif est d'utiliser le DSP pour décharger l'architecture parallèle implanter sur le FPGA. L'objectif est de programmer les fonctions complexes difficiles à réaliser par une architecture câblée au du DSP et de l'utiliser comme calculateur spécialisé.

Suivant un superviseur représenté sur la figure 3.16 par un microcontrôleur NIOS, le DSP est sollicité pour effectuer ponctuellement les fonctions difficiles à réaliser au sein d'une architecture câblée. Le superviseur sollicite une action du DSP par le biais d'un signal d'interruption. La relation maître/esclave est établie suivant deux chemins de données. Le port EMIFB du DSP est utilisé pour échanger des para-

8. Direct Memory Access

mètres et des résultats avec le superviseur. La masse de données qui est nécessaire au traitement programmé sur le FPGA est transférée par l'intermédiaire d'une RAM. Le rôle de cette mémoire est de permettre au DSP d'être maître de l'acquisition des

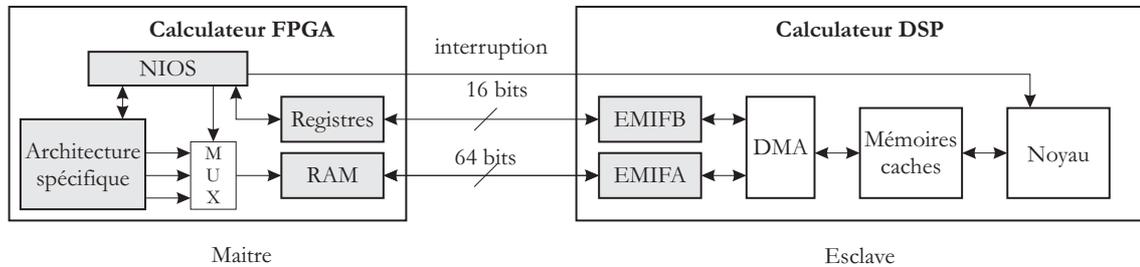


FIG. 3.16 – Utilisation du calculateur DSP comme co-processeur.

données. Car, le transfert d'information en mode flots de données vers le DSP nécessite l'utilisation d'interruption pour forcer le processeur à consulter promptement un périphérique. Ce type de transfert utilise une partie conséquente de la bande passante du processeur. Pour effectuer un transfert de données efficace, le DSP doit être maître des transferts effectués via ces périphériques. Pour cette raison, les données à traiter transitent par une RAM qui est pilotée par le DSP.

3.6 Conclusion

La plate-forme de recherche développée est un outil d'expérimentation qui peut être exploité selon les divers aspects de la problématique d'adéquation algorithmique et architecturale. L'architecture programmable de la carte de traitement principale permet l'étude d'adéquation architecturale fonctionnelle à partir d'une algorithmie. Au contraire, le calculateur logiciel de la carte de traitement secondaire permet d'évaluer l'adéquation d'implantations algorithmiques sur une architecture matérielle figée performante. L'association de ces deux cartes forme une architecture hétérogène qui permet de considérer les problèmes de partitionnement algorithmique et de co-traitement. D'autre part, la structure fagot modulaire permet d'assembler divers éléments matériels pour former une architecture physique spécifique. Ainsi, des systèmes de vision embarqués peuvent être implémentés sans nécessiter la reconstruction globale du dispositif. La carte sensorielle développée permet la mise

en œuvre de systèmes de vision active novateurs exploitant un imageur CMOS. Le dispositif permet une double rétroaction sensorielle. Le mode de lecture de l'imageur permet d'exploiter l'échantillonnage spatio-temporel du signal visuel. Et, une rétroaction peut être implantée sur les paramètres de quantification du signal analogique issu de la matrice photosensible. D'autre part, la carte sensorielle délivre un ensemble de mesures inertielles afin d'étudier les intérêts d'un couplage de données visuelles et inertielles. Ainsi, cette plate-forme constitue un outil d'expérimentation riche et versatile.

Chapitre 4

Méthodologie d'implantation

4.1 Introduction

Une différence majeure entre les systèmes de perception visuelle biologique et artificielle est sûrement la flexibilité. En effet, le dispositif de perception visuelle des vivants est engagé dans une grande variété d'activités qui nécessitent une grande flexibilité. Pour mener à bien l'ensemble de ces activités, les systèmes visuels biologiques sont à la fois extrêmement réactifs, précis et minutieux. Réactif, car la vision est une ressource de base pour identifier toute sorte de situation dangereuse nécessitant une réponse vive. D'un autre côté, la vision permet la réalisation d'une variété de tâches exigeants beaucoup d'attention et de soin. Une autre faculté que peuvent envier les dispositifs artificiels est la capacité d'adaptation des systèmes de vision biologique. En effet, l'oeil est capable de s'adapter à la nature de l'environnement observé et à tâche de perception réalisée (chap. 1). Il s'accommode aux variations de luminosité, aux distorsions optiques et prend part au processus d'attention et de focalisation.

Afin de développer la flexibilité et la faculté d'adaptation à l'environnement des systèmes de vision artificiels, une méthodologie est proposée pour implanter des systèmes de bas niveaux selon le concept de vision active (chap. 1). Dans le chapitre précédent, une architecture matérielle dédiée a été conçue pour satisfaire les contraintes fonctionnelles de la méthodologie d'implantation proposée. La singularité de l'approche proposée réside dans l'association d'un dispositif de traitement

embarqué et d'un imageur CMOS (chap. 2) pour instaurer des mécanismes de vision active bas niveaux.

4.1.1 Rétroactions visuelles

La plupart des systèmes de vision active exploitent les paramètres optiques (zoom, iris, vergence,...) et mécaniques (vitesse, position,...) pour définir des rétroactions visuelles. Des dispositifs de commande mécaniques motorisés sont utilisés pour contrôler ces paramètres. Ces dispositifs mécaniques présentent des temps de réponse relativement lents et sont confrontés à des problèmes de stabilisation de l'image (dus aux vibrations générées par le déplacement rapide de la caméra). Afin d'éviter ces difficultés, il est suggéré d'instaurer un mécanisme de retour de l'information visuelle au niveau de la transduction électrique de l'éclairage et au niveau des mécanismes de pré-traitement du signal.

4.1.1.1 Les paramètres sensoriels

Les caractéristiques du circuit imageur définissent l'échantillonnage spatio-temporel et quantitatif du signal d'éclairage (fig. 4.1). L'ensemble des paramètres

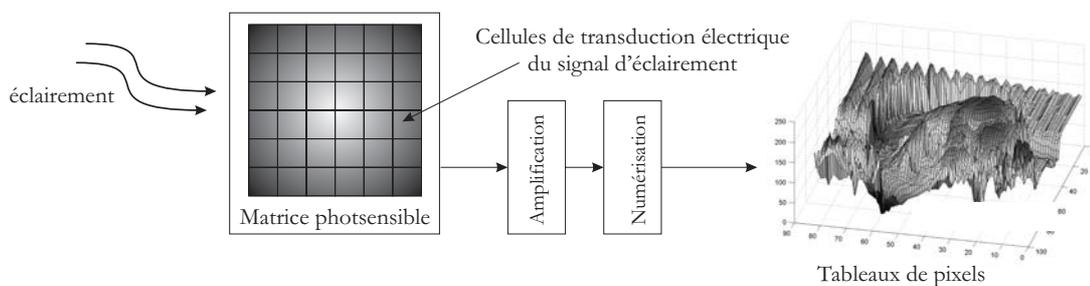


FIG. 4.1 – Échantillonnage et transduction électrique de l'éclairage.

qui influent sur la transduction du signal d'éclairage sont les paramètres sensoriels de la chaîne de perception. Dans le cas des imageurs CMOS, leur structure à accès aléatoire confèrent une propriété essentielle pour introduire l'imageur dans une boucle de perception active. Le flot de pixels délivré par l'imageur est contrôlé

par son dispositif d'adressage qui constitue la commande sur l'échantillonnage du signal projeté sur la matrice photosensible.

L'exploitation de l'échantillonnage spatial permet d'instaurer un dispositif de focalisation locale. La tâche de focalisation est réalisée en sélectionnant des zones d'intérêt sur la matrice photosensible au lieu de contrôler les mouvements de la caméra pour sélectionner successivement une cible d'intérêt. La modification de l'échantillonnage spatial a une incidence directe sur l'échantillonnage temporel du signal visuel. La fréquence d'échantillonnage temporelle peut être considérablement augmentée en réalisant une sélection des pixels utiles par le biais de fenêtrage et/ou de sous résolution. La sélection de zones locales et l'accroissement de la cadence d'acquisition simplifie alors généralement les algorithmes de traitement d'images. Un compromis optimal entre la quantité d'information utile et la fréquence de rafraîchissement de l'information doit être considéré lors du développement d'algorithmie. Évidemment, il est nécessaire de pouvoir exploiter les informations fournies par le capteur. La notion de fonctionnement en temps réel devient primordiale et l'adéquation architecturale et algorithmique vient prendre part au compromis initial.

D'autre part, une rétroaction peut être implémentée sur le dispositif de quantification du signal issu du capteur. Un réglage dynamique du dispositif de quantification permet d'adapter de conversion du signal. Un contraste optimal de l'image peut être maintenu malgré des variations de luminosité (chap. 5). En environnement réel, ce type de traitements bas niveaux est essentiel pour assurer l'intégrité du signal fournit aux mécanismes de traitements de la chaîne de perception.

4.1.1.2 Les paramètres algorithmiques

L'interprétation de la représentation de l'environnement fournie par le capteur se traduit par une implantation algorithmique. L'interprétation du signal image est certainement l'étape la plus difficile pour un dispositif de vision artificielle. Les paramètres algorithmiques de la chaîne de perception sont les éléments qui influencent le processus d'interprétation. L'exploitation des paramètres algorithmiques bas niveau conduit à instaurer une rétroaction visuelle sur les mécanismes de pré-traitement utilisés. Un exemple de rétroaction algorithmique largement utilisée est le seuillage adaptatif. Bien sûr, d'autres paramètres peuvent être ciblés tels que les coefficients

d'un filtre de convolution ou ceux d'un filtre récuratif.

D'autre part la mise œuvre de rétroactions algorithmiques est envisageable à un niveau de représentation de l'information plus élevé. Dans ce cas, les dispositifs développés donnent lieu à des mécanismes de raisonnement et de prise de décision qui s'apparentent alors à l'intelligence artificielle.

Dans l'approche proposée les paramètres algorithmiques sont utilisés selon deux objectifs. Le premier est d'adapter les mécanismes de pré-traitement selon l'évolution de l'environnement et le second objectif est de doter le système embarqué d'un caractère autonome afin de décharger le système hôte de la gestion des tâches bas niveaux.

4.1.2 Une architecture modulaire supervisée

L'approche fonctionnelle proposée s'inspire des méthodologies cognitives de plus hauts niveaux [85, 86, 87] pour définir un système de vision bas niveau. Elle est basée sur une architecture supervisée (fig. 4.2) qui exerce un contrôle sur la chaîne de traitement. Suivant les requêtes du système haut niveau, le dispositif superviseur

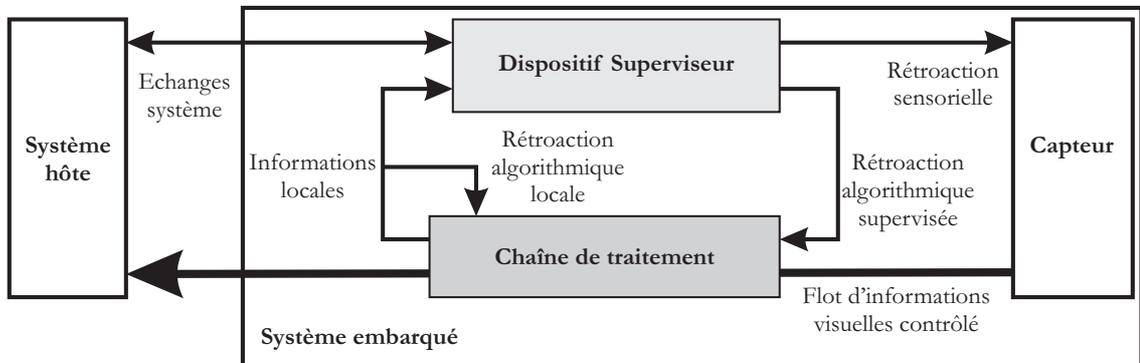


FIG. 4.2 – *Système de vision active bas niveau basé sur une architecture supervisée.*

doit paramétrer et séquencer la chaîne de traitement suivant une stratégie de vision intentionnelle (chap. 1). Les échanges avec l'hôte permettent de spécifier quelle tâche doit être réalisée et quelles sont les informations recherchées. Le but est d'adapter la stratégie d'interprétation suivant l'objectif visuel considéré. La chaîne de traitement est basée sur une architecture modulaire composé d'unités élémentaires chargées

d'extraire des informations locales. Plusieurs niveaux de rétroaction visuelle sont instaurés pour développer les capacités d'adaptation du système selon les concepts de la vision active (chap. 1). L'architecture fonctionnelle intègre trois types de rétroactions visuelles (fig. 4.2):

- les rétroactions sensorielles, qui sont exercées sur les paramètres du capteurs.
- les rétroactions algorithmiques locales, qui représentent un retour d'information visuelle sur des paramètres locaux de la chaîne de traitement.
- les rétroactions algorithmiques supervisées qui permettent d'exercer un contrôle actif de la chaîne de traitement à partir d'une interprétation globale des informations extraites.

La proximité physique entre le capteur et le système est essentielle pour implémenter des mécanismes de rétro-actions performants. La délocalisation auprès du capteur évite les contraintes de communication qui viendraient pénaliser l'interaction du système de perception avec le dispositif sensoriel. De plus, l'implantation du système sur une architecture matérielle embarquée, mais aussi dédiée, permet une distribution des ressources de calcul en adéquation avec les besoins de chaque élément du système.

La présentation de la méthodologie d'implantation débute par la description de la structure fonctionnelle adoptée. Les différents modules ainsi que leurs interactions sont exposés. La notion de détecteur actifs qui constitue l'originalité majeure de l'approche est présentée. Ensuite, un certain nombre contraintes, liées à l'approche de vision sont abordées pour justifier l'emploi d'un dispositif superviseur. Enfin, un squelette d'implantation qui s'appuie sur les fonctionnalités de la plate-forme de recherche (chap. 3) est présenté.

4.2 Structure fonctionnelle

Afin de développer la flexibilité et la faculté d'adaptation du système de vision, une organisation fonctionnelle a été établie (fig. 4.3). La chaîne de traitement est constituée d'un assemblage d'unités conceptuelles de détection appelées **Détecteurs Actifs**. Ces détecteurs sont liés à des régions d'intérêt dont les paramètres sont contrôlés pour instaurer une rétroaction sensorielle sur le capteur d'image. La

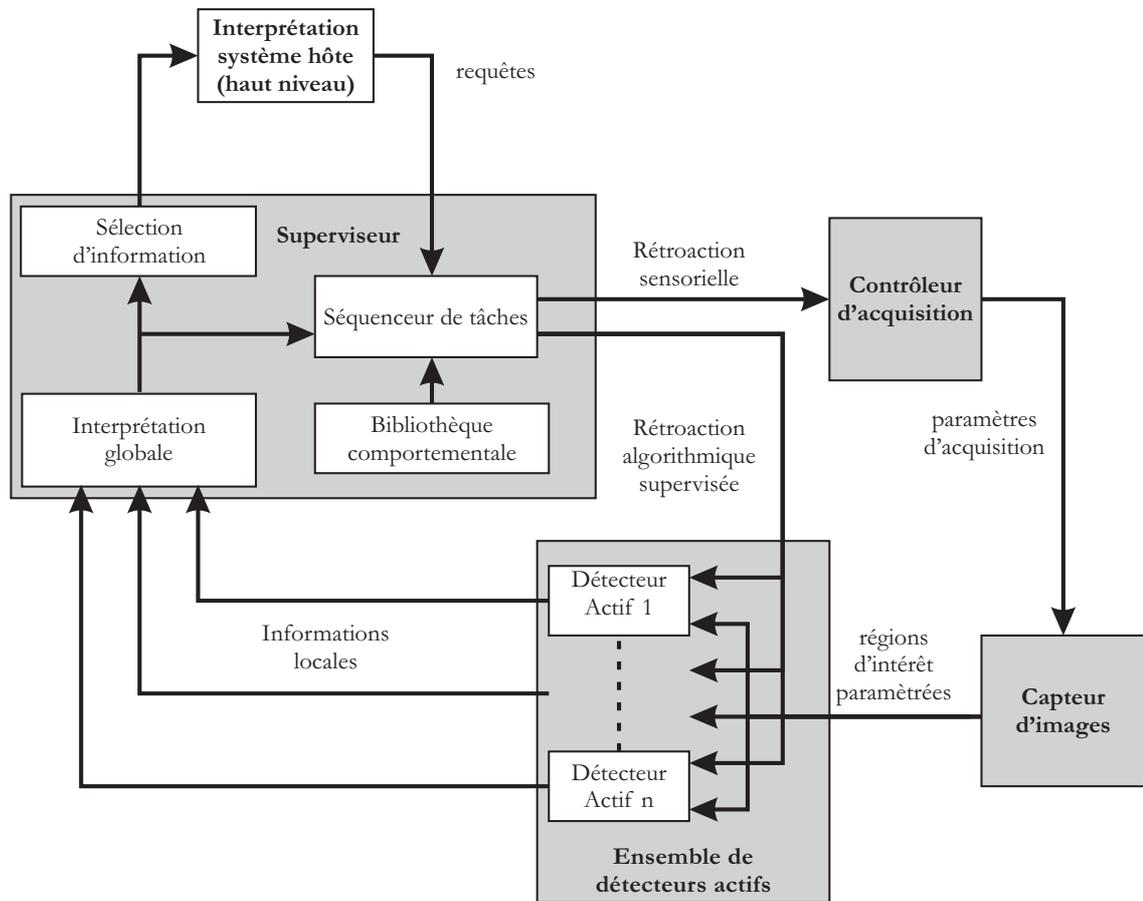


FIG. 4.3 – Schéma fonctionnel global.

rétroaction sensorielle des détecteurs est réalisée par l'intermédiaire du superviseur qui gère l'accès à la ressource critique que constitue l'imageur. Les détecteurs actifs adoptent une stratégie de détection locale à partir d'une rétroaction algorithmique interne et d'une rétroaction sensorielle. Ils fournissent un ensemble d'informations locales au dispositif superviseur. Suivant une interprétation globale de l'information, le superviseur séquence l'accès à l'information visuelle des détecteurs actifs et exerce une rétroaction algorithmique pour paramétrer la chaîne de perception. Une configuration initiale des détecteurs est spécifiée suivant la tâche mise en œuvre et les informations perçues. Suivant les requêtes du système haut niveau, le séquenceur de tâches exécute une routine prédéfinie dans une bibliothèque. Cette routine traduit une stratégie de détection pour réaliser une tâche spécifique. La bibliothèque com-

portementale doit intégrer des routines adaptées aux objectifs visuels du système hôte et les informations qui doivent lui être communiquées.

Dans cette partie, les différents modules qui constituent le schéma fonctionnel global sont détaillés. Tout d'abord, la notion de détecteur actif est définie afin de préciser l'approche modulaire adoptée pour implémenter la chaîne de perception. Ensuite, le rôle fonctionnel du dispositif superviseur et les interactions mises en jeu sont détaillés.

4.2.1 Détecteur actif

Les stratégies de perception visuelle abordent le problème de vision suivant des méthodologies ascendantes ou descendantes. La stratégie ascendante construit une représentation abstraite de l'information sensorielle pour la confronter à des connaissances *a priori* sur l'environnement. Alors que la stratégie descendante déduit de ces connaissances *a priori* les informations bas niveau à confronter avec les primitives détectées dans l'image. La stratégie de perception proposée allie ces deux méthodologies au sein d'un processus de vision intentionnelle bas niveau. Le principe est comparable à la notion de routines visuelles (chap. 1) qui définit une stratégie de perception haut niveau (fig. 4.4). Cette approche propose d'établir un processus

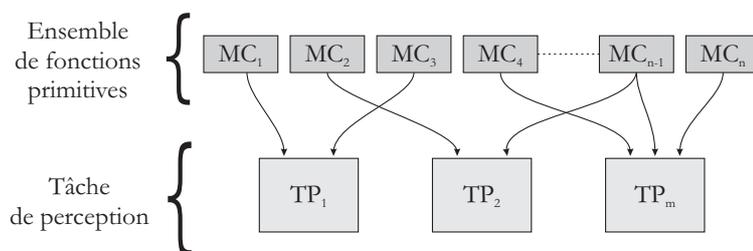


FIG. 4.4 – Décomposition du mécanisme de perception selon la notion de routines visuelles.

de vision "intelligent" en utilisant une bibliothèque de routines de perception qui définissent la manière d'exploiter un ensemble de routines primitives. Ces routines primitives sont des tâches élémentaires qui permettent l'extraction d'informations spécifiques. Pour implémenter notre système de vision bas niveau, il est proposé d'adop-

ter un raisonnement analogue. La notion de détecteur actif redéfinit les fonctions primitives comme des procédés de détection réactifs locaux basés sur le concept de vision active.

4.2.1.1 Définition

Un détecteur actif se définit comme une solution unifiée relevant d'une adéquation globale entre un comportement réactif local, une solution algorithmique et son implantation. Un détecteur est constitué d'une zone d'intérêt dans l'image qui est associée à une solution algorithmique de détection. L'aspect "logiciel" de la notion de détecteur actif est défini par une algorithmie dédiée dont l'objectif est d'extraire une information locale. Cette solution algorithmique doit exploiter une unité de traitement et une région d'intérêt sur l'imageur pour définir une stratégie de détection relative au concept de vision active. L'aspect "matériel" réside dans l'élaboration d'une architecture d'implantation dédiée à la solution "logicielle" envisagée. Dans notre cas, deux éléments matériels concrétisent cet élément conceptuel : une partie physique du FPGA et une partie de l'espace photosensible de l'imageur CMOS.

De manière générale le détecteur actif représente une **adéquation stratégie, algorithmie, architecture** et locale d'un procédé de vision élémentaire. Ainsi, un détecteur actif est caractérisé par la spécificité de sa fonction et des contraintes inhérentes à celle-ci. Malgré la spécificité de chaque détecteur actif, les trois types de rétroactions génériques énumérés précédemment sont identifiables au sein de sa structure fonctionnelle (fig. 4.5).

Un détecteur actif possède un niveau de rétroaction locale afin d'adopter un comportement autonome. À partir du contrôle des paramètres algorithmiques et sensoriels qui lui sont propres, il adapte son procédé de détection et sa zone d'intérêt dans l'image. Afin de poursuivre continuellement l'extraction d'une information spécifique, le détecteur peut exploiter l'échantillonnage spatial et quantitatif de sa région d'intérêt. Par contre, le contrôle de l'échantillonnage temporel du signal est géré par le superviseur afin de distribuer la bande passante du capteur selon les informations critiques pour le système global. Ainsi, les rétroactions sensorielles des détecteurs transitent par le dispositif superviseur qui séquence leurs accès à l'information visuelle suivant un mécanisme préemptif.

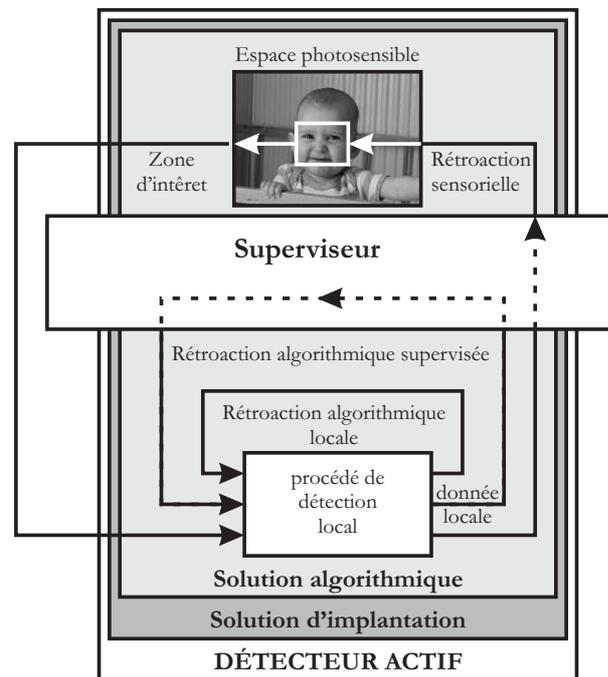


FIG. 4.5 – Schéma synoptique d'un détecteur actif.

Le dernier type de rétroaction utilisé est une rétroaction algorithmique supervisée. Ce niveau de retour de l'information permet au superviseur d'exercer un contrôle sur les détecteurs. Suivant la stratégie d'observation établie à partir d'une interprétation globale des informations extraites, le superviseur est susceptible d'exercer un contrôle sur le détecteur actif. Ainsi, lors de l'initialisation d'une tâche perception, le superviseur paramètre le détecteur suivant une configuration initiale cohérente. Ce type de rétroaction de nature discret traduit les changements stratégiques adoptés par le système de vision active globale.

4.2.1.2 Identification

Afin de mettre en œuvre la chaîne de perception, une première étape consiste à discerner l'ensemble des informations qui vont être exploitées par le système suivant ses objectifs visuels. A partir des tâches de perception identifiées, les objectifs et la dépendance des informations, un ordonnancement de tâches doit être défini. Suivant une approche active de la problématique de vision ces tâches vont être réalisées

suivant une succession de procédés d'attention, de focalisation et d'identification. L'ordonnement de ces procédés va définir une première ébauche des routines qui vont traduire les stratégies d'observation du système. Le niveau d'abstraction de l'information utilisée par les procédés d'attention, de focalisation et d'identification doit être décomposé pour déterminer l'ensemble des procédés de détection bas niveau nécessaires au système. La fonction visuelle de ces procédés va définir les contraintes qui vont caractériser un détecteur actif. Suivant les informations critiques qui conditionnent la stratégie d'observation adoptée, des priorités entre les détecteurs doivent être établies pour définir un mécanisme préemptif et identifier les points critiques d'implantation.

4.2.2 Supervision

Pour constituer une chaîne de perception à partir de détecteurs actifs qui réalisent chacun une fonction de détection élémentaire locale, il faut adjoindre au système un dispositif superviseur. Ce dispositif doit permettre au système global de faire évoluer la stratégie de perception suivant les observations réalisées par les détecteurs. Les informations disponibles doivent être centralisées pour décider du comportement à adopter suivant une interprétation globale. Afin de mettre en place une stratégie de perception, le superviseur doit disposer d'un contrôle sur les détecteurs pour permettre leurs coopérations et établir leurs dépendances fonctionnelles. Suivant les préoccupations instantanées du système, le superviseur doit sélectionner dynamiquement les détecteurs actifs pertinents et leur donner une configuration initiale. Ce contrôle définit le comportement actif et intentionnel qui traduit la stratégie de perception. Suivant la commande du superviseur, les détecteurs actifs exploitent leurs zones d'intérêt à la recherche d'une information ou d'un événement qu'ils leur sont propres. Ils adoptent un comportement autonome à partir de rétroactions algorithmiques et sensorielle locales. L'évolution simultanée des détecteurs doit adapter l'état du système suivant l'environnement observé et la spécificité de la tâche à accomplir.

4.2.2.1 Prémption des procédés de détection

De manière générale, parmi les processus de perception nécessaires à l'implantation d'un système de vision active, on distingue la succession fonctionnelle des tâches d'attention, de focalisation et d'identification (fig. 4.6).

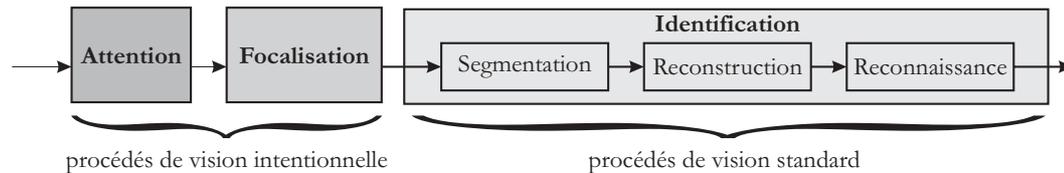


FIG. 4.6 – *Séquençement des tâches de perception d'un système de vision active.*

Les tâches d'attention et de focalisation traduisent le caractère intentionnel de l'approche active. Quant à la tâche d'identification, elle s'apparente généralement aux étapes de segmentation, reconstruction et reconnaissance présentes dans les systèmes classiques de vision. Ainsi, la chaîne de perception doit distinguer ces procédés suivant leur vocation perceptuelle. D'un point de vue fonctionnel, deux aspects majeurs vont intervenir pour identifier les priorités des procédés de perception :

- l'importance fonctionnelle,
- la validité temporelle des informations recueillies.

Généralement, les procédés d'attention et de focalisation devront disposer d'un accès plus fréquent à l'information que la plupart des processus d'identification. Par définition la vision active est considérée comme un processus continu. La fonction des procédés d'attention et focalisation est de détecter les variations de l'environnement pour déclencher une mise à jour des connaissances du système. En considérant cette observation dans la conception d'un système de vision, les priorités et la rapidité des processus de détection doivent être intégrés dans la stratégie de perception. Les opérations réalisées par certains processus nécessitent d'être garanties suivant une quantité de temps critique. Alors que le système peut admettre la prise en compte différée de certains résultats. Par exemple, la détection du mouvement d'un objet doit conduire à une réaction brusque du système, cependant l'identification de l'objet peut être différée grâce au processus de focalisation qui correspond à une fonction critique du système de perception. Ainsi, un mécanisme de prémption de

tâches doit favoriser ces fonctions critiques. La priorité d'un procédé conduit à accorder à ce dernier un accès à l'information visuelle plus fréquent et d'autre part à fournir aux autres procédés les ressources de calcul nécessaires pour exploiter l'information. Ainsi, les mécanismes préemptifs devront guider l'implantation matérielle afin d'optimiser la distribution des ressources de calcul qui va influencer l'efficacité fonctionnelle du dispositif. L'implantation matérielle idéale doit déporter le goulot d'étranglement de l'information vers le capteur afin d'optimiser l'usage de la source d'information.

De manière intrinsèque, développer la capacité de réaction du dispositif de perception conduit aux problématiques des systèmes temps réel. L'approche fonctionnelle proposée étant tributaire de cette contrainte temporelle, son implantation matérielle devient déterminante. Sans considérer l'efficacité fonctionnelle du mécanisme de préemption de tâches, la réactivité du système est limitée par ces ressources de calcul. Pour cette raison, l'approche proposée considère de manière unifier la stratégie de perception visuelle et l'implantation matérielle du système. L'objectif d'accroître la réactivité du système de perception vis à vis de l'environnement, tout en disposant d'une chaîne de traitement en adéquation avec les objectifs visuels considérés. L'efficacité du dispositif est conditionnée par la gestion des liens spatio-temporels des mécanismes d'extraction de l'information. Bien que le système soit tributaire de l'acuité de ces mécanismes, la capacité du superviseur à mettre en place une interface adaptée entre le système hôte et le dispositif sensoriel est déterminante. Ainsi, l'ordonnancement des tâches de perception est primordiale et va déterminer l'aptitude du système à résoudre le problème de vision posé.

4.2.2.2 Interaction système

L'interaction entre le système d'exploitation hôte et le superviseur va constituer un boucle de rétroaction située au sommet de la structure hiérarchique (fig. 4.3). Les requêtes systèmes vont définir un comportement stratégique initial défini par une routine de contrôle des procédés de la chaîne de traitement. Ce comportement stratégique possède un caractère global et discret qui est défini par un objectif du système hôte sous forme d'une requête. Le retour d'information est l'interprétation globale effectuée par le superviseur. Le système hôte effectue un contrôle global du capteur pour guider sa perception suivant l'interprétation haut niveau réalisée. Le

superviseur va établir une interface vis à vis de l'hôte pour définir un comportement semi-autonome du système bas niveau. Le point fondamental de la vision active est la relation établie entre le processus de vision et le capteur. La boucle de rétroaction matérialisant un comportement actif vis à vis de l'information perçue peut être implantée à divers niveaux du système. L'asservissement du système par l'information visuelle donne lieu à des stratégies d'observation discrète ou continue qui sont appliquées à un niveau global ou local du système.

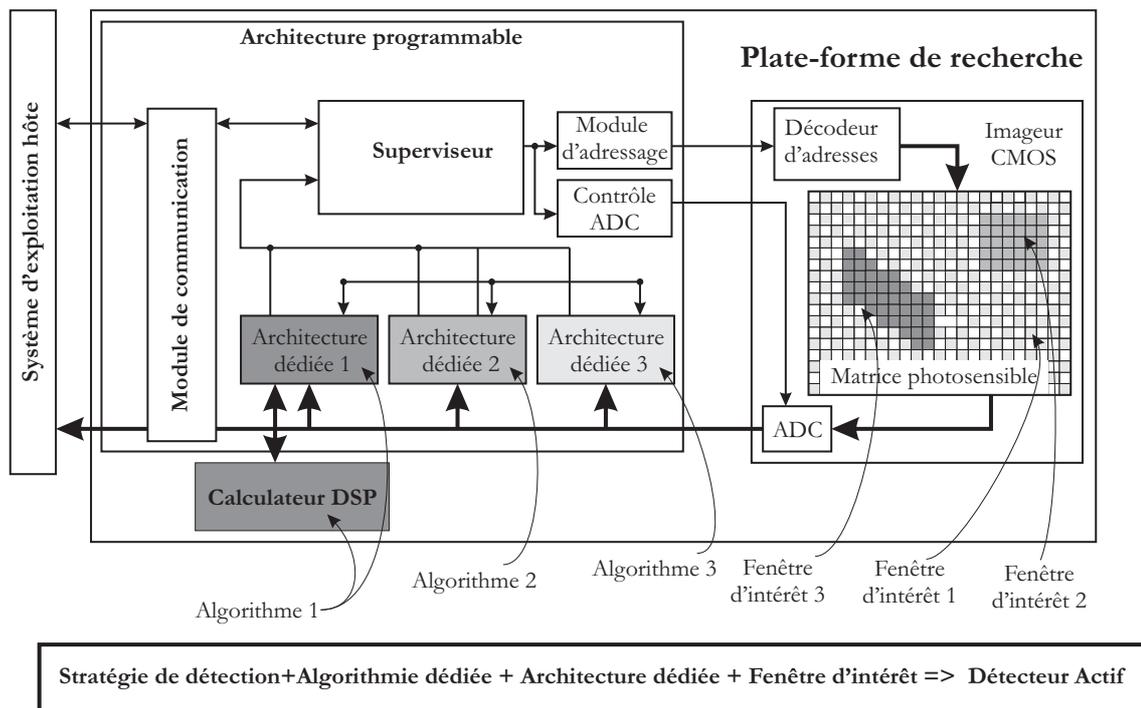


FIG. 4.7 – *Synoptique d'implantation.*

En conclusion, un schéma synoptique de l'approche fonctionnelle est proposée figure 4.7. Ce schéma représente l'implantation de notre approche fonctionnelle au sein de notre plate-forme de recherche. Les composants matériels qui constituent les détecteurs sont représentés pour faire les liens entre les zones d'intérêt sur l'imageur CMOS et l'ensemble des architectures dédiées des détecteurs actifs implantés sur la plate-forme. On note la présence du calculateur DSP qui peut être utilisé en coprocesseur pour faciliter l'implantation d'une architecture de traitement critique.

4.3 Squelette d'implantation

La méthodologie d'implantation sur FPGA s'accorde avec l'organisation modulaire proposée (chap. 2). Le parallélisme de tâches inhérent au réseau de détecteurs actifs s'accorde avec les capacités d'implantation des FPGAs. Notre plate-forme de recherche (chap. 3) a été conçue à partir de cette technologie pour permettre une implantation aisée de l'approche fonctionnelle présentée. À partir des fonctionnalités de la plate-forme un squelette d'implantation est proposé (fig. 4.8).

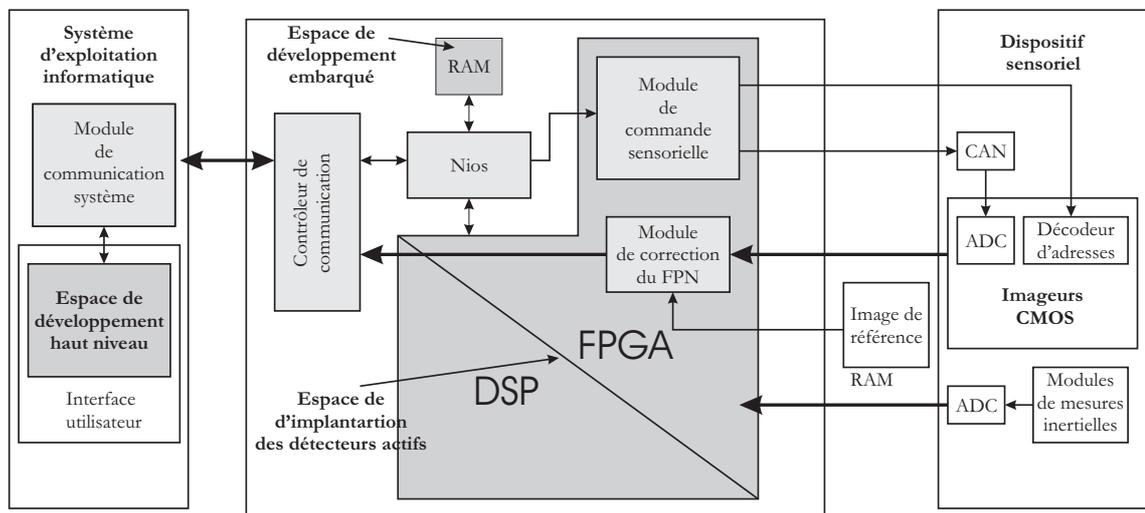


FIG. 4.8 – Architecture générale

Ce squelette d'implantation est divisé en deux catégories de modules architecturaux. Il est constitué d'une fraction (gris foncé) qui est définie suivant les spécificités d'une application et de modules communs (gris clair) qui traduisent l'architecture minimale de l'approche fonctionnelle. Ces deux parties sont discutées dans les paragraphes suivants. Dans une première partie, les solutions matérielles choisies pour implanter les modules génériques sont détaillées. Puis dans une seconde partie, les moyens de caractérisation du squelette sont abordés.

4.3.1 Les modules génériques

Quelque que soit l'application étudiée, l'approche fonctionnelle fait apparaître des modules architecturaux récurrents. Parmi ces modules, on retrouve le dispositif superviseur, le module de communication avec le système hôte et des modules de contrôle de l'imageur CMOS (fig. 4.8). Une solution d'implantation est proposée pour chacun de ces modules dans les sections suivantes.

4.3.1.1 Le dispositif superviseur

Le dispositif superviseur est basé sur la technologie ©Nios qui est une famille de processeurs embarqués dit "soft core" (fig. ??). Ces processeurs sont paramétrés de manière logicielle pour être ensuite synthétisés au sein d'un FPGA. Cet outil d'implantation permet de bénéficier d'une architecture de traitement logicielle tout en spécifiant l'architecture qui caractérise ce processeur. Bien que ce type de processeur ne possède pas une capacité de calcul élevée (~ 150 mips), il permet de définir de manière aisée une interface spécifique avec ces périphériques. Une implémentation du dispositif superviseur basée sur cette technologie, permet aussi de définir les actions du superviseur de manière logicielle tout en définissant les liens avec les détecteurs actifs de manière spécifique. Le bus de communication qui lie le processeur et ces périphériques est le bus Avalon. Ce bus autorise des modes transfert variés selon des protocoles simples. Les transferts asynchrones sont gérés par une simple ligne de contrôle. Quant aux transferts synchrones, ils sont paramétrés de manière logicielle (SOPC Builder) avant la synthèse du processeur Nios. Ce type de protocole flexible s'accorde avec la nature diverse des détecteurs actifs implémentés. Ce procédé d'interface est utilisé pour lier le superviseur Nios aux détecteurs actifs, au module de commande sensorielle et au contrôleur de communication liant le système embarqué au système hôte (fig. 4.9).

D'autre part, la technologie Nios offre une alternative pour améliorer l'implantation de traitements critiques. L'unité arithmétique et logique du processeur peut être surchargée par des instructions définies par l'utilisateur. Ces instructions sont réalisées par une architecture spécialisée instanciée dans le FPGA (module rayé fig. 4.9).

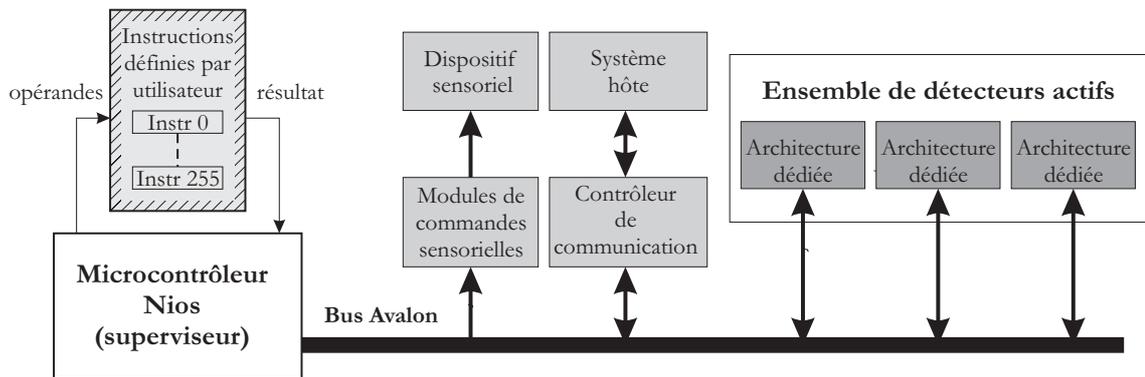


FIG. 4.9 – Les modules constituant le système de perception sont interfacés comme des périphériques d'un processeur Nios qui assure la fonction de superviseur.

4.3.1.2 Le module de communication système

Afin qu'un utilisateur puisse exploiter le système embarqué, il est nécessaire de disposer d'un moyen de communication avec un système hôte. Pour cela deux solutions matérielles ont été proposées dans le chapitre 3. Le module de communication système est chargé de gérer les flots d'entrée/sortie entre le système embarqué et l'hôte. Des données sont échangées avec une application supportée par un système d'exploitation informatique. Deux types de transferts sont mis en place. Tout d'abord, le flot de pixel contrôlé issu du capteur est transmis vers l'hôte pour visualiser les fenêtres d'intérêt sélectionnées sur la matrice CMOS. D'autre part des données sont échangées entre le dispositif superviseur (Nios) et l'hôte. Les transferts allant vers le superviseur sont des requêtes et des paramètres qui traduisent une commande de l'hôte, comme par exemple la position d'une fenêtre d'intérêt. L'architecture proposée pour gérer ces flots d'information est donnée figure 4.10.

4.3.1.2.1 Transfert de données entrantes Du côté hôte, ce module présente un flot de données entrant cadencé par l'hôte à 20 MHz. Les données entrantes sont synchronisées sur l'horloge Nios et stockées dans une mémoire double port de 128 mots de 16 bits. L'accès en lecture du Nios sur le bus Avalon n'assure pas la validité d'un mode de lecture destructive des données. En d'autres termes, les accès en lecture peuvent se produire même si aucune instruction de lecture n'est spécifiée selon les accès aux autres périphériques du Nios. Ainsi, l'utilisation d'une FIFO

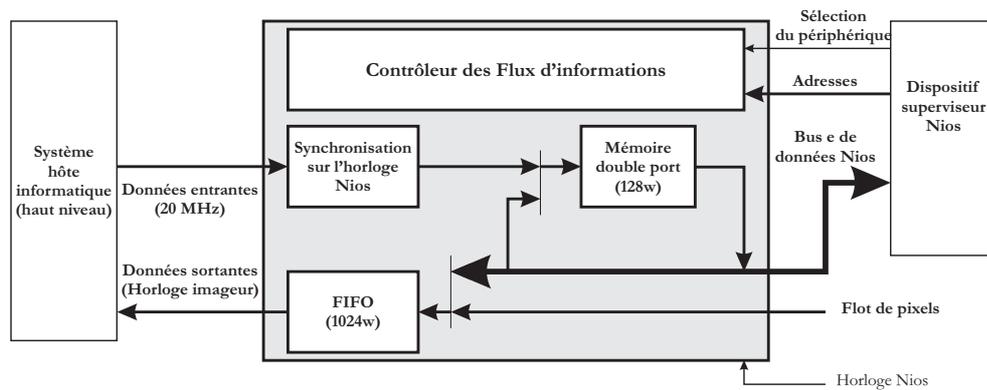


FIG. 4.10 – Architecture du module de communication système.

est prohibée et une solution basée sur une RAM interne a été mise en place. Pour s'assurer de ne pas lire plusieurs fois les données stockés, le Nios vient écrire un code dans un emplacement spécifique de la RAM qui indique si les données ont déjà été consultées. La RAM double port ainsi permet au Nios d'utiliser le Bus de lecture de manière indépendante à tout moment et de gérer ces commandes d'écriture sans perturber les transferts des données entrantes.

4.3.1.2.2 Transfert de données sortantes Le flot de données sortantes est cadencé par l'horloge de l'imageur afin d'établir un transfert direct du flot de pixels. L'objectif est d'éviter l'utilisation d'une RAM de stockage de l'image et de réduire les délais de communication au minimum. Les données sortantes transitent par un FIFO de 1024 mots de 16 bits. Le rôle de cette FIFO est de bufferiser le flot de pixels provenant de la caméra lorsque le système hôte n'est pas disponible pour recevoir l'information. Suivant une cadence d'acquisition de 8 MHz, cette FIFO autorise au système hôte une latence de $128 \mu s$ pour venir récupérer les données sortantes. Le Nios possède une adresse réservée pour venir écrire ses propres données dans cette FIFO. Ainsi, il peut adjoindre toute sorte d'informations au flot de données sortantes.

4.3.1.3 Module de commande sensorielle

Le module de commande sensorielle réalise une interface de commande qui permet de générer les fenêtres d'intérêt successivement consultées sur la matrice à partir des paramètres communiqués par le dispositif superviseur. L'architecture proposée (fig. 4.11) comporte deux bases de registres qui stockent les paramètres de transduction et les paramètres géométriques de la région d'intérêt considérée. Ainsi,

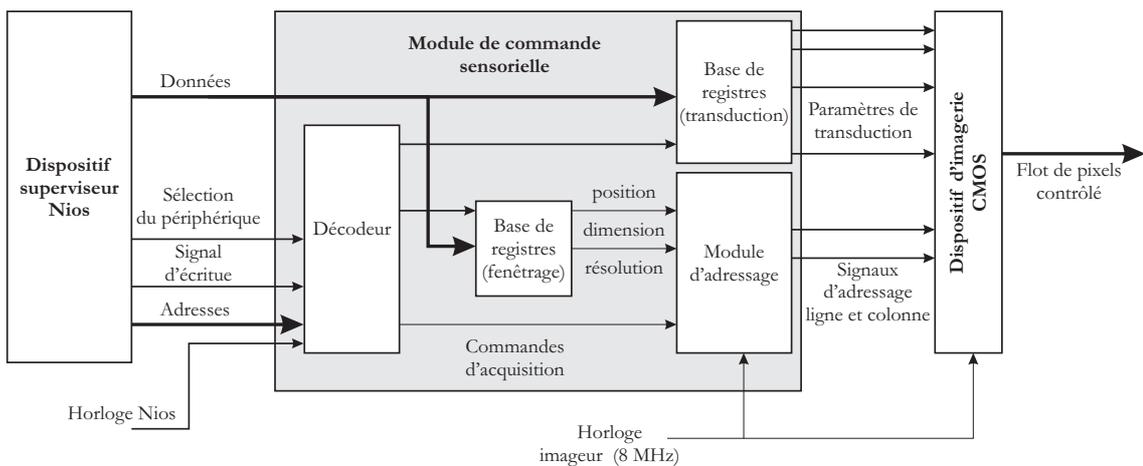


FIG. 4.11 – Architecture fonctionnelle du module de commande sensorielle.

une région est caractérisée par un ensemble de paramètres qui définissent le mode d'échantillonnage utilisée et les paramètres de transduction qui conditionne le signal issu des pixels. A partir d'un module de décodage, le superviseur Nios vient remplir les deux bases de registres et déclencher l'acquisition à partir d'une instruction d'écriture à une adresse spécifique. Suivant le mode d'échantillonnage que l'on désire réaliser, le module de fenêtrage qui génère les adresses des pixels consultés peut adopter différente architecture. Deux architectures de fenêtrage sont proposées. La première permet de consulter des zones d'intérêt rectangulaires et sous échantillonnées. Et la seconde propose une solution pour définir des zones d'intérêt inclinées.

4.3.1.3.1 Module fenêtrage sous échantillonné : La première architecture proposée génère les adresses des pixels d'une fenêtre rectangulaire définies par sa position, ses dimensions et des pas d'échantillonnage horizontale et verticale. Les adresses de ce type de fenêtrage sont générées à partir d'une architecture simple (fig.

4.12). Deux registre C_x et C_y sont utilisés pour accumuler les valeurs des résolutions

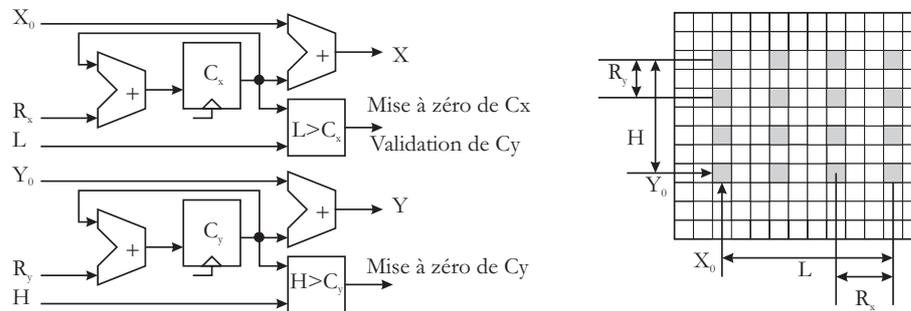


FIG. 4.12 – *Module Fenêtrage sous échantillonné*

horizontale et verticale caractérisées par les paramètres R_x et R_y . Suivant les dimensions L et H , des signaux de contrôle sont générés pour synchroniser l'évolution des adresses horizontales et verticales. Enfin, ces adresses sont translatées à partir de la position de la fenêtre caractérisée par les paramètres X_0 et Y_0 . Cette architecture élémentaire suffit pour appliquer une sélection d'information directement sur la matrice photosensible. L'utilisation de fenêtre de petites dimensions ou de faibles résolutions permet d'accroître considérablement la cadence de rafraîchissement de l'information. Par exemple, une fenêtre de 128×128 avec pas d'échantillonnage de 2 pixels comporte seulement 4096 pixels. Suivant une fréquence d'acquisition des pixels de 8 Mhz, la cadence de rafraîchissement de la région atteint environ 2000 images/s.

4.3.1.3.2 Module de fenêtrage "évolué": Afin d'exploiter différemment l'échantillonnage du signal image, un module de fenêtrage générant les adresses des pixels contenu dans un parallélogramme quelconque sur l'imageur est proposé. Cette architecture est basée sur l'algorithme de tracé de droites discrète de Bresenham [88]. Sa méthode permet de définir une droite discrète en manipulant uniquement des valeurs entières et sans utiliser de multiplications. Cet algorithme est une méthode itérative basée sur l'estimation d'une erreur ε (fig. 4.13) qui permet de détecter les ordonnées discrètes les plus proches de la droite réelle considérée. L'erreur ε caractérise la distance entre le point discret courant et l'intersection de la droite réelle avec l'axe vertical aligné sur le point. Les abscisses sont considérées une à une et lorsque ε est supérieure à $1/2$, l'ordonnée du prochain point de la droite discrète est

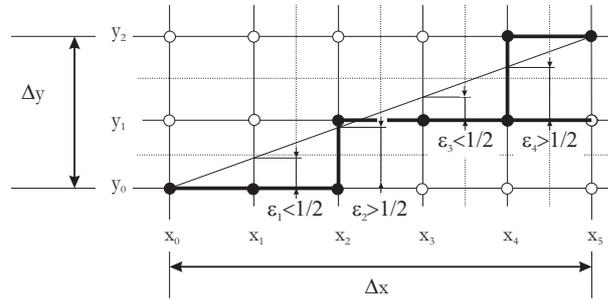


FIG. 4.13 – Tracé d'une droite discrète selon l'algorithme de Bresenham.

incrémentée. Durant le processus itératif, la valeur ε est mise à jour selon l'équation suivant :

$$\varepsilon_{x+1} = \varepsilon_x + \frac{\Delta y}{\Delta x} - 1 \quad \text{sinon} \quad \varepsilon_{x+1} = \varepsilon_x + \frac{\Delta y}{\Delta x}$$

ou les valeurs Δy et Δx définissent le segment à discrétiser. Afin de manipuler uniquement des valeurs entières, on réalise le changement de variable $\epsilon = 2 \cdot \Delta x \cdot \varepsilon$ qui permet d'écrire l'algorithme suivant :

1. Initialisation :
 $y = y_0; x = x_0; \epsilon = 0; \Delta x = x_5 - x_0; \Delta y = y_5 - y_0$
2. Processus itératif :
 $x = x + 1; \epsilon = \epsilon + 2 \cdot \Delta y;$
 Lorsque $\epsilon > \Delta x$ alors $y = y + 1; \epsilon = \epsilon - 2 \cdot \Delta x;$
3. Test :
 Tant que $x < x_5$ on recommence le processus itératif.

La technique précédente permet de générer les adresses d'un ligne discrète de pixels sur la matrice photosensible. Cependant, tel qu'il est présenté l'algorithme se limite au tracé de droites dont la pente est positive et inférieure à un. Pour généraliser cette méthode à toutes les droites, il est nécessaire d'introduire quelques modifications. Pour étendre le processus itératif aux pentes supérieures à un, on peut dupliquer le processus afin de faire évoluer les deux coordonnées simultanément. De plus en contrôlant, l'incrément et la décrémentation de x et y , on étend la méthode à l'ensemble des vecteurs du plan discret. L'algorithme final est le suivant :

1. Initialisation :
 $y = y_0; x = x_0; \epsilon_{y/x} = 0; \epsilon_{x/y} = 0; \Delta x = x_5 - x_0; \Delta y = y_5 - y_0$

- $s_x=1$ si $\Delta x < 0$ sinon 0; $s_y=1$ si $\Delta y < 0$ sinon 0;
 $\Delta x = |\Delta x|$; $\Delta y = |\Delta y|$;
 2. Processus itératif :
 $\epsilon_{y/x} = \epsilon_{y/x} + 2 \cdot \Delta y$; $\epsilon_{x/y} = \epsilon_{x/y} + 2 \cdot \Delta x$;
 Lorsque $\epsilon_{y/x} > \Delta x$ alors $y = y + (-1)^{s_y}$; $\epsilon_{y/x} = \epsilon_{y/x} - 2 \cdot \Delta x$;
 Lorsque $\epsilon_{x/y} > \Delta y$ alors $x = x + (-1)^{s_x}$; $\epsilon_{x/y} = \epsilon_{x/y} - 2 \cdot \Delta y$;
 3. Test :
 Tant que $x \neq x_5$ (ou $y \neq y_5$) on recommence le processus itératif.

L'architecture matérialisant cet algorithme utilise deux registres accumulateur pour faire évoluer les quantités $\epsilon_{y/x}$ et $\epsilon_{x/y}$ (fig. 4.14).

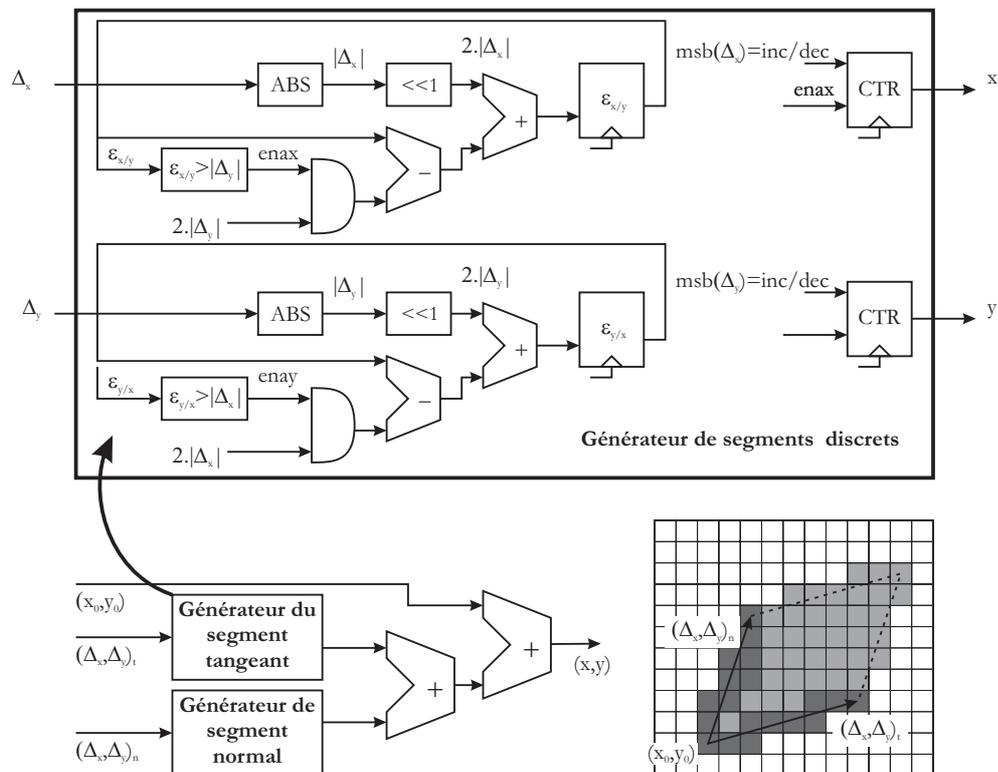


FIG. 4.14 – Architecture d'adressage d'une fenêtre "évoluée".

Un comparateur permet de définir quelle quantité doit être accumulée et de contrôler l'évolution de deux compteurs qui génèrent les coordonnées. Suivant les signes des quantités Δx et Δy , ces compteurs sont incrémentés ou décréments.

Cette architecture permet le calcul des adresses d'une fenêtre d'intérêt formant un parallélogramme quelconque. Les lignes de points discrets remplissant la fenêtre sont paramétrées par trois vecteurs. Un premier vecteur (x_0, y_0) définit la position de la fenêtre dans l'espace. Le vecteur tangentiel $(\Delta_x, \Delta_y)_t$ et le vecteur normal $(\Delta_x, \Delta_y)_n$ caractérisent le parallélogramme. $(\Delta_x, \Delta_y)_t$ représente la direction d'acquisition des pixels et $(\Delta_x, \Delta_y)_n$ le changement de ligne d'acquisition. L'architecture développée manipule uniquement des données entières et permet une implantation efficace au sein du FPGA de la plate-forme de recherche :

- 10 bits => $\sim 100\text{Mhz}$, 343/57120 LEs(0.6%).

4.3.1.4 Module de correction du FPN

Le dernier module générique indispensable pour implanter un système est le module de correction du bruit spatiale fixe du capteur CMOS (chap. 2). Le flot image brut fourni par l'imageur CMOS présente un bruit important du à la disparité des réponse de chaque pixel. Pour corriger ce bruit, une image dite de référence est stockée dans une mémoire externe (fig. 4.15). Ces valeurs sont estimées en soumettant

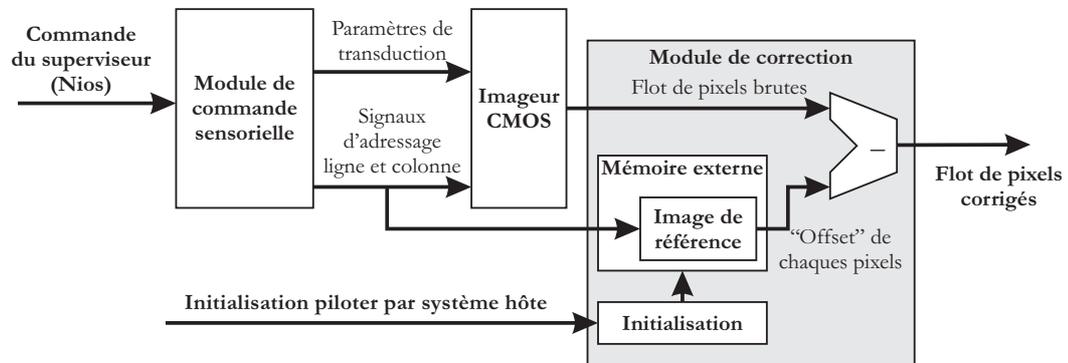


FIG. 4.15 – *Module de correction du bruit spatiale fixe du capteur CMOS.*

le capteur à un éclairage homogène moyen. L'image recueillie I est utilisée pour caractériser "offset" de chaque pixel k . Une mesure différentielle est réalisée sur chaque valeur de pixel $I(k)$ de l'image suivant l'équation :

$$I_{ref}(k) = I(k) - \frac{1}{N} \sum_{i=0}^{N-1} I(k)$$

L'image I_{ref} constituée par l'ensemble de ces différences d'"offset" forme l'image de référence. Un chemin de données est intégré au système pour permettre à l'hôte d'écrire cette image dans une mémoire externe. Durant l'acquisition, les signaux d'adresses sont utilisés pour soustraire ces valeurs d'"offset" aux pixels qui leurs correspondent. Ce module permet d'uniformiser artificiellement la réponse des pixels pour disposer d'un flot de pixels "utilisables".

4.3.2 Espaces de caractérisation du squelette

Les divers modules présentés précédemment constituent le squelette d'implantation basique. Pour instancier une application spécifique, l'utilisateur doit alors définir les modules de l'espace dédié (fig. 4.8). La fraction du squelette d'implantation qui constitue l'espace d'adéquation à une application donnée comporte trois niveaux de caractérisation :

- un espace d'implantation matériel,
- un espace de développement embarqué et
- un espace de développement haut niveau.

Ces trois niveaux spécifient les interactions entre le capteur, la chaîne de traitement, le superviseur et le système hôte.

4.3.2.1 Espace d'implantation matériel

Cette espace d'implantation est réservé à la définition de détecteurs actifs qui se matérialisent sous la forme d'architecture dédiée à l'extraction d'une information locale. Ainsi, les architectures utilisées ne peuvent être spécifiées que dans un cas concret. La plate-forme de recherche présente un espace d'implantation librement configurable grâce au composant à architecture programmable (Stratix). La technologie FPGA permet d'implanter une architectures dédiée pour chacun de ces détecteurs actifs. En plus des fonctionnalités inhérente au FGPA (Nios, blocs DSP, RAM TriMatrix) et des divers modules de mémorisation externes (SRAM, SDRAM), l'utilisateur dispose d'un processeur DSP pour définir une architecture hétérogène adaptée à ses besoins. Selon l'approche fonctionnelle proposée, les flots de données à mettre en place sont conditionnée par une rétroaction algorithmique locale, une

rétroaction algorithmique supervisée et une rétroaction sensorielle et l'information locale recherchée. La rétroaction algorithmique locale doit être intégrée à l'architecture du détecteur. La rétroaction sensorielle qui le lie à une zone d'intérêt sur l'espace photosensible doit être transmise au superviseur. Une rétroaction algorithmique doit être instaurée pour permettre au superviseur de donner une configuration initiale au détecteur. Le dernier flot de données à considérer est le résultat fourni par le détecteur qui est transmis au superviseur.

4.3.2.2 Espace de développement embarqué

Le second niveau de caractérisation de caractérisation du système de vision (active) est l'espace de développement embarqué. Cette espace est réservé à la définition de la conduite qu'adopte le superviseur pour réaliser un tâche spécifique. L'interprétation globale des informations que réalise le superviseur peut être implémentée de manière logicielle ou faire appel à une architecture co-traitement implémentée sur le FPGA et/ou déportée vers le DSP. La définition d'instructions "spécifiques" (chap. 3) peut alors être utilisée pour réaliser les points critiques des traitements.

Selon l'approche fonctionnelle, un ordonnancement de procédés d'attention, de focalisation et d'identification est déterminé pour satisfaire les contraintes de l'application envisagée et des priorités entre les détecteurs sont identifiées suivant leur importance fonctionnelle et la validité temporelle de leurs résultats. Suivant ces considérations, l'accès à la ressource critique que représente l'imageur est partagé suivant des règles préemptives.

4.3.2.3 Espace de développement haut niveau

Sans considérer les spécificités d'un système haut niveau qui exploite le système embarqué, il nécessaire de définir une interface entre les deux systèmes suivant les caractéristiques de l'applications. Les échanges de commandes, de paramètres et de résultats sont définis pour utiliser au mieux les moyens disponibles entre la plateforme et le système d'exploitation hôte.

4.3.3 Fenêtrage pour implémenter un suivi

Cette partie illustre un suivi de motif rapide à partir d'un fenêtrage réalisé sur un imageur CMOS ???. Dans cette expérimentation, une corrélation entre un motif de référence et le motif couramment acquis est réalisée pour estimer la translation qui permet de mettre jour la position de la région d'acquisition courante sur l'imageur.

Une séquence d'images est proposée 4.16 pour illustrer les résultats obtenus. Pour que l'expérimentation puisse être visuellement interprétée, la totalité de la matrice CMOS est enregistrée toutes les 400 acquisitions. Cependant, en réalité uniquement la zone marquée en rouge sur les images est adressée. Suivant ces conditions expérimentales, on obtient pour un région d'intérêt de 32×32 , une cadence de suivi de 550 "images"/s. La corrélation est réalisée dans le domaine spectrale à partir d'un algorithme de transformée de Fourier rapide. Le calcul de translation est implémenté sur un PC standard (PC AMD Athlon 1700). Pour donner une idée des cadences de suivi, les performances temporelles obtenues pour quatre tailles de fenêtre sont répertoriées.

Fenêtre	16x16	32x32	64x64	128x128
Acquisition.	200 μ s	570 μ s	2 ms	8 ms
Correction FPN.	450 μ s	1.8 ms	8.6 ms	78 ms
Suivi. rate	2220 f/s	550 f/s	117 f/s	12 f/s

4.4 Perspectives

L'objet du squelette d'implantation proposé est de démontrer l'adéquation entre l'approche fonctionnelle choisie et la plate-forme de recherche proposée. Suivant la plate-forme de recherche conçue, l'objectif sera alors la création d'une bibliothèque de détecteurs actifs génériques et d'établir un outil capable d'extraire des squelettes d'implantation spécifiques aux systèmes de vision active.

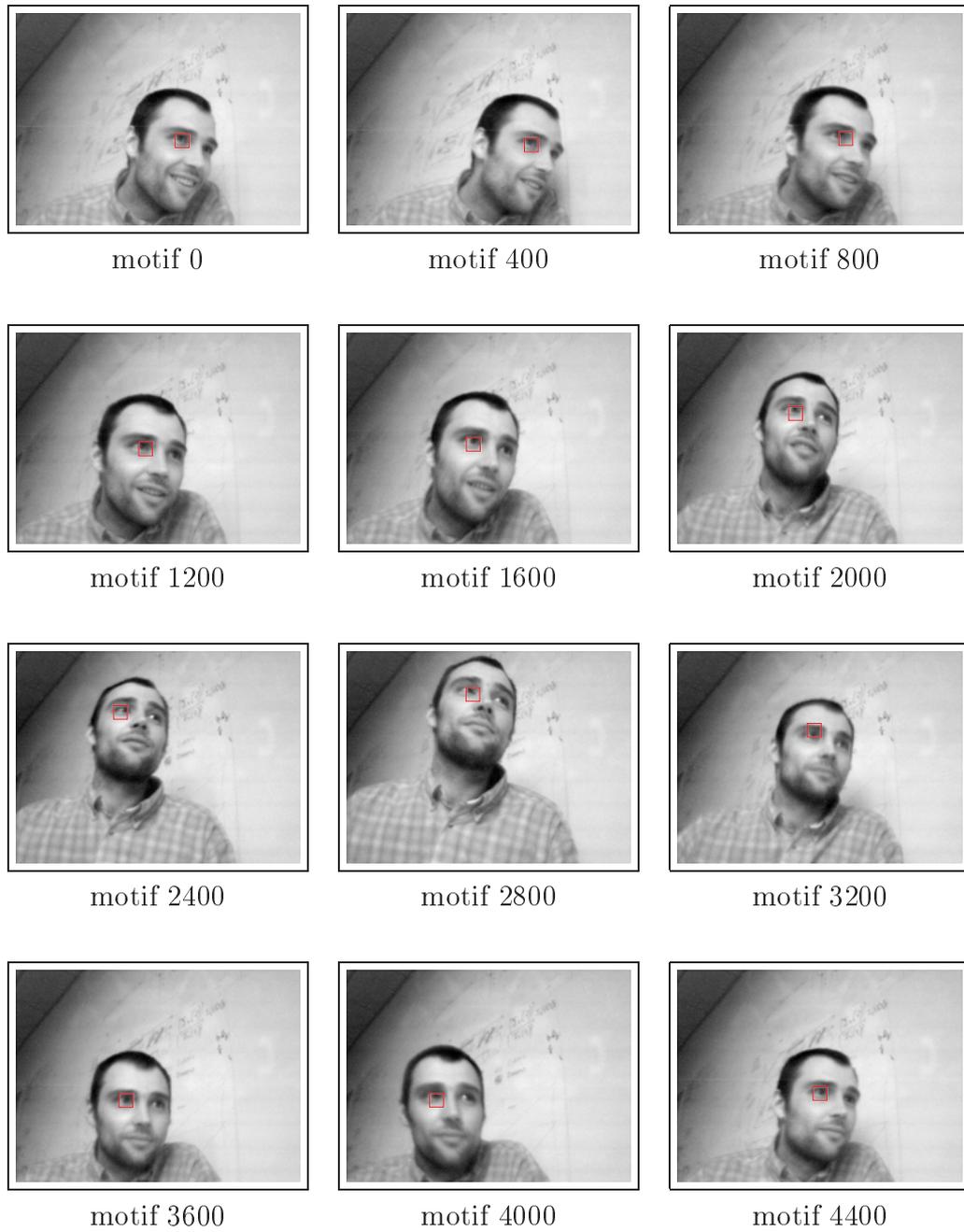


FIG. 4.16 – Suivi d'un motif de niveau de gris (32×32) suivant temps de 720 ms entre deux acquisitions consécutives.

Chapitre 5

Exemples de détecteurs actifs

Afin d'illustrer la méthodologie d'implantation et l'approche fonctionnelle adoptées, deux détecteurs actifs sont présentés dans ce chapitre. Ils exploitent tout deux l'imageur CMOS de manière différente pour définir des procédés de vision active originaux.

Le premier détecteur actif instaure une rétroaction sensorielle sur les paramètres de transduction du flot de pixels. Il vise une adéquation dynamique du dispositif de transduction avec le signal d'éclairement pour permettre au système de s'adapter à des conditions de luminosité changeantes. La qualité du signal fourni par le capteur est un point qui conditionne toute la chaîne de perception. Cette faculté d'adaptation est essentielle pour définir une application dans un environnement non contrôlé (réel). Selon cet objectif, le détecteur actif présenté asservit les paramètres de transduction du flot de pixels à partir du signal d'éclairement. Une mesure de la dynamique de la région d'intérêt associé au détecteur est utilisé pour instaurer une rétroaction sensorielle adéquate. Ce détecteur ne fournit pas directement une information au système mais accroît la capacité de détection d'un second qui partage cette région d'intérêt.

Le second détecteur proposé illustre une rétroaction sensorielle appliquée sur l'échantillonnage spatio-temporel du signal. La fonction de détection considérée est l'estimation du déplacement d'un motif de référence sur la matrice. L'objectif du détecteur est de contraindre la position de sa fenêtre d'intérêt sur le motif. Ce détecteur offre une solution pour permettre au système de focaliser son attention sur

une région d'intérêt locale de son environnement. Il démontre l'intérêt de l'association du concept de vision active et du mode de lecture des imageurs CMOS pour concevoir un système de vision rapide. Dans le cas de ce détecteur, la rétroaction algorithmique supervisée (chap. 4) consiste à définir les paramètres d'une région d'intérêt initiale. Ensuite, le détecteur adopte un comportement "autonome" à partir d'une rétroaction sensorielle.

Ces deux expérimentations n'ont pas la prétention de démontrer l'ensemble des aspects fonctionnelles de l'approche proposée. Cependant, elles illustrent une utilisation du squelette d'implantation pour mettre en œuvre une tâche spécifique à partir des postulats de la vision active et d'un imageur CMOS.

5.1 Détecteur actif anti-éblouissement

Ce détecteur actif réalise une adaptation du dispositif de transduction avec le signal d'éclairage pour permettre au système de s'adapter à des conditions de luminosité changeantes. La fonction de détection utilise la mesure de la dynamique des niveaux de gris de sa fenêtre d'intérêt qui ne pose pas de réelles difficultés d'implantation. Ici, le problème abordé est le contrôle de la chaîne de transduction du signal visuel suivant les mesures effectuées dans l'image. La chaîne de transduction doit être asservie suivant les caractéristiques locales de chaque zone d'intérêt considérées par les détecteurs actifs. Chacun doit posséder ces propres paramètres de transduction susceptibles d'évoluer dynamiquement. La difficulté consiste à élaborer un schéma d'asservissement qui agit de manière cohérente sur les divers points paramétrables de la chaîne de transduction du signal.

Afin d'exploiter la caractéristique de réponse logarithmique des pixels de l'imageur CMOS, une méthode de calibration dynamique du capteur est proposée. Son objectif est de faire évoluer l'espace de conversion numérique sur la caractéristique de réponse des pixels pour optimiser le contraste de chaque fenêtre d'intérêt (fig. 5.1). Il est proposé d'asservir les paramètres de transduction de l'imageur CMOS suivant les mesures des minima et maxima des niveaux de gris de chaque zone d'intérêt. Cet asservissement nécessite une analyse de la chaîne de transduction de l'imageur afin d'établir une commande de ces paramètres. Ensuite, un procédé d'asservissement est établi. Enfin, des résultats expérimentaux sont proposés.

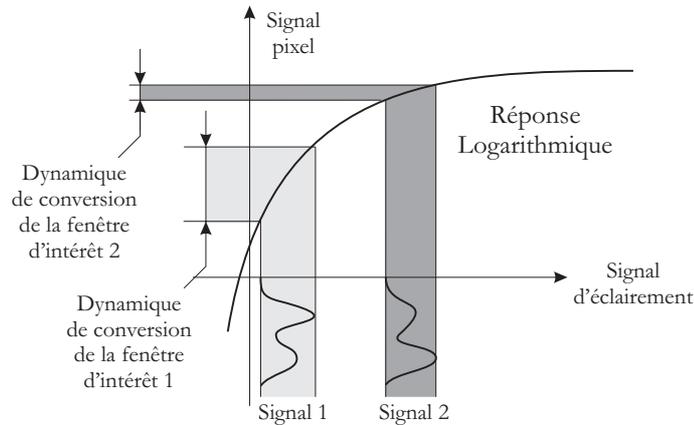


FIG. 5.1 – Adaptation de la conversion du signal visuel de chaque fenêtre d'intérêt.

5.1.1 Analyse de la chaîne de transduction

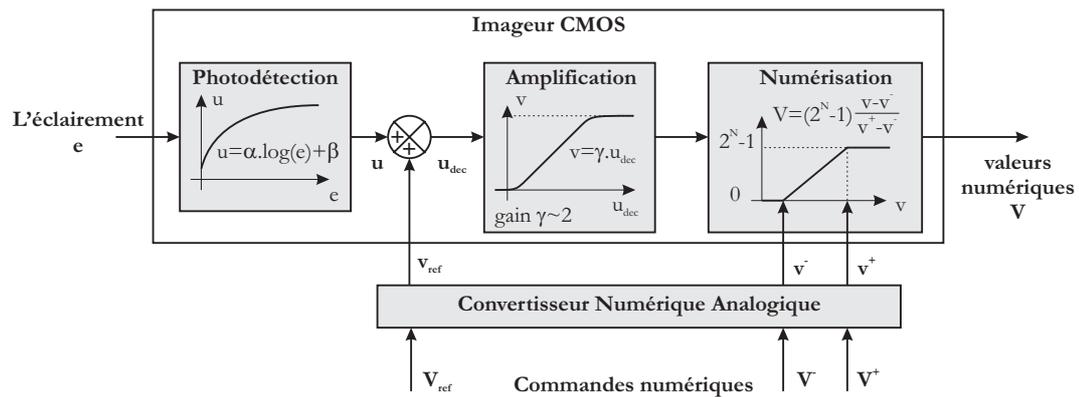


FIG. 5.2 – Chaîne de transduction de l'imager.

La chaîne de transduction de l'éclaircement en une valeur numérique comporte trois étapes majeures (fig. 5.2) :

- la photodétection,
- l'amplification,
- la numérisation.

Ces trois étapes sont étudiées pour définir quels paramètres vont prendre part à l'asservissement et quelles consignes devront être utilisées.

La première étape de la chaîne de transduction (fig. 5.2) est la photodétection. Le signal d'éclairement e est traduit en une tension analogique u suivant la caractéristique de réponse des pixels de l'imageur. Suivant deux paramètres α et β , la réponse du pixel est caractérisée par le modèle suivant :

$$u = \alpha \cdot \log(e) + \beta \quad (5.1)$$

La seconde étape de la chaîne de transduction (fig. 5.2) est l'amplification. Au signal analogique u issu des pixels, est additionnée une tension continue v_{ref} qui décale la valeur moyenne du signal visuel par rapport à la caractéristique d'un amplificateur analogique. Cette tension v_{ref} analogique est commandée suivant la valeur numérique V_{ref} à partir d'un convertisseur numérique/analogique. La tension u_{dec} qui résulte de l'addition de v_{ref} et de u constitue le signal d'entrée d'un amplificateur dont le gain γ ($\gamma \simeq 2$) est fixe. Dans la zone de linéarité de l'amplificateur (fig. 5.3),

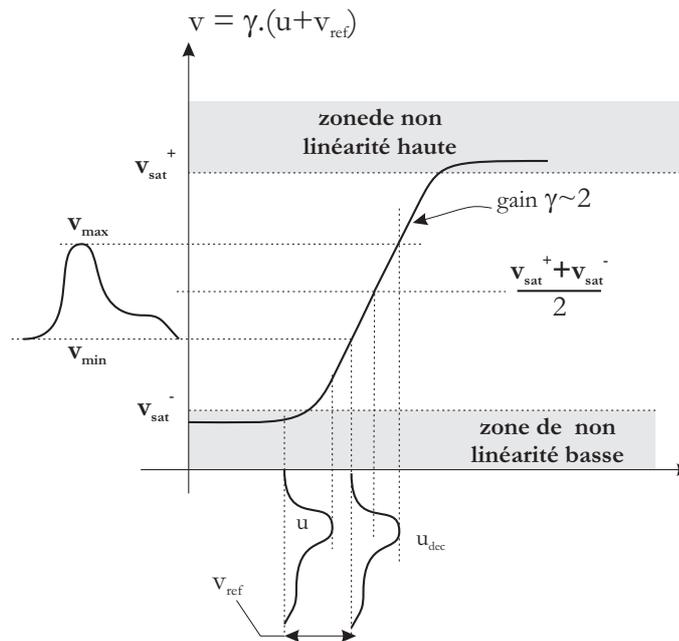


FIG. 5.3 – Caractéristique de l'amplificateur analogique de l'imageur.

la tension amplifiée v est caractérisée par l'équation :

$$v = \gamma u_{dec} = \gamma (u + v_{ref}) \quad (5.2)$$

Lorsque la condition de linéarité haute n'est pas remplie ($\gamma(u + v_{ref}) > v_{sat+}$), on a $v \simeq v_{sat}^+$. De la même façon, lorsque la condition de linéarité basse n'est pas remplie ($\gamma(u + v_{ref}) < v_{sat}^-$), on a $v \simeq v_{sat}^-$. Pour éviter la saturation de l'amplificateur, le paramètre v_{ref} doit être asservi pour maintenir la dynamique du signal u au centre de la caractéristique d'amplification. Afin de limiter la saturation de l'amplificateur, l'erreur ε_{vref} doit être minimisée :

$$\varepsilon_{vref} = \frac{v_{max} + v_{min}}{2} - \frac{v_{sat}^+ + v_{sat}^-}{2} \quad (5.3)$$

La dernière étape de la chaîne de transduction (fig. 5.2) est la numérisation. Le signal analogique amplifié v est envoyé vers un convertisseur analogique/numérique de N bits dont la bande de conversion est paramétrable. Sa limite haute v^+ et sa limite basse v^- dépendent de deux tensions analogiques de commande. La première v_{cm} fixe le centre de la bande conversion et la seconde v_{bg} définit sa largeur. Les relations liant v^+ , v^- et v_{cm} , v_{bg} sont les suivantes :

$$v^+ = v_{cm} + \frac{2}{5}v_{bg} \quad v^- = v_{cm} - \frac{2}{5}v_{bg} \quad (5.4)$$

Les tensions analogiques v_{cm} et v_{bg} sont commandées suivant les valeurs numériques V_{cm} et V_{bg} à partir de deux convertisseurs numérique/analogique. Suivant ces tensions, la bande conversion est asservie pour exploiter de manière optimale la dynamique de numérisation de N bits. Dans la zone de linéarité (fig. 5.4), l'équation liant la tension analogique v et la valeur numérique qui lui correspondent est :

$$V = (2^N - 1) \frac{v - v^-}{v^+ - v^-} \quad (5.5)$$

L'exploitation optimale de la bande conversion conduit à asservir v^- et v^+ respectivement sur les valeurs v_{min} et v_{max} du signal (analogique) amplifié v . Suivant cette objectif, les erreurs à minimiser sont :

$$\epsilon_{v+} = v^+ - v_{max} \quad \epsilon_{v-} = v^- - v_{min}$$

Les mesures disponibles pour réaliser les asservissements de v_{ref} , v^- et v^+ sont les extremum du signal numérique fourni par le capteur. Ces valeurs sont notées V_{max} et V_{min} . La minimisation des erreurs ϵ_{v+} et ϵ_{v-} tend à faire converger les tensions v^+ et v^- vers les limites de linéarité de la caractéristique du convertisseur. Hors

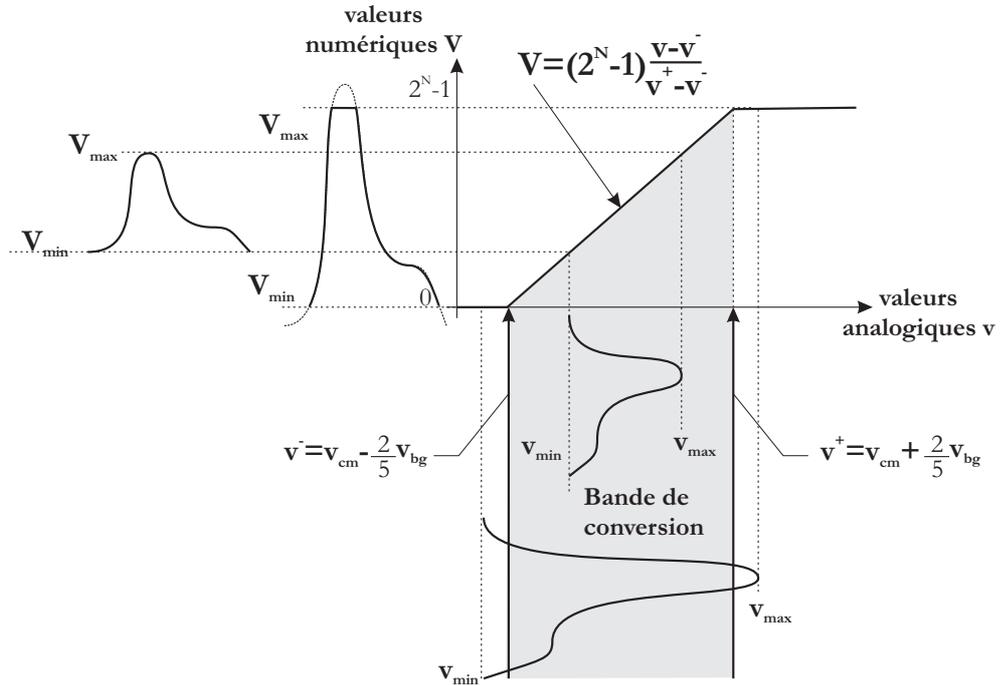


FIG. 5.4 – Caractéristique du convertisseur analogique numérique.

de la zone de linéarité, les valeurs V_{max} et V_{min} ne sont plus représentatives des tensions analogiques v_{max} et v_{min} (fig. 5.4) lorsque la valeur maximale du signal analogique v_{max} est supérieure à v^+ ($V_{max} < (2^N - 1) \frac{v_{max} - v^-}{v^+ - v^-}$) ou lorsque sa valeur minimale v_{min} est inférieure à v^- ($V_{min} > (2^N - 1) \frac{v_{min} - v^-}{v^+ - v^-}$). La mesure effectuée sur le système est faussée et les asservissements de v_{ref} , v^- et v^+ sont alors efficace. Ainsi, il est nécessaire de contraindre le système pour assurer la validité de l'équation 5.5. Le convertisseur doit être utilisé dans la zone linéaire de sa caractéristique afin d'effectuer des mesures pertinentes de v_{max} et v_{min} . Une marge Δv est adjointe à la consigne d'asservissement au détriment de la dynamique de conversion numérique pour assurer un fonctionnement linéaire. Les erreurs à minimiser pour asservir le convertisseur analogique numérique deviennent :

$$\epsilon_{v+} = (v^+ - \Delta v) - v_{max} \quad \epsilon_{v-} = (v^- + \Delta v) - v_{min} \quad (5.6)$$

5.1.2 Le procédé d'asservissement

Dans un souci de clarté, le procédé d'asservissement est établi en fonction des tensions analogiques v^+ , v^- et v_{ref} bien que celles-ci ne soient pas physiquement les entrées du système. Les équations d'asservissement réelles seront calculées par la suite le raisonnement établi. Selon les 3 consignes établies précédemment (equ. 5.6, 5.3), un schéma d'asservissement de la chaîne de transduction de l'imageur est proposé figure 5.5. L'ensemble du dispositif imageur qui inclut l'amplificateur et le

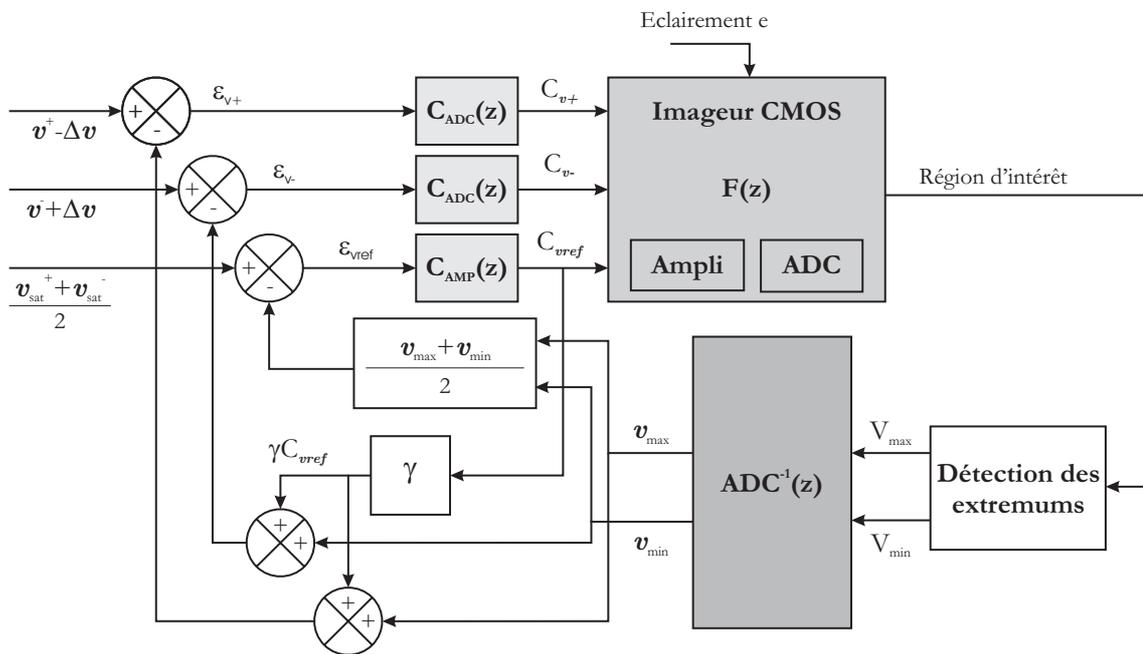


FIG. 5.5 – Schéma d'asservissement de la chaîne de transduction de l'imageur.

convertisseur analogique/numérique intégrés est représenté par la fonction transfert $F(z)$. Le signal d'éclairage e est considéré comme une perturbation appliquée à ce dispositif. La mesure disponible à la sortie de l'imageur est une région d'intérêt ou plus précisément une liste de valeur numérique représentative de la valeur de chaque pixel. La première étape consiste à extraire les extremums V_{max} et V_{min} de ce signal numérique. Les mesures à comparer aux consignes du procédé sont les valeurs analogiques v_{max} et v_{min} qui correspondent à ces extremums. Le modèle inverse ADC^{-1} est utilisée pour calculer ces valeurs. Suivant l'équation 5.5, cette

fonction de transfert est caractérisée par la relation :

$$v(n) = \frac{V(n)}{2^N - 1} ((v^+(n) - v^-(n)) + v^-(n)) \quad (5.7)$$

Suivant la mesure $\frac{v_{max} + v_{min}}{2}$ la correction appliquée sur v_{ref} est $C_{Amp}(z) \cdot \varepsilon_{v_{ref}}$. Il faut noter que cette correction influence l'asservissement du convertisseur analogique/numérique. Les modifications de la tension v_{ref} de l'amplificateur se répercutent sur la tension v , donc sur l'asservissement de v^+ et v^- . Une variation Δ_{ref} de la tension v_{ref} se traduit à la sortie de l'amplificateur par la variation $\gamma \cdot \Delta_{ref}$ (eq. 5.2). Ainsi, la dépendance de v vis à vis de v_{ref} doit être intégrée dans les procédés d'asservissement du convertisseur. Une compensation de cette perturbation est réalisée en additionnant aux mesures v_{max} et v_{min} la correction $\gamma \cdot C_{v_{ref}}$. Cette compensation anticipe l'effet de la correction qui est appliquée sur v_{ref} pour l'intégrer dans l'asservissement des tensions v^- et v^+ .

5.1.2.1 Équations d'évolution des tensions analogiques

Suivant le correcteur $C_{AMP}(z)$ et l'erreur $\varepsilon_{v_{ref}}$, l'équation d'évolution de la tension de commande v_{ref} est :

$$v_{ref}(n+1) = v_{ref}(n) - C_{AMP}(n) \cdot \left(\frac{v_{max}(n) + v_{min}(n)}{2} - \frac{v_{sat}^+ + v_{sat}^-}{2} \right) \quad (5.8)$$

Les procédés d'asservissement des tensions v^+ et v^- étant de même nature, les correcteurs qui leur sont associés sont supposés identiques. Suivant le correcteur C_{ADC} , les équations d'évolution des tensions v^+ et v^- sont :

$$v^+(n+1) = v^+(n) - C_{ADC}(n) \cdot ((v^+(n) - \Delta v) - (\gamma \cdot C_{ref}(n) + v_{max}(n))) \quad (5.9)$$

$$v^-(n+1) = v^-(n) - C_{ADC}(n) \cdot ((v^-(n) + \Delta v) - (\gamma \cdot C_{ref}(n) + v_{min}(n))) \quad (5.10)$$

avec

$$C_{ref}(n) = C_{AMP}(n) \cdot \left(\frac{v_{max}(n) + v_{min}(n)}{2} - \frac{v_{sat}^+ + v_{sat}^-}{2} \right)$$

Les entrées réelles de l'imageur sont les tensions analogiques v_{cm} , v_{bg} et v_{ref} . Suivant les résultats précédents et la relation 5.4, les équations d'évolution de v_{cm}

v_{bg} sont calculées par :

$$v_{cm}(n+1) = v_{cm}(n) - C_{ADC}(n) \cdot \left(v_{cm}(n) - \left(\gamma \cdot C_{ref}(n) + \frac{v_{max}(n) + v_{min}(n)}{2} \right) \right) \quad (5.11)$$

$$v_{bg}(n+1) = v_{bg}(n) - C_{ADC}(n) \cdot \left(v_{bg}(n) - \frac{5}{2} \left(\Delta v + \frac{v_{max}(n) - v_{min}(n)}{2} \right) \right) \quad (5.12)$$

5.1.2.2 Équations d'évolution numériques

Les tensions v_{cm} , v_{bg} et v_{ref} sont commandées numériquement par le biais de convertisseurs numérique/analogique de N bits. Suivant la tension de référence du convertisseur v_{alim} , les valeurs numériques caractérisant les tensions analogiques de commande sont notées :

$$V_{ref} = \frac{(2^N - 1)}{v_{alim}} \cdot v_{ref} \quad V_{cm} = \frac{(2^N - 1)}{v_{alim}} \cdot v_{cm} \quad V_{bg} = \frac{(2^N - 1)}{v_{alim}} \cdot v_{bg} \quad (5.13)$$

Les tensions de saturation de l'amplificateur sont considérées comme symétriques par rapport à sa tension d'alimentation v_{alim} , ainsi on peut écrire la relation :

$$\frac{v_{sat}^+ + v_{sat}^-}{2} = \frac{v_{alim}}{2} \quad (5.14)$$

En choisissant pour C_{AMP} un simple correcteur proportionnel, l'équation d'évolution de la valeur numérique de commande V_{ref} est calculée à partir des équations 5.14, 5.8, 5.13 :

$$V_{ref}(n+1) = V_{ref}(n) - C_{AMP} \left(a - \frac{2^N - 1}{2} \right) \quad (5.15)$$

avec

$$a = \frac{2^N - 1}{2} \left(\frac{v_{max}(n) + v_{min}(n)}{v_{alim}} \right)$$

Dans le cas du convertisseur numérique, en choisissant pour C_{ADC} un correcteur proportionnel et suivant 5.11, 5.12 et 5.13, les équations de commande numériques de V_{cm} et V_{bg} sont :

$$V_{cm}(n+1) = V_{cm}(n) - C_{ADC} \left(V_{cm}(n) - \left(\gamma \cdot C_{AMP} \left(a - \frac{2^N - 1}{2} \right) + a \right) \right) \quad (5.16)$$

$$V_{bg}(n+1) = V_{bg}(n) - C_{ADC}(n) \cdot (V_{bg}(n) - (c+b)) \quad (5.17)$$

avec

$$b = \frac{5(2^N - 1)}{4} \frac{v_{max}(n) - v_{min}(n)}{v_{alim}} \quad c = \frac{5(2^N - 1)}{2} \frac{\Delta v}{v_{alim}}$$

A partir de 5.7, 5.4 et 5.13, les quantités a et b sont calculées en fonction de V_{max} et V_{min} .

$$a = \frac{V_{bg}(n)}{5} \left(\frac{V_{max}(n) + V_{min}(n)}{2^N - 1} - 1 \right) + \frac{V_{cm}(n)}{2}$$

$$b = \frac{4}{5} \frac{V_{max}(n) - V_{min}(n)}{2^N - 1} V_{bg}(n)$$

5.1.3 Correction du FPN

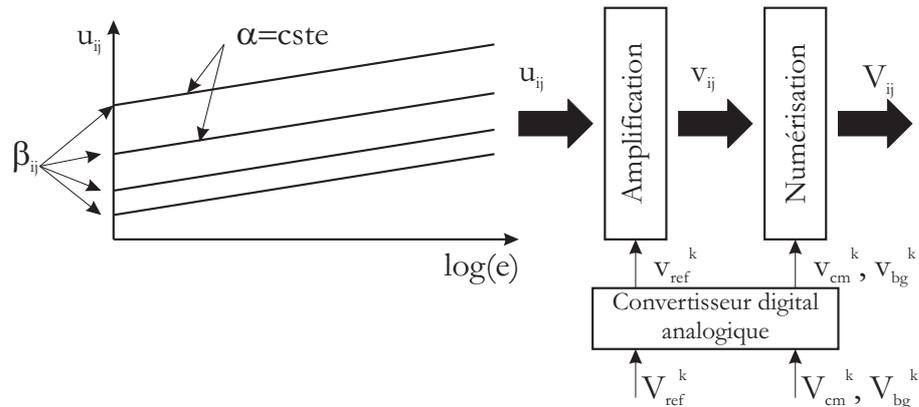
Pour intégrer le processus de correction l'image dans la boucle de rétroaction proposée. La mesure d'une image de référence est effectuée suivant des paramètres d'acquisition spécifiques. Ainsi, la modification dynamique de ces paramètres exigent une adaptation de l'image de référence. Le module de correction proposé dans le chapitre 4 est légèrement modifié pour réévaluer les différences d'"offset" de pixels.

La réévaluation est effectuée en supposant que le paramètre α de l'équation de réponse des pixels 5.1 est identique pour chacun d'entre eux. Cette hypothèse est dans l'absolu fautive, cependant la principale composante des disparités des réponses provient du paramètre β (fig. 5.6). Les différences d'"offset" évaluées sont des mesures différentielles et le gain γ de l'amplificateur est constant. En admettant que l'amplificateur et le convertisseur soient utilisés dans leurs plage de linéarité, les mesures d'"offset" dépendent uniquement de la largeur de la bande de conversion. Ainsi seul le paramètre V_{bg} influence la mesure de la disparité des pixels. L'image de référence utilisée est ainsi rendue indépendante de V_{bg} , et calculée selon l'équation :

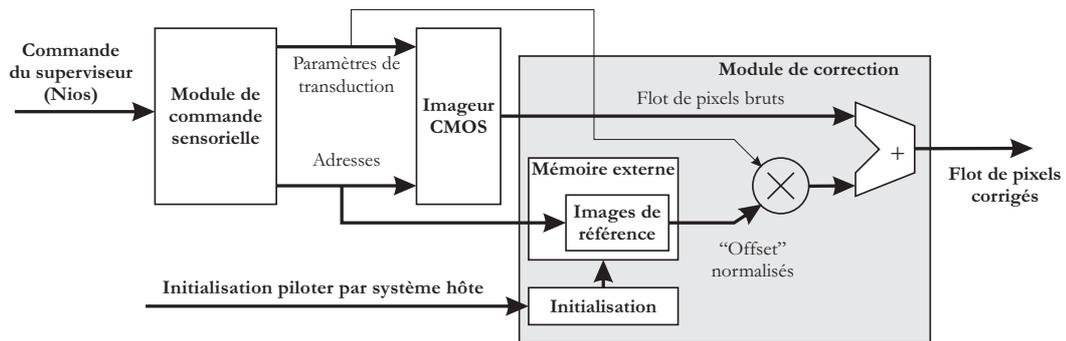
$$I_{ref}(k) = \frac{I(k) - \frac{1}{N} \sum_{i=0}^{N-1} I(k)}{V_{bg}^{init}}$$

Et durant l'acquisition, le flot de pixels bruts issu de l'imageur est corrigé à partir de l'équation :

$$V_{cor}(k) = V_{brut}(k) - I_{ref}(k) \cdot V_{bg}$$

FIG. 5.6 – *Modèle de la disparité des pixels de l'imageur CMOS.*

où $V_{cor}(k)$ et $V_{brut}(k)$ sont respectivement les valeurs numériques corrigée et brute de pixel k . La seule différence avec le module de correction présenté au chapitre 4 est la présence d'un multiplieur à la sortie de la mémoire (fig. 5.7).

FIG. 5.7 – *Module de correction du FPN suivant les variation de V_{bg} .*

5.1.4 Résultats

Dans ce paragraphe, différentes prises de vue sont proposées afin d'illustrer les résultats obtenus en contrôlant la chaîne de transduction de l'imageur. La première séquence d'images (fig. 5.8) montre une scène fortement éclairée. Une lampe a été placée face au livre observé. L'image de gauche est issue d'un réglage moyen des paramètres de la chaîne de transduction. La seconde image représente la même scène



FIG. 5.8 – *Scène soumise à un fort éclairage indirect*

mais en utilisant le procédé d'asservissement. On constate l'égalisation de l'histogramme de l'image et l'élimination des saturations observées précédemment. La dernière image montre le résultat obtenu sans correction du FPN de l'imageur. Malgré la simplification du modèle de pixel utilisé, on constate un résultat satisfaisant.

D'autres conditions expérimentales sont proposées figure 5.9. Cette fois-ci l'imageur est soumis à un éclairage direct. Ces conditions d'éclairage sature complètement l'imageur lors d'un réglage moyen des paramètres de transduction. Afin de visualiser la scène considérée un réglage manuel approximatif a été effectué pour obtenir l'image en haut à gauche. L'image à sa droite représente la même scène mais en utilisant le procédé d'asservissement. Enfin, la dernière image montre le résultat obtenu en approchant la lampe vers le capteur. Malgré l'éclairage très élevé, l'image n'est pas saturée et on visualise aisément le filament de la lampe utilisée.

Plusieurs perspectives sont envisagées pour améliorer cette méthode d'optimisation de contraste. Tout d'abord, il est envisagé de coupler les correcteurs de l'asservissement avec le nombre de pixels de la fenêtre d'intérêt considérée qui détermine le temps réponse du système. Le réglage des correcteurs dépend de ce temps ainsi il est envisagé d'intégrer ce paramètre dans l'asservissement afin d'optimiser sa réponse temporel. D'autre part, les réglages dynamiques considérés sont spécifiés pour un région d'intérêt sur l'imageur. Il est envisagé d'utiliser le même type de procédé mais cette fois-ci au niveau du pixel. Pour pouvoir modifier les paramètres de transduction de chaque pixels indépendamment, les convertisseurs analogiques/numériques qui commandent de l'imageur ont été conçus pour présenter des temps de réponse inférieurs à ceux de l'imageur. L'idée consiste à définir une carte de transduction qui

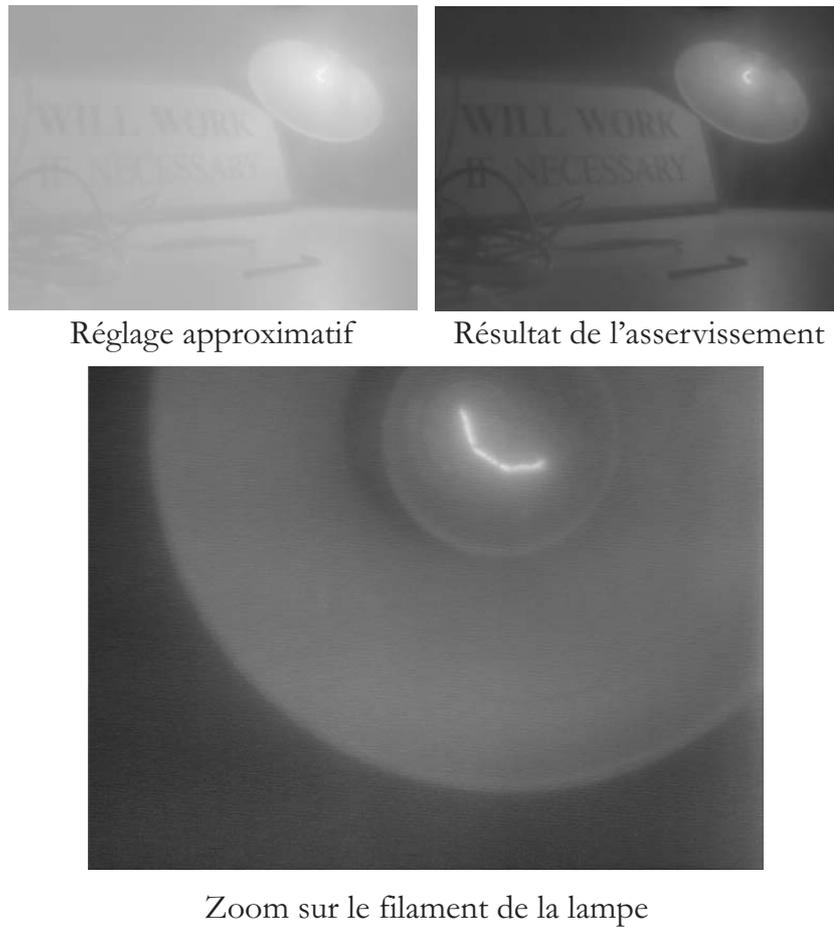


FIG. 5.9 – *Scène soumise à un fort éclairage direct*

évolue dynamiquement. Pour cela il faudra définir quelles caractéristiques locales de l'images vont spécifier la forme de cette carte de transduction.

5.2 Détecteur actif de focalisation

Suivant la méthodologie d'implantation proposée, la focalisation de l'attention du système représente une fonction temporellement critique qui va délimiter la réactivité globale du système de vision. Le détecteur actif proposé exploite une zone d'intérêt dynamique sur l'imageur. L'acquisition contrôlée de l'imageur CMOS est exploitée pour élaborer un dispositif de suivi rapide. La fonction détection réalisée

est une estimation globale du déplacement d'un motif de référence dans la fenêtre d'acquisition courante. Une rétroaction sensorielle est utilisée pour mettre à jour dynamiquement la position de la fenêtre d'acquisition suivant l'évolution de l'environnement.

5.2.1 Description fonctionnelle globale

Le schéma fonctionnel (fig. 5.10) représente le détecteur proposé au sein du squelette d'implantation du chapitre 3. Le détecteur est composé d'un module de mémo-

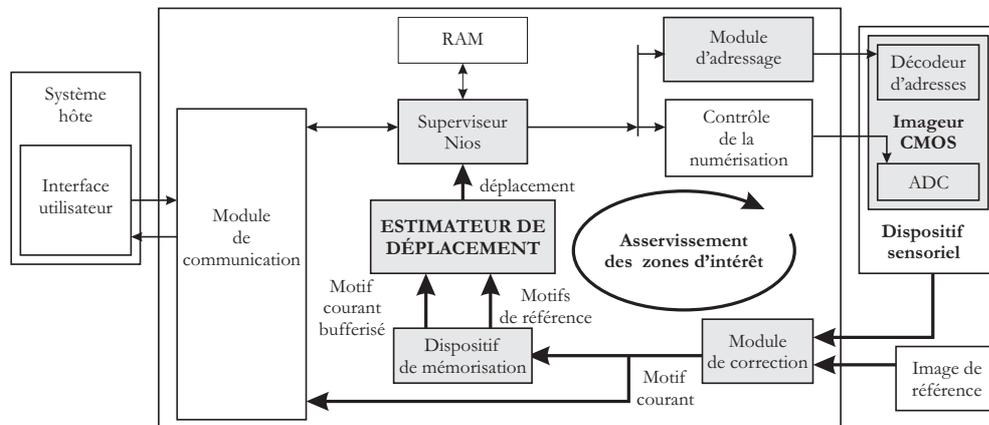


FIG. 5.10 – Architecture fonctionnelle du détecteur actif de focalisation.

risation et d'un module d'estimation de déplacement. Le dispositif de mémorisation assure deux fonctions. Tout d'abord, il permet de stocker plusieurs motifs de référence pour étendre le dispositif de suivi à plusieurs fenêtres d'intérêt. D'autre part, il bufferise la zone d'intérêt couramment acquise pour permettre au module d'estimation de déplacement de travailler indépendamment de la cadence d'acquisition de l'imageur. L'ensemble fonctionnelle constitue une boucle d'asservissement de la position de fenêtres d'intérêt qui fait appel aux modules génériques détaillés dans le chapitre 4.

Dans les sections suivantes, l'algorithme utilisé pour le module d'estimation de déplacement est présenté. Ensuite, les deux modules constituant le détecteur actif sont détaillés.

5.2.2 Algorithme d'estimation du déplacement

L'algorithme utilisé pour réaliser l'estimation du déplacement est celui Kanade Lucas Tomasi [89, 90] qui est largement utilisé par la communauté. Sans considérer les problèmes de déformations, d'occultation et d'illumination, cet algorithme modélise la dissimilitude entre deux images consécutives par une translation globale des pixels. La fonction de passage d'une image $I(t + \tau)$ à une image $I(t)$ est la suivante :

$$I(\mathbf{x} + d, t + \tau) = I(\mathbf{x}, t)$$

Afin d'estimer le vecteur de translation $d = (d_x, d_y)^t$, l'erreur ε est minimisée sur l'ensemble de l'espace image W :

$$\varepsilon = \sum^W (I(\mathbf{x} + d, t + \tau) - I(\mathbf{x}, t))^2$$

L'algorithme propose de linéariser $I(\mathbf{x} + d, t + \tau)$ en réalisant une approximation au premier ordre :

$$I(\mathbf{x} + d, t + \tau) = I(\mathbf{x}, t + \tau) + \left(\frac{\partial I(\mathbf{x}, t + \tau)}{\partial \mathbf{x}} \right)^t \cdot d$$

avec

$$\frac{\partial I(\mathbf{x}, t + \tau)}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial I(t+\tau)}{\partial x} \\ \frac{\partial I(t+\tau)}{\partial y} \end{pmatrix}$$

On obtient :

$$\varepsilon = \sum^W \left(I(\mathbf{x}, t + \tau) - I(\mathbf{x}, t) + \left(\frac{\partial I(\mathbf{x}, t + \tau)}{\partial \mathbf{x}} \right)^t \cdot d \right)^2$$

Afin de déterminer d , l'expression est différenciée pour évaluer le minimum de la fonction ε :

$$\begin{aligned} \frac{\partial \varepsilon}{\partial d} &= \frac{\partial}{\partial d} \sum^W \left(I(\mathbf{x}, t + \tau) - I(\mathbf{x}, t) + \left(\frac{\partial I(\mathbf{x}, t + \tau)}{\partial \mathbf{x}} \right)^t \cdot d \right)^2 = 0 \\ \sum^W \frac{\partial I(\mathbf{x}, t + \tau)}{\partial \mathbf{x}} \cdot \left(I(\mathbf{x}, t + \tau) - I(\mathbf{x}, t) + \left(\frac{\partial I(\mathbf{x}, t + \tau)}{\partial \mathbf{x}} \right)^t \cdot d \right) &= 0 \end{aligned}$$

On obtient finalement la relation :

$$\sum^W \left(\frac{\partial I(\mathbf{x}, t + \tau)}{\partial \mathbf{x}} \cdot \left(\frac{\partial}{\partial \mathbf{x}} I(\mathbf{x}, t + \tau) \right)^t \right) \cdot d = \sum^W \left(\frac{\partial I(\mathbf{x}, t + \tau)}{\partial \mathbf{x}} \cdot (I(\mathbf{x}, t + \tau) - I(\mathbf{x}, t)) \right)$$

En posant :

$$G = \sum^W \left(\frac{\partial I(\mathbf{x}, t + \tau)}{\partial \mathbf{x}} \cdot \left(\frac{\partial}{\partial \mathbf{x}} I(\mathbf{x}, t + \tau) \right)^t \right)$$

$$G = \begin{pmatrix} \sum^W \left(\frac{\partial I(t+\tau)}{\partial x} \right)^2 & \sum^W \left(\frac{\partial I(t+\tau)}{\partial x} \cdot \frac{\partial I(t+\tau)}{\partial y} \right) \\ \sum^W \left(\frac{\partial I(t+\tau)}{\partial x} \cdot \frac{\partial I(t+\tau)}{\partial y} \right) & \sum^W \left(\frac{\partial I(t+\tau)}{\partial y} \right)^2 \end{pmatrix}$$

$$e = \begin{pmatrix} \sum^W \frac{\partial I(t+\tau)}{\partial x} \cdot (I(\mathbf{x}, t + \tau) - I(\mathbf{x}, t)) \\ \sum^W \frac{\partial I(t+\tau)}{\partial y} \cdot (I(\mathbf{x}, t + \tau) - I(\mathbf{x}, t)) \end{pmatrix}$$

On obtient l'équation matricielle :

$$G \cdot d = e \quad (5.18)$$

Cette équation est résolue suivant une procédure itérative (Newton Raphson) à partir de la fonction $\hat{d} = G^{-1} \cdot e$. Ce raisonnement effectué dans le cas d'un modèle de déformation simple peut être réalisé pour des modèles plus élaborés. Par exemple, l'algorithme de suivi de Shi Tomasi [91] utilise un modèle composé d'une transformée affine ou encore les travaux de G. D. Hager et P. N. Belhumeur [92] qui généralise l'approche en intégrant l'illumination.

5.2.3 Module de mémorisation

La première fonction de ce module consiste à stocker un ensemble de motifs de référence suivant la commande du superviseur. Ces motifs représentent les zones locales d'intérêt du système et sont stockés dans une SRAM externe (fig. 5.11). Le second rôle de ce dispositif est de permettre à l'estimateur de disposer simultanément de la zone d'acquisition courante et du motif de référence qui lui correspond. À partir d'un mécanisme d'échange basé sur deux mémoires internes (double port), le dispositif réalise deux opérations simultanées suivant des cadences de fonctionnement différentes. Le futur motif courant est stocké suivant la cadence d'acquisition du capteur. Et simultanément, le motif courant et son motif de référence sont utilisés par le calculateur de déplacement suivant une cadence de travail plus élevée.

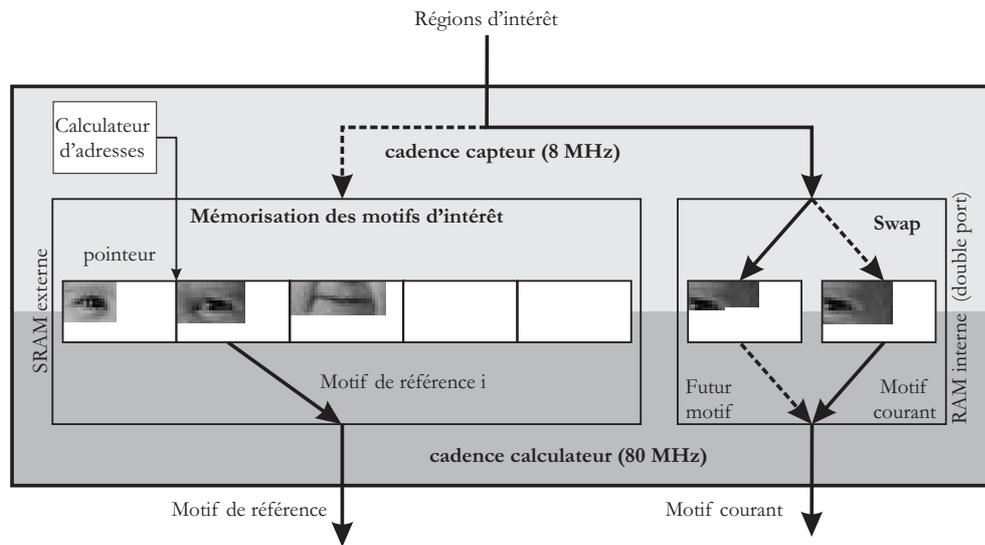


FIG. 5.11 – Mécanisme d'échange permettant au calculateur de déplacement de travailler à sa propre fréquence de traitement.

5.2.4 Module d'estimation du déplacement

Le processus de suivi basé sur le module d'estimation du déplacement est constitué de deux niveaux fonctionnels itératifs imbriqués. Le premier provient de l'acquisition continue de la fenêtre d'intérêt sur l'imageur et le second du procédé d'estimation de la translation liant le motif courant et un motif de référence. L'architecture proposée pour implémenter le dispositif de suivi est basé sur l'algorithme de Lucas Kanade Tomasi est proposé sur la figure 5.12.

Le module de mémorisation présenté précédemment (fig. 5.11) permet de traiter le motif courant et de stocker le prochain motif simultanément. Il permet d'effectuer ces deux opérations suivant deux cadences de travail différentes. Ainsi, le traitement du motif courant est effectué à une fréquence 10 fois plus élevée que la cadence d'acquisition (8 MHz). Selon ce choix, le dispositif dispose de 10 cycles d'horloge par pixels pour effectuer l'estimation de la translation. L'architecture fonctionnelle proposée pour implanter l'algorithme est composée d'une fonction pipelinée qui calcule les gradients du motif courant I et la différence entre I et le motif de référence I_{ref} . Le motif de référence est translaté à chaque itération selon le vecteur de déplacement estimé $d = (dx, dy)^t$. Lors de la première itération, les valeurs accumulées g_{xx}, g_{xy}

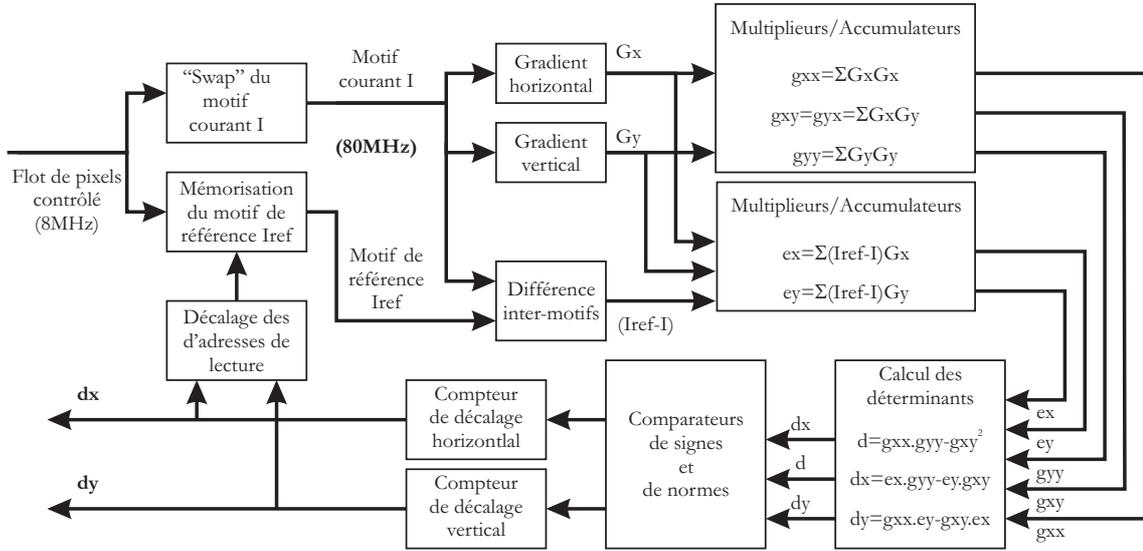


FIG. 5.12 – Architecture fonctionnelle de l'estimateur de translation.

et g_{yy} qui compose la matrice G sont stockées jusqu'au traitement du prochain motif courant I . A chaque itération, les valeurs accumulées e_x et e_y sont recalculées en fonction du motif de référence I_{ref} traduit suivant la valeur précédente de d . Ainsi le motif de référence traduit converge vers le motif courant et le vecteur de translation est accumulé dans d .

5.2.4.0.1 Accumulation des éléments de G et e Le dispositif accumulateur calcule les paramètres de la matrice G et du vecteur e suivant le motif courant I et le motif de référence I_{ref} . Les gradients de I sont calculés en appliquant sur l'image un masque de convolution $m(i,j)$ de dimensions $M \times M$. L'architecture fonctionnelle développée (fig. 5.13) réalise un filtre à réponse impulsionnelle finie qui permet d'appliquer n'importe quel masque de convolution m_{ij} sur une image paramétrée par sa largeur L . Ce filtre est défini par l'équation suivante :

$$f(k) = \sum_{i=0}^M \sum_{j=0}^M m(M-i, M-j) \cdot I(k - (i \cdot L) + j)$$

ou bien

$$F(z) = \sum_{i=0}^M \left(z^{-i \cdot L} \left(\sum_{j=0}^M m(M-i, M-j) z^{-j} \right) \right)$$

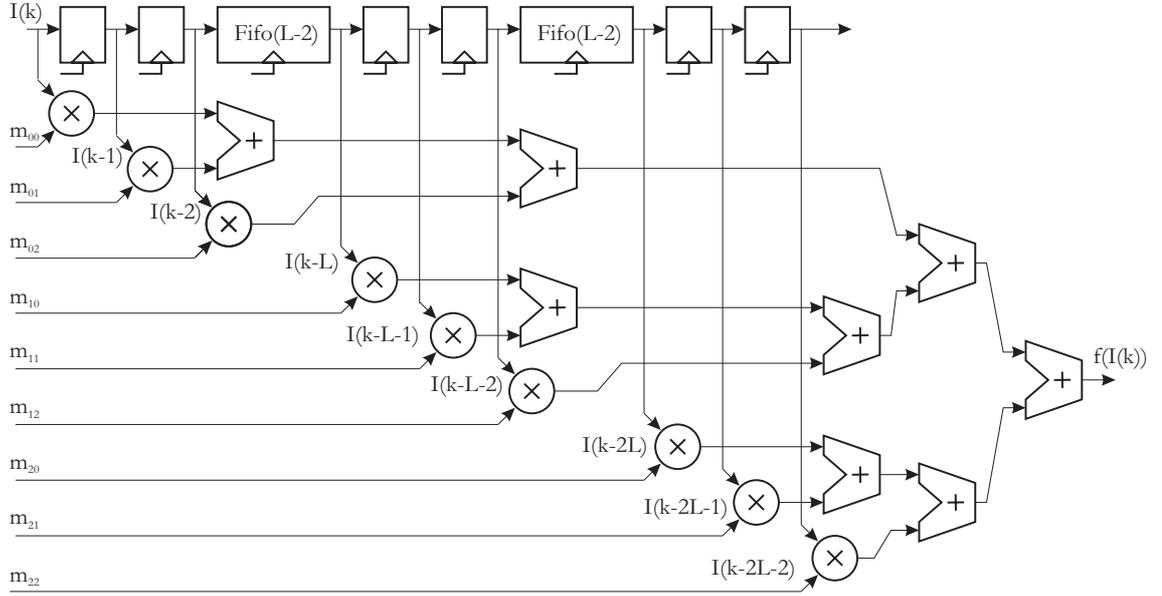


FIG. 5.13 – Architecture fonctionnelle de l'estimateur de translation.

Cependant, les contraintes d'implantation limite les dimensions du masque de convolution à 3, 5 ou 7 car le nombre de multiplieurs et d'additionneurs croît rapidement. Cette architecture n'a pas été optimisée pour l'application d'un masque spécifique, car elle permet d'effectuer un changement dynamique des paramètres du masques. Cette possibilité pourra être utilisée lors de futures expérimentations pour intégrer une rétroaction visuelle sur ces paramètres algorithmiques. Cette architecture est utilisée pour calculer les gradients horizontal et vertical de l'image courante (g_{xx} , g_{xy} , g_{yy} , e_x et e_y) ce nécessite une batterie de multiplieurs accumulateurs. Ces éléments constituent la matrices G et le vecteur e de la section précédente. La figure 5.14 montre les résultats de simulation de l'architecture pour différents types de masques de convolution. L'image en haut à droite est l'image initiale moyennée suivant le masque de convolution suivant :

$$M_{moy} = \frac{1}{9} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad M_{g_x} = \frac{1}{3} \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \quad M_{g_y} = \frac{1}{3} \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}$$

5.2.4.0.2 Accumulation du vecteur de translation Suivant les valeurs de G et de e , il faut calculer le vecteur de translation d établi par l'équation matricielle

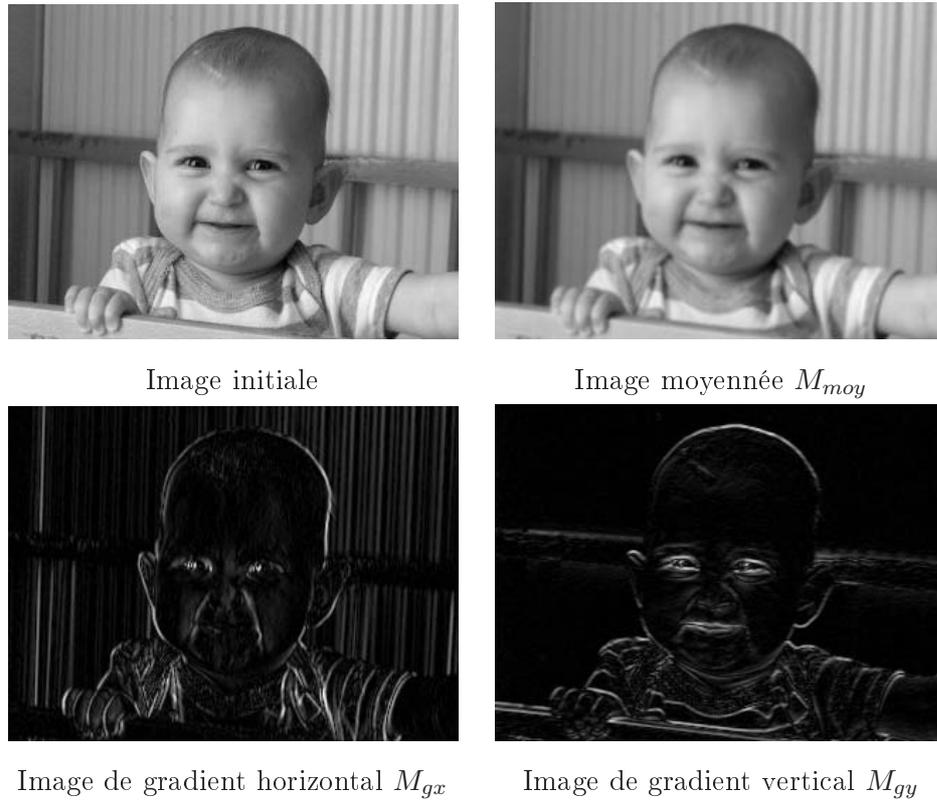


FIG. 5.14 – Résultats de simulation de l'architecture de convolution.

$\hat{d} = G^{-1}.e$. Cependant, la zone d'acquisition et le motif de référence sont tous deux des signaux discrets. Ainsi, le calcul d'une valeur décimale de d qui implique l'inversion de la matrice G n'est pas nécessaire. Afin, d'établir à chaque itération, une valeur discrète de d , les déterminants (fig. 5.12) suivant sont calculés :

$$\Delta = gxx.gyy - gxy^2 \quad \Delta x = ex.gyy - ey.gxy \quad \Delta y = gxx.ey - gxy.ey$$

Les valeurs décimales des déplacements horizontal et vertical sont :

$$dx = \frac{\Delta x}{\Delta} \quad dy = \frac{\Delta y}{\Delta}$$

Du fait des caractéristique du processus d'estimation itératif, ces valeurs restent proches de un. Au lieu d'effectuer la division, les normes et les signes de ces déterminants sont comparés pour mettre à jour des compteurs qui accumulent dx la translation horizontale et dy la translation horizontale. Afin, d'évaluer la nécessité

d'une mise à jour des compteurs, les normes des déterminants sont comparées selon la validité de ces équations, et sont alors incrémentés ou décréments :

$$|\Delta|/2 > \Delta x \quad |\Delta|/2 > \Delta y$$

Ainsi, lorsque dx (ou dy) est supérieur à 0.5, le compteur est incrémenté dans le cas où $\Delta \cdot \Delta x > 0$ (ou $|\Delta|/2 > \Delta y$) sinon il est décréments. De cette manière, les dx et dy sont accumulées par les compteurs de déplacements.

Suivant ces deux dispositifs d'accumulation, l'estimation du vecteur de la translation liant deux motifs est réalisée de manière itérative. L'acuité de ce procédé dépend du nombre d'itération réalisée et de la valeur du déplacement recherché. Pour garantir un fonctionnement temps réel de ce dispositif de suivi, il faut limiter le nombre d'itérations. Sachant que la cadence de traitement est dix fois que la cadence d'acquisition ont dispose de 10 cycles par pixels pour effectuer la mesure. Du fait du retard introduit par les divers modules de traitements, le déplacement maximal mesurable en temps réel est de 9 pixels. En considérant, des zones d'intérêt locales, la translation observée entre deux motifs reste minime du fait d'un cadence de mise à jour du motif élevée. La cadence de traitement choisie paraît satisfaisante pour assurer à la fois un fonctionnement temps réel et une estimation correcte du vecteur de translation.

5.2.5 Conclusion

Une étude en simulation a été réalisée à partir du motif de référence figure 5.15. Ce motif a été artificiellement translaté pour simuler un déplacement dans la zone d'intérêt et injecter dans la description VHDL de l'architecture. On constate que pour toutes les valeurs de déplacement testées, le motif de référence est recalé parfaitement. Ces essais étant concluant, la prochaine étape sera l'implantation sur la plateforme de ce détecteur.

Cette architecture instaure un asservissement sur les caractéristiques géométriques de la fenêtres d'intérêt de l'imageur. L'objectif est de considérer des zones d'intérêt locales ou sous-échantillonnées pour accroître l'échantillonnage temporel du signal. En minimisant, les variations entre deux acquisitions consécutives, l'im-

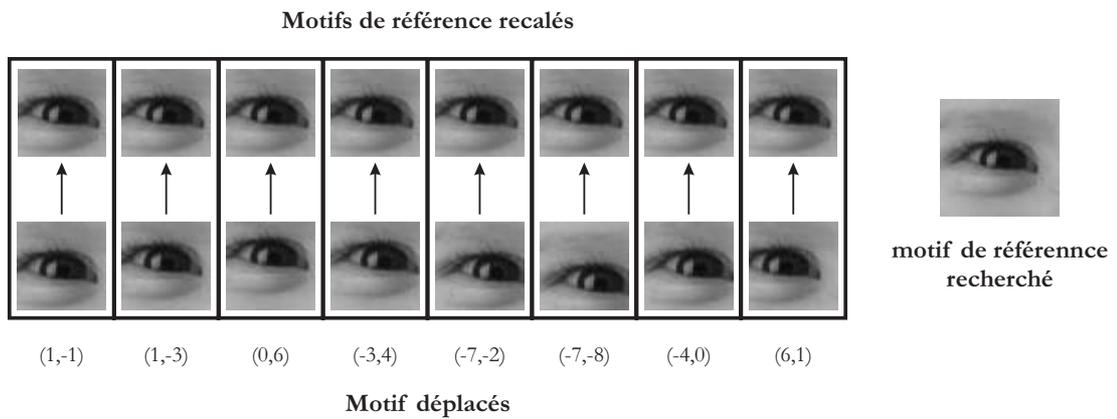


FIG. 5.15 – Résultats de simulation de l'architecture du détecteur actif de focalisation.

plémentation d'un processus de suivi est facilitée et sa robustesse est accrue. L'architecture d'implantation du calculateur de déplacement doit permettre d'exploiter pleinement la cadence d'acquisition de l'imageur.

Conclusion

5.3 Conclusion

La vision par ordinateur est un domaine qui nécessite la conception de machines dédiées à la fois très puissantes et très coûteuse. En s'appuyant sur des technologies d'imagerie et d'intégration récentes, une manière originale d'aborder ce problème a été proposée. La contribution principale de cette thèse est la conception d'une plateforme de recherche destinée à l'implantation de systèmes de vision active embarqués. Ce dispositif a été développé pour satisfaire les contraintes d'une méthodologie d'implantation dans un soucis de flexibilité optimum. Ce dispositif intègre un imageur CMOS pour exploiter l'échantillonnage spatio-temporel du signal image dans une stratégie de perception bas niveau.

Une méthodologie d'implantation est proposée pour définir une adéquation stratégique, algorithmique et architecturale. La notion de détecteur actif a été proposée pour unifier ses différents aspects au sein d'une même entité conceptuelle. Selon une approche locale, on recherche à décorréler les tâches élémentaire de détection pour déterminer une solution dédiée réutilisable. Le but est d'être capable de construire une chaîne de traitement réactive supervisée pour définir un système selon ses objectifs visuels et capable de s'adapter à son environnement.

Les travaux qui seront réalisés dans la continuité de cette thèse devront développer une bibliothèque de détecteurs actifs afin de confronter la validité de notre approche et de la solution matérielle proposées à différentes tâches de perception relatives à la vision active devront être abordées. Un ensemble de détecteurs actifs doit être mis au point pour être capable de mettre en place des procédés d'attention, de focalisation et d'identification variés. Malgré des objectifs visuels diversifiés, un

ensemble de fonctions élémentaires de détection communes doivent être identifiées pour concevoir une bibliothèque de détecteurs actifs qui permet de constituer une chaîne de traitement réactives adaptées à toutes sortes de situation visuelle.

Dans un second temps, il est prévu d'aborder le problème de prototypage rapide sur la plate-forme hétérogène proposée. De nombreux travaux de recherche [93, 94, 95] proposent des méthodes pour implémenter efficacement et rapidement des dispositifs de traitement d'image. Il est envisagé de développer ce type de méthodes dans le cas spécifique de la vision active. L'objectif est d'identifier des squelettes d'implantation et d'interactions récurrents pour réaliser un dispositif de développement automatisé.

Enfin, un dernier axe de recherche est issu de l'adéquation entre les composants reprogrammables dynamiquement et le caractère adaptatif des systèmes de vision active. Actuellement, les membres du projet ARDOISE [96] (Architecture Reconfigurable Dynamiquement Orientée Image et Signal Embarquable) travaillent sur la reconfiguration dynamique des FPGAs. L'objectif est de reconfigurer l'architecture matérielle du système au cours du traitement pour exécuter successivement différents algorithmes de traitement d'images. Cette possibilité est extrêmement intéressante pour le développement de systèmes de vision active. Au lieu d'adapter la manière d'exploiter une chaîne de perception donnée suivant une stratégie d'observation, la capacité de reprogrammation dynamique permet de faire évoluer matériellement la chaîne de perception. Suivant les objectifs momentanés du système et l'environnement observé, l'architecture de traitement peut être complètement redéfinie afin d'assurer continuellement adéquation stratégique, algorithmique et matérielle.

Bibliographie

- [1] Richard M. Wilkie and John P. Wann. Eye-movements aid the control of locomotion. *Journal of Vision*, 3:677–684, November 2003.
- [2] Jochen Triesch, Dana H. Ballard, Mary M. Hayboe, and Brian T. Sullivan. What you see is what you need. *Journal of Vision*, 3:86–94, November 2003.
- [3] W. H. Warren and D. J. Hannon. Eye movements and optical flow. *Journal of the optical society of america A*, 7(1):160–169, 1990.
- [4] S. Grossberg and E. Mingolla. "neural dynamics of surface perception: boundary webbs, illuminants and shape-from-shading". *Computer Vision, Graphics and Image Processing*, pages 116–165, 1987.
- [5] P. Gaussier et J.P. Cocquerez. "utilisation des réseaux de neurones pour la reconnaissance de scènes complexes: simulation d'un système visuel comprenant plusieurs aires corticales". *Traitement du Signal*, 8(6):441–466, 1991.
- [6] D. Marr. *Vision*. Freeman editors, San Francisco, 1982.
- [7] Y. Demazeau. A distributed artificial intelligence approach to integration of visual modules. In *European working week on vision, Workshop on integration of visual modules*, Heraklion, Greece, September 1990.
- [8] Y. Demazeau. *Niveaux de représentation pour la vision par ordinateur. Indices d'image et indices de scène*. PhD thesis, Thèse de doctorat, INP Grenoble, 1996.
- [9] J.L. Crowley. Knowledge, symbolic reasoning and perception. In *ntelligent Autonomous Systems*, Amsterdam, Nederland, 1990.
- [10] B. Neumann. Niveaux de représentation en vision par ordinateur. communication personnelle, 1990.

- [11] A.R. Hanson and E.M. Riseman. *VISIONS: a computer system for interpreting scenes*. Computer Vision Systems - Academic Press, a.r. hanson et e.m. riseman, editeurs edition, 1978.
- [12] V. Clement. *Raisonnements cognitifs appliqués au pilotage d'algorithmes de traitement d'images*. PhD thesis, Thèse INRIA, Nice-Sophia-Antipolis, 1990.
- [13] O. Baujard and C. Garbay. "kiss: A multi-agent segmentation system". *SPIE Optical Engineering*, 32(6):1235–1249, 1993.
- [14] Aloimonos J.Y., Weiss I., and Bandyopadhyay A. Active vision. In *Proceedings of the 1st International Conference on Computer Vision*, pages 35–54, London, England, June 1987.
- [15] Dana H. Ballard. Animate vision. *Artif. Intell.*, 48(1):57–86, 1991.
- [16] D. H. Ballard and M Brown. *Active Perception*, volume 1, chapter Principles of Animate Vision, pages 245–282. Y. Aloimonos, 1993.
- [17] Bajcsy R. Active perception. *IEEE Proceedings*, 76(8):996–1006, August 1988.
- [18] D.A. Rosenthal and R. Bajcsy. Visual And Conceptual Hierarchy: A Paradigm For Studies Of Automated Generation Of Recognition Strategies. *T-PAMI*, 6:319–325, 1984.
- [19] J. Y. Aloimonos. Purposive and qualitative vision. In *Proc. AAAI-90 Workshop on Qualitative Vision*, pages 1–5, San Mateo, CA, 1990. Morgan Kaufmann.
- [20] Y. Aloimonos. Special issue on purposive and quantitative active vision. *CVGIP B: Image Understanding*, 56, 1992.
- [21] Aloimonos Y. *Active perception*. Erlbaum, Hillsdale, NJ, 1993.
- [22] D.H. Ballard, M.M. Hayhoe, P.K. Pook, and R.P.N. Rao. Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20, 1997.
- [23] J.K. Tsotsos. Active vs. passive visual search: which is more efficient. *Rapport, University of Toronto*,, 1990.
- [24] Pahlavan K. and Uhlin T. and Eklund J.O. *Active vision as a methodology*, pages 19–46. Robotics and Automated Systems, ISBN 981-02-1364-6. Lawrence Erlbaum Associates Publishers, 1993.
- [25] Vieville T. A few steps towards 3d active vision. *Springer Series in Information Sciences*, 33, 1997.

-
- [26] B. Madden and U. von Seelen. Penneyes: A binocular active vision system. Technical report, GRASP Laboratory, Department of Computer and Information Science, University of Pennsylvania, Philadelphia, 1995.
- [27] H. I. Christensen. *Active Robot Vision: Camera Heads, Model Based Navigation and Reactive Control*. K W Bowyer, H. I. Christensen, Kevin Bowyer, 1993.
- [28] Andersen C.S. *A Framework for control of a camera head*. PhD thesis, Laboratory of image analysis, Aalborg Universtity, Denmark, 1996.
- [29] W.W. Mayol, B.J. Tordoff, and D.W. Murray. Wearable visual robots. In *Proc Int Symposium on Wearable Computing*, 2000.
- [30] Vladimir Brajovic and Takeo Kanade. A vlsi sorting image sensor: Global massively parallel intensity-to-time processing for low-latency, adaptive vision. *IEEE Transactions on Robotics and Automation*, 15(1):67–75, February 1999.
- [31] C. Koch and S. Ullman. Shifts in selective visual-attention toward the underlying neural circuitry. *Human Neurobiology*, 4(4):219–227, 1985.
- [32] G. Indiveri. Neuromorphic selective attention systems. In *IEEE International Symposium on Circuits and Systems*, 2003.
- [33] Indiveri G. and Mürer R. and Kramer J. Active vision using an analog vlsi model of selective attention. *IEEE Transactions on Circuits and Systems-II: Analog and digital signal processing*, 48:492–500, May 2001.
- [34] Bederson, B. B. *A Miniature Space-Variant Active Vision System: Cortex-I*. PhD thesis, New York University, New York, NY, 1992.
- [35] Park D.-S., Kim J.-H., Kim H.-S., Park J.-H., Shin J.-K., and Lee M. A foveated-structure cmos retina chip for edge detection with local light adaptation. *Sensors and Actuators A: Physical*, 108(1):75–80, 2003.
- [36] M. Tistarelli and G. Sandini. On the advantage of polar and log-polar mapping for direct estimation of time-to-impact from optical flow. *PAMI*, 15(4):401–410, April 1993.
- [37] Rojer A. and Schwartz E. L. Design considerations for a space-variant visual sensor with complex-logarithmic geometry. In *10th International Conference on Pattern Recognition*, volume 2, Nashville, TN, 1990.
- [38] J. L. Crowley. Integration and control of reactive visual processes. *Robotics and Autonomous Systems*, 15(1), 1995.

- [39] J. L. Crowley and H.I. Christensen. Integration of visual processes. *Pattern Recognition and Artificial Intelligence*, 1993.
- [40] D.H. Ballard. *An approach to knowledge-directed image analysis*, volume Computer Vision Systems, pages 271–281. A.R.Hanson et E.M. Riseman, éditeurs, 1978.
- [41] R.D. Rimey and C.M. Brown. Task-Oriented Vision with Multiple Bayes Nets in Active Vision. *chapitre 13, MIT press*, 1992.
- [42] R.D. Rimey et C.M. Brown. Where to look next using a bayes net : Incorporating geometric relations. *G. Sandini, editeur, European Conference on Computer Vision, SPIE Optical Engineering*, pages 542–550, 1992.
- [43] D. Litwiller. CCD vs CMOS: Facts and Fiction. *Photonics Spectra* © Laurin Publishing Co. Inc., January 2001.
- [44] N. Blanc. CCD versus CMOS - has CCD imaging come to end? *Photogrammetric Week*, 2001.
- [45] E. R. Fossum. CMOS Image Sensors: Electronic Camera-On-A-Chip. *IEEE Transactions on Electron Devices*, 44(10):1689–98, october 1997.
- [46] C. S. Hong. *On Chip Spatial Image Processing with CMOS Active Pixel Sensors*. PhD thesis, University of Waterloo, Electrical and Computer Engineering, 2001.
- [47] W.S. Boyle and G.E. Smith. Charge-coupled semiconductor devices. *Bell Syst. Tech. J.*, 49:587–93, 1970.
- [48] B. Cassadei. *Conception et réalisation d'un capteur en technologie CMOS pour l'observation d'impulsions lumineuses brèves voisines de 1ns*. PhD thesis, Université Louis Pasteur ũ Strasbourg, Sciences pour l'Ingénieur, 2003.
- [49] Xinqiao Liu SukHwan Lim Abbas El Gamal Ali Ozer Ercan, Feng Xiao and Brian Wandell. Experimental High Speed CMOS Image Sensor System and Applications. *Proceedings of IEEE Sensors*, 2002.
- [50] X. Liu S. Kleinfelder, S. Lim and A. El Gamal. A 10 000 Frames/s CMOS Digital Pixel Sensor. *IEEE journal of solid-state circuits*, (12), december 2001.
- [51] Y. Huang. *Current-Mode CMOS Image Sensor*. PhD thesis, University of Waterloo, Electrical and Computer Engineering, 2002.
- [52] D. Joseph. *Modelling and calibration of logarithmic CMOS image sensors*. PhD thesis, University of Oxford, Engineering Science, 2002.

-
- [53] C. Cavadore. *Conception et caractérisation de capteurs d'images à pixels actifs CMOS-APS*. PhD thesis, Ecole Nationale Supérieure de l'Aéronautique et de l'Espace (SUPAERO) and Thomson TSC CCD, Toulouse, 1998.
- [54] Y. Degerli. *Etude, modélisation des bruits et conception des circuits de lecture dans les capteurs d'images à pixels actifs*. PhD thesis, Ecole Nationale Supérieure de l'Aéronautique et de l'Espace (SUPAERO), Toulouse, Octobre 2000.
- [55] M. Tabet. *Double Sampling Techniques for CMOS Image Sensors*. PhD thesis, University of Waterloo, Electrical and Computer Engineering, 2002.
- [56] J. S. Lee. *Photoresponse of CMOS Image Sensors*. PhD thesis, University of Waterloo, Electrical and Computer Engineering, 2003.
- [57] Y. Ni K. Matou. A Precise FPN Compensation Circuit for CMOS APS. *Electronics Letters, Vol.38 Issue.1*, pages 1078–1079, 2002.
- [58] W. Ogiers. Survey of CMOS Imagers. Technical report, IMEC report P60280-MS-RP-002, September 1997.
- [59] E. R. Fossum. Ultra-low-power imaging systems using CMOS image sensor technology. *Proc. SPIE Vol. 2267, Advanced Microdevices and Space Science Sensors*, pages 107–111, 1994.
- [60] B. Dierickx G. Meynants and D. Scheffer. CMOS active pixel image sensor with CCD performance. In *AFPAEC Europto/SPIE, proc. SPIE*, volume 3410, pages 68–76, Zurich, may 1998.
- [61] Y. Tanabe T. Adachi S. Shigematsu, H. Morimura and K. Machida. A Single-Chip Fingerprint Sensor and Identifier. *IEEE journal of Solid-State Circuits*, 34(12):2104–2111, december 1999.
- [62] Alireza Moini. *Vision chips*. International Series In Engineering and Computer Science. Kluwer Academic Publisers, 2000.
- [63] S. Vinayagamorthy. *An Implementation of Gaussian Image Processing for a CMOS Image Sensor*. PhD thesis, University of Waterloo, Electrical and Computer Engineering, 2002.
- [64] P. Dudek and P.J. Hicks. A General-Purpose Vision Chip with a Processor-Per-Pixel SIMD Array. pages 228–231, september 2001.
- [65] C. M. Higgins and C. Koch. An integrated vision sensor for the computation of optical flow singular points. *MIT Press, Advances in Neural Information*

- Processing Systems*, 11, 1998.
- [66] Shyh Yih Ma and Liang-Gee Chen. A single-chip CMOS APS camera with direct frame difference output. *IEEE Journal of Solid-State Circuits*, 34(10):1415–1418, october 1999.
- [67] D. Navarro. *Architecture et Conception de Rétines Silicium CMOS: Application à la mesure de flot optique*. PhD thesis, Université de Montpellier II, Electronique, Optronique et Systèmes, 2003.
- [68] R. D. Burns. *Improved Techniques for Object Location with CMOS Image Sensors*. PhD thesis, University of Waterloo, Electrical and Computer Engineering, 2003.
- [69] P. Gorria G. Cathebras B. Bellach L.F.C. Lew Yan Voon, B. Lamalle and D. Navarro. Real time pattern recognition retina in CMOS technology. *International Conference on Quality Control by Artificial Vision*, 1:238–242, february 2001.
- [70] L.F.C. Lew Yan Voon P. Gorria B. Bellach B. Lamalle, G. Cathebras and O. Aubreton. A Correlation Retina for Real-Time Pattern Recognition. In *proceedings of the 2001 8th IEEE International Conference on Emerging Technologies and Factory Automation*, pages 367–372, Antibes ũ Juan les Pins, FRANCE,, October 2001.
- [71] S. Matsunaga. A Single-Chip 10,000 frame/s CMOS Tracking Image Sensor for complex targets. *ESSCIRC*, 2002.
- [72] Vladimir Brajovic. An Object Tracking Computational Sensor. Technical Report CMU-RI-TR-01-40, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, december 2001.
- [73] Haruyoshi Toyoda Yoshihiro Nakabo, Masatoshi Ishikawa and Seiichiro Mizuno. 1ms Column Parallel Vision System and Its Application of High Speed Target Tracking. *IEEE Int. Conf. Robotics and Automation (San Francisco)*, pages 650–655, 2000.
- [74] Ziyi Lu and Bertram E. Shi. Visual Tracking with Subpixel Resolution using an Analog VLSI Computational Sensor. *IEEE Intl. Conf. on Robotics and Automation*, 2(8):1676–1681, april 2000.
- [75] Vladimir Brajovic and Takeo Kanade. Computational sensor for visual tracking with attention. *IEEE Journal of Solid State Circuits*, 33(8):1199 – 1207, August 1998.

- [76] N. Schibli S. Fischer and F. Moscheni. Design and development of the smart machine vision sensor (SMVS). *Proc. SPIE Vol. 3410, Advanced Focal Plane Arrays and Electronic Cameras II*, Thierry M. Bernard; Ed., pages 186–192, 1998.
- [77] C. Lavarenne T. Grandpierre and Y. Sorel. Optimized rapid prototyping for real-time embedded heterogeneous multiprocessors. In *7th International Workshop on Hardware/Software Co-Design*, pages 5195–5200, Rome, May 1999.
- [78] D. Ginhac J. Serot and J.P. Derutin. Skipper: A skeleton-based parallel programming environment for real-time image processing applications. In *Parallel Computing Technologies*, pages 296–305, 1999.
- [79] D. Muir and J. Sitte. Seeing cheaply: flexible vision for small devices. *Proceedings of the 2nd International Symposium on Autonomous Minirobots for Research and Edutainment*, February 2003.
- [80] Peter Lang Ulrich Muehlmann, Miguel Ribo and Axel Pinz. A New High Speed CMOS Camera for Real-Time Tracking Applications. In *International Conference on Robotics and Automation*, pages 5195–5200, New Orleans, LA, April 2004.
- [81] Virginie Fresse and Olivier Deforges. Arial: rapid prototyping for mixed and parallel platforms. *Parallel Comput.*, 28(7-8):1179–1202, 2002.
- [82] O. Déforges V. Fresse and J. F. Nezan. Avsyndex: A rapid prototyping process dedicated to the implementation of digital image processing applications on multi-dsps and fpga architectures. *EURASIP journal on Applied Signal Processing, special issue on Implementation of DSP and Communication Systems*.
- [83] A. Huster. *Relative Position Sensing by Fusing Monocular Vision and Inertial Rate Sensors*. PhD thesis, University of Standford, july 2004.
- [84] Jorge Lobo. *Inertial sensor data integration in computer vision system*. PhD thesis, University of Coimbra, Faculty of Science and Tecnology, Coimbra, february 2002.
- [85] C. Le Gal A. Guarda and A. Lux. Evolving visual features and detectors. 1998.
- [86] O. Boissier. *Problème du contrôle dans un système intégré de vision utilisation d'un système multi agents*. PhD thesis.
- [87] J. L. Crowley. Context Driven Observation of Human Activity. *European Symposium on Ambient Intelligence*, November 2003.

- [88] J. E. Bresenham. Algorithm for computer control of a digital plotter. Number 1, pages 25–30, 1965.
- [89] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.
- [90] Carlo Tomasi and Takeo Kanade. Detection and tracking of point features. *Carnegie Mellon University Technical Report CMU-CS-91-132*, April 1991.
- [91] J. Shi and C. Tomasi. Good features to track. *IEEE Conference on Computer Vision and Pattern Recognition*, June 1994.
- [92] G. D. Hager and P. N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE Transactions on pattern analysis and machine intelligence*, (10), October 1998.
- [93] D. Crookes K. Benkrid and A. Benkrid. Towards a general framework for fpga based image processing using hardware skeletons. *Parallel Computing*, 28:1141–1154, August 2002.
- [94] W. Böhm J. Hammes R. Rinker C. Ross M. Chawathe B. Draper, W. Najjar and J. Bins. Compiling and optimizing image processing algorithms for fpga's. *Int. Workshop on Computer Architecture for Machine Perception.*, pages 1141–1154, September 2000.
- [95] K. F. D. Alotaibi. *A high level hardware description environment for FPGA-based image processing applications*. PhD thesis, Faculty of science of the Queen's university of Belfast, 1999.
- [96] Riad Bourguiba Didier Demigny, Nassima Boudouani and Lounis Kessal. Vers une méthodologie pour la programmation des architectures à reconfiguration dynamique. *In Actes du workshop Adéquation Algorithmes Architectures En Traitement Du Signal et de L'image*, pages 15–20, Janvier 2000.