

Systeme temps réel de vision périphérique 3D pour l'analyse de scènes dynamiques

Peripheral Vision by a Real Time 3D System for Analysing Dynamic Scenes

par Skander GUETARI*, Jean DEVARs*, Ryad BENOSMAN*, Thierry MANIERE**

*Perception, Automatique, Réseaux Connexionnistes (PARC)

**Laboratoire d'Electronique Analogique et Micro-ondes (LEAM)

Université Pierre et Marie Curie, PARC, Boite courrier 164,
4, place Jussieu F-75252 Paris cedex 05

résumé et mots clés

De nombreuses applications en robotique exigent des systèmes de vision de plus en plus performants permettant de restituer l'espace 3D observé le plus fidèlement possible et dans un contexte de scène dynamique. Dans cette contribution on propose un système complet temps réel (comprenant la conception des capteurs, leur disposition, et le traitement informatique de l'information) visant à observer une scène de façon périphérique, c'est à dire couvrant 360° d'angle, et à évaluer les distances des objets observés.

Stéréovision, Temps réel, Mouvement, Vision périphérique, Caméra tournante.

abstract and key words

Many applications in robotics need enhanced vision systems able to give a reliable 3D information of the focused objects in a dynamic scene context. In this paper we propose a complete system (including the conception of sensors, their disposition and the information processing) for observing a scene in a peripheral way, this means that it will cover 360° of angle, and for evaluating the observed objects' distances.

Stereovision, Real Time, Motion, Peripheral Vision, Rotating Camera.

1. introduction

La vision 3D est un sujet souvent abordé. La démarche classique repose sur le principe dit de triangulation, qui consiste à déterminer les coordonnées 3D d'un point à partir d'au moins deux images prises sous deux angles différents [6]. Dans ce qui suit, on mettra en évidence les caractéristiques qui différencient le système proposé des systèmes classiques.

La stéréovision implique, dans un premier temps, de disposer plusieurs capteurs, généralement des caméras matricielles, observant la scène à analyser sous des angles différents. On dispose ainsi de deux images (ou plus, mais le problème de base repose sur deux prises de vue), donc de deux projections dans deux repères 2D de la scène à analyser. Les constructeurs ne fournissent d'information que sur les caméras elles-mêmes, et souvent avec une précision insuffisante. L'assemblage de celles-ci n'est pas totale-

ment maîtrisé. C'est pourquoi il est nécessaire de déterminer les paramètres de l'ensemble par calibrage du système.

1.1. transformation de perspective directe

Elle consiste en la projection de la scène sur les matrices des caméras [figure 1]. Cependant afin d'exploiter l'information issue des deux caméras, il faut avoir fait au préalable un étalonnage géométrique des deux capteurs. Cette opération est appelée calibrage, ou calibration, de caméras. Le calibrage dans le cas des capteurs classiques implique le calcul des 11 paramètres intrinsèques et extrinsèques du système de vision [13]. Pour ce faire on choisit N points indépendants de l'espace, avec $N \geq 6$, et on effectue un calcul d'inversion de la matrice de calibrage à $2N$

lignes et à 11 colonnes. Cette matrice n'étant pas carrée, on calcule sa pseudo-inverse et en minimisant l'erreur commise selon l'un des critères d'optimisation connus [6].

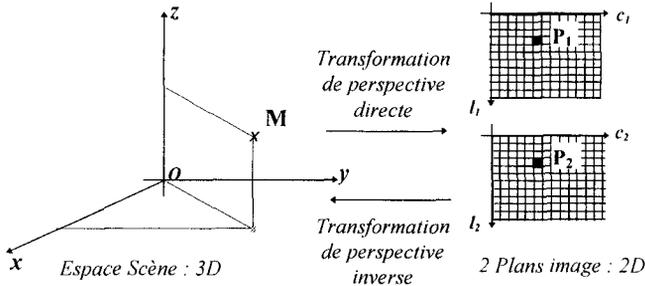


Figure 1. – Transformations de perspective directe et inverse.

1.2. transformation de perspective inverse

Les paramètres du système étant ainsi déterminés, il s'agit de restituer les positions 3D des objets de la scène à partir d'un couple d'images stéréoscopiques. Pour ce faire, il faut chercher pour chaque primitive de la première image, la primitive correspondante dans l'autre image. Cette recherche des points homologues est l'opération la plus délicate à réaliser en vision binoculaire. Dans la littérature [1][6][9][13], on cherche pour chaque point les couples de droites appelées droites épipolaires conjuguées sur lesquelles se trouvent les points homologues [figure 2]. Dans un système idéal, les épipolaires se confondent avec les lignes des matrices CCD, mais en réalité cette hypothèse n'est généralement pas vérifiée, et un calcul est nécessaire pour résoudre le problème de géométrie 3D. Ce calcul des droites épipolaires n'est pas dépourvu d'erreurs. De plus, à cause de l'échantillonnage et de l'orientation des plans images, on peut ne pas trouver la primitive homologue sur la droite épipolaire conjuguée, mais au voisinage de celle-ci. La mise en correspondance nécessite donc une

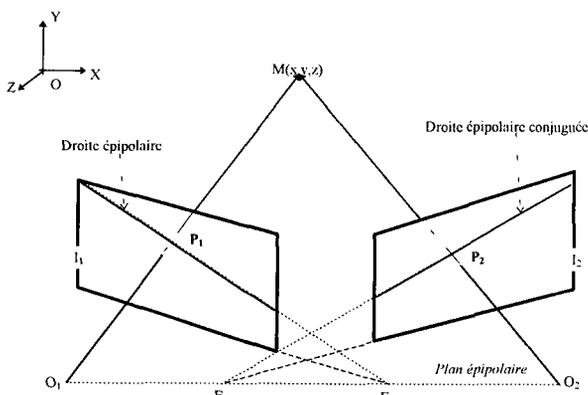


Figure 2. – Les droites épipolaires sont les traces du plan épipolaire dans les plans image.

recherche dans un voisinage 2D centré sur le point courant de la droite épipolaire.

Le point M est projeté par le système optique sur les plans I_1 et I_2 en P_1 et P_2 . Cette projection est régie par les lois de l'optique géométrique, liant la distance du point M , ses coordonnées relatives et la distance focale du système optique [13, 6]. Ce sont les inverses de ces relations qui permettent de calculer les coordonnées 3D du point M , dans le repère de la scène.

1.3. conclusion sur la vision 3d classique

Les difficultés liées à cette démarche peuvent se résumer comme suit :

- Le calcul de triangulation nécessite de déterminer la matrice 4×4 , déduite du calibrage. Le calcul des coordonnées 3D du point observé implique le calcul d'une pseudo-inverse, calcul lourd et sensible aux incertitudes dans la détermination des coefficients de cette matrice.
- Pour une prise de vue, la recherche des homologues s'effectue sur un nombre élevé de pixels.
- Pour obtenir une vision périphérique par cette méthode, il faut faire tourner les capteurs et, par voie de conséquence, déterminer à chaque position angulaire de nouvelles matrices de passage.

Notre idée est de repenser complètement le système de prise de vue avec des contraintes de vision périphérique et de temps réel. Il est nécessaire de redéfinir complètement le système de vision afin d'assurer sa parfaite adéquation avec le traitement informatique qui lui sera appliqué, ceci afin de simplifier ce dernier au maximum dans un souci de temps réel, donc de réalisation câblée. Ce système s'inscrit dans la famille des dispositifs passifs.

2. conception du système

Lorsque l'on parle de vision périphérique 3D, on cherche implicitement à projeter l'espace sur un cylindre. Ce problème étant de géométrie cylindrique, il est naturel d'imaginer des capteurs respectant cette symétrie et de les inclure dans un système lui-même de symétrie de révolution [5]. Notre solution pour réaliser un capteur d'images de géométrie cylindrique, consiste à mettre en rotation une barrette CCD [figure 3].

Un tel dispositif a déjà été proposé [11], mais un déplacement radial de la caméra était nécessaire pour obtenir une vision stéréoscopique. Ceci est inacceptable pour un système temps réel puisque, pour une position angulaire donnée du système, il faut attendre que la caméra se déplace entre les positions radiales afin d'obtenir une paire d'images.

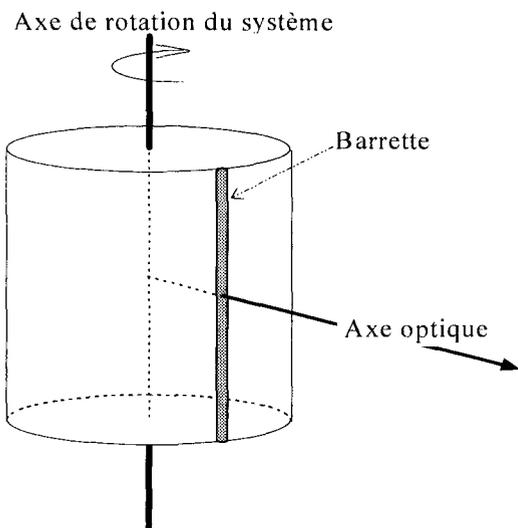


Figure 3. – Cylindre de projection.

Dans notre cas, afin d'avoir simultanément deux prises de vues et sachant que ce système doit admettre une symétrie cylindrique, il faut positionner deux capteurs l'un au-dessus de l'autre sur un même axe de rotation [figure 4]. Cette disposition permet des simplifications dans l'algorithme du calcul ultérieur.

Le choix de barrettes CCD par rapport aux matrices CCD est motivé par leur relative simplicité de lecture, mais aussi par le fait qu'on peut avoir une meilleure résolution pour une dimension donnée du capteur.

y: Axe de rotation

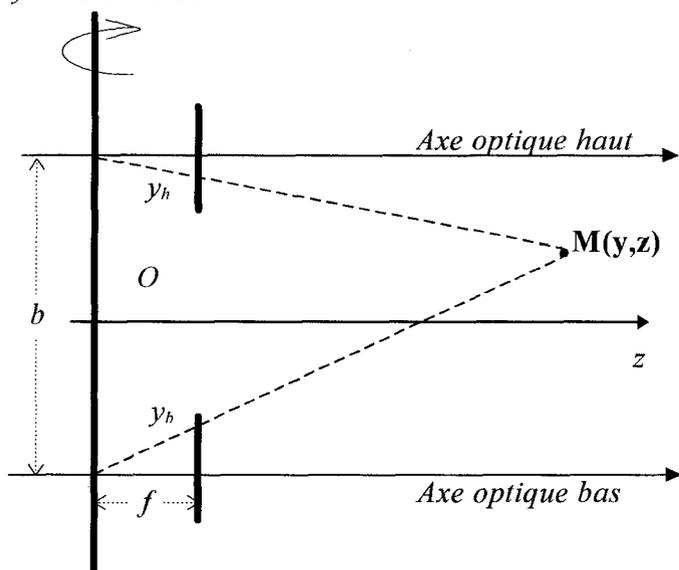


Figure 4. – Système global.

2.1. exploitation des images

Ce système donne simultanément deux images. Sachant qu'on maîtrise totalement les constituants du capteur et leur disposition, on n'effectue pas de calibrage classique, mais un réglage initial à partir d'une mire élémentaire [3]. En outre, si un point est perçu par les deux barrettes, il est simultanément sur les deux barrettes. Le problème de la recherche des points homologues se ramène à une dimension, et on s'intéresse au calcul des coordonnées d'un point de la scène dans le plan de vision. Par triangulation planaire, on peut exprimer les coordonnées 3D d'un point *M* en fonction de ses projections, en prenant comme origine du repère le milieu de la base, et en considérant un modèle sténopé pour la barrette et son optique associée :

$$\begin{aligned} z &= \frac{fb}{|y_h - y_b|} \\ y &= \frac{|y_h + y_b|b}{|y_h - y_b|2} \end{aligned} \quad (1)$$

Où *y* et *z* sont les coordonnées du point observé, *f* est la distance focale équivalente, *y_h* et *y_b* sont les coordonnées des images haute et basse de *M*, *b* est la base du système stéréo [figure 4].

Des prises de vues, obtenues par un prototype de ce capteur montrent que la projection particulière de l'espace scène crée des déformations géométriques des lignes horizontales, voir les images des figures 14.a et 14.b. De façon plus rigoureuse, les segments horizontaux sont transformés, dans le plan image développé, en arcs de sinusoides dont les paramètres dépendent de leur position 3D [figure 5]. Cette particularité peut être exploitée pour obtenir une information qualitative de position (à un coefficient près), en vision monoculaire.

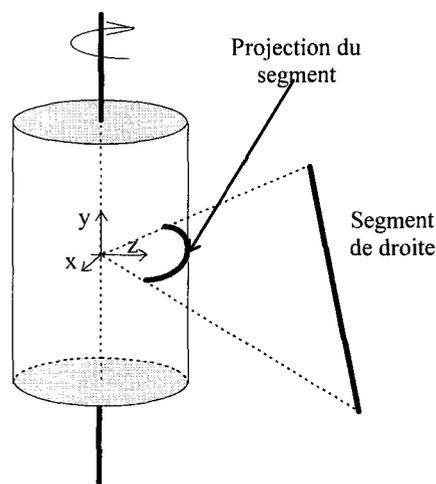


Figure 5. – Projection d'un segment de droite de l'espace sur le cylindre.

2.2. avantages du système envisagé

- Elimination du calcul des droites épipolaires : si un point est perçu par les deux caméras, ses projections se trouvent simultanément sur les barrettes. Donc, la recherche des points homologues se fait sur une seule dimension, ce qui restreint le domaine de recherche et réduit la probabilité des faux appariements.
- Elimination du calibrage classique, remplacé par un réglage à l'assemblage des barrettes et des optiques correspondantes. Les imprécisions qui subsistent deviennent du second ordre alors que pour les systèmes classiques, les données sont pratiquement inexploitable sans étalonnage géométrique.
- La triangulation ne fait plus appel qu'à un simple calcul scalaire, d'où gain en temps de calcul, et la possibilité d'une réalisation électronique.

2.3. objets mobiles et détection d'obstacles

La géométrie de prise de vue facilite l'analyse de trajectoire du mobile. La régularité de la rotation du capteur est exploitée afin d'estimer la vitesse d'un objet en mouvement. On assimile un mobile non déformable à l'un de ses points caractéristiques (éventuellement le centre de gravité du mobile) et son mouvement entre deux vues consécutives à un mouvement rectiligne uniforme. Ainsi pour chaque couple de vues consécutives, on peut calculer le segment approximant la trajectoire. L'angle effectué entre deux vues consécutives d'un point caractéristique correspond à $\Delta\theta$. La vitesse angulaire ω , imposée par le moteur, est quant à elle directement mesurée. Connaissant la position de l'objet à deux instants, on en déduit sa vitesse relative entre ces deux instants [figure 6]. La discrétisation des images saisies par les barrettes fixe la précision de localisation. La période de rotation, limitée par la vitesse des traitements, détermine l'écart de temps maximum entre les deux observations. Ce sont des éléments du cahier des charges qu'il convient de préciser.

Calculons le temps Δt entre deux vues consécutives du mobile.

$$\Delta\theta = \omega \Delta t \Rightarrow \Delta t = \frac{\Delta\theta}{\omega}$$

C'est aussi le temps nécessité pour que le mobile se déplace de la position P vers la position P' .

La norme de la vitesse est déterminée par la formule suivante :

$$\|\vec{V}\| = \frac{\|\overrightarrow{PP'}\|}{\Delta t} = \frac{\|\overrightarrow{PP'}\|}{\Delta\theta} \omega$$

Le sens du vecteur \vec{V} est donné par le sens du vecteur $\overrightarrow{PP'}$.

Sachant que l'axe optique à $t = 0$ est confondu à l'axe z , les angles θ_i et θ_j sont parfaitement connus. Après la mise en

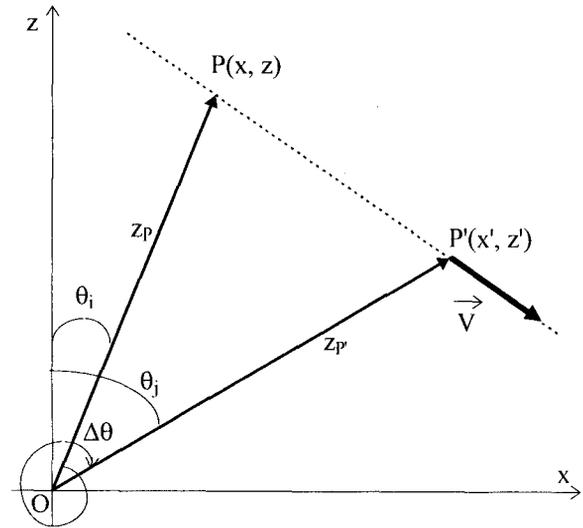


Figure 6. - Représentation d'un mobile à deux instants successifs dans le repère caméra (vue de dessus).

correspondance et la triangulation, on détermine les coordonnées z_P et $z_{P'}$ des points P et P' . Les coordonnées des points P et P' dans le repère (O, x, z) sont alors :

$$P \begin{cases} x = z_P \sin(\theta_i) \\ z = z_P \cos(\theta_i) \end{cases} \quad \text{et} \quad P' \begin{cases} x' = z_{P'} \sin(\theta_j) \\ z' = z_{P'} \cos(\theta_j) \end{cases}$$

2.4. méthode de détection

C'est une méthode statistique servant à détecter les objets en mouvement dans une séquence d'images. Notre méthode, s'appuyant sur l'occurrence des images en fonction du temps, combine une construction du fond en niveau de gris et une comparaison des gradients calculés sur l'image courante avec ceux d'une image de référence[7]. L'image de référence est fabriquée par mémorisation des gradients constamment présents.

- Les images de référence I_{ref} sont initialisées à 0.
- Elles sont mises à jour selon G_c , image actuelle des gradients, d'après l'équation de récurrence suivante :

$$\text{Si } G_c < 0 \Rightarrow I_{ref} = I_{ref} - 1$$

$$\text{Si } G_c > 0 \Rightarrow I_{ref} = I_{ref} + 1$$

$$\text{Si } G_c = 0 \Rightarrow I_{ref} = I_{ref} + \alpha \text{ signe}(I_{ref})$$

où I_{ref} est l'état actuel de l'image de référence. α est un coefficient de pénalité, dépendant de la nature des images, qui permet de régler le facteur « d'oubli » de la mémorisation des gradients lorsque le signe devient indéterminé; donc en présence d'une zone homogène qui pourrait être une zone en mouvement. L'image de référence, après une phase initiale de latence, se stabilise donc sur une carte des gradients permanents. Il s'agit a priori des gradients correspondant aux zones fixes de l'image, donc du fond.

3. définition et choix des paramètres

Les contraintes sont liées aux caractéristiques statiques et dynamiques de la scène à observer et à la précision que l'on se fixe. Les paramètres du système sur lesquels on peut jouer sont au nombre de six : la vitesse de rotation (qui n'intervient que dans l'estimation du mouvement), la résolution angulaire, l'ouverture optique, la focale, la base, la taille et le nombre de pixels des barrettes.

Une simulation du comportement du système, avec des valeurs cohérentes de ces paramètres, a fourni des abaques représentant les erreurs quadratiques relatives que l'on commet sur l'évaluation de la distance. En effet, l'erreur augmente bien sûr en fonction de la distance de vision, mais fluctue à cause de la discrétisation des images [figure 7], une erreur quadratique moyenne est plus facilement exploitable. C'est en fonction de la distance de vision, élément le plus pertinent dans le contexte d'une vision d'alerte, que ces courbes sont représentées [figures 8 et 9]. Sur des images de synthèse, on a vérifié que le système stéréoscopique proposé, conçu avec des composants standards, peut percevoir jusqu'à 40 mètres avec une précision de l'ordre du mètre satisfaisant le cahier des charges.

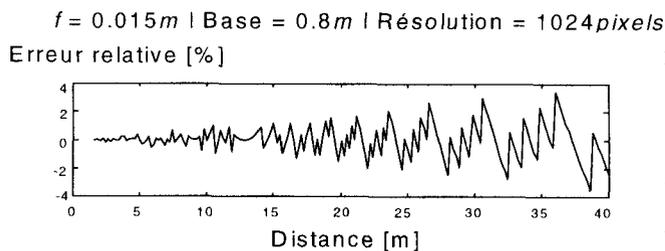


Figure 7. – Effet de l'échantillonnage, variation de l'erreur en fonction de la distance de vision en prenant pour modèle de caméra le modèle sténopé.

- **Base** : La précision de la mesure est bien sûr d'autant meilleure que l'espacement entre les capteurs est grand.
- **Focale** : Une augmentation de la distance focale équivalente conduit à une meilleure précision de la localisation.
- **Résolution optique** : Le champ observé par la barrette est fonction de la focale du système optique choisi : plus la distance focale est grande, plus l'angle d'observation est petit, ce qui mène à une meilleure résolution de l'échantillonnage. Ce paramètre est donc cohérent avec le précédent.
- **Résolution angulaire** : En général fixé par la résolution optique, le pas mécanique de rotation peut être choisi inférieur à l'angle d'observation, la redondance de l'information pouvant alors être mise à profit pour une approche de type sub-pixel. Ceci implique cependant une augmentation du volume d'information à traiter. Inversement, une diminution de la résolution angulaire

$f = 0.015m$ - Résolution = 1024 pixels

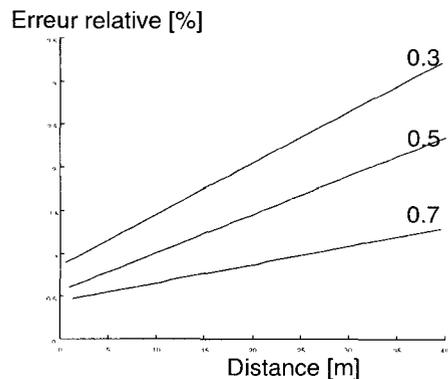


Figure 8. – Evolution de l'erreur relative en fonction de la distance à base constante.

Base = 0.8m - Résolution = 1024 pixels

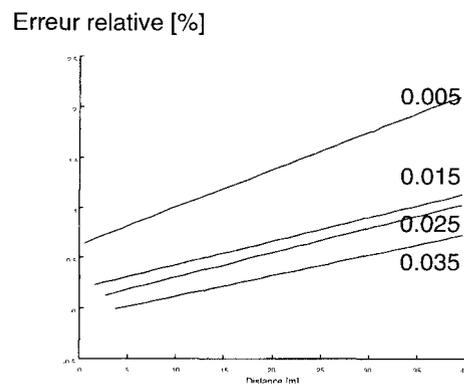


Figure 9. – Evolution de l'erreur relative en fonction de la distance à focale constante.

sous-échantillonne l'espace, mais permet d'accroître la vitesse de rotation. En prenant comme critère la séparation optique de points distants de 0,1 m à 40 m, une résolution de 0.1 degrés est suffisante, ce qui correspond à des images de taille 1024 × 3600 pixels.

4. accélération de la mise en correspondance

La simplicité du système de vision permet d'ajouter quelques considérations géométriques pour réduire le temps de calcul, plus particulièrement pour une mise en correspondance rapide et efficace des primitives images[8]. La complexité de la mise en correspondance dépend de la méthode utilisée, ainsi que des propriétés physiques des images. Dans notre cas, la mise en correspondance se fait entre pixels, par une technique de

corrélation basée sur la radiométrie et le contraste, qui peut être implantée dans un circuit électronique compte tenu du caractère linéaire des capteurs.

La figure 10 illustre les trois paramètres à discuter :

- Le sens de recherche de l'homologue d'un pixel,
- L'intervalle de recherche,
- La zone interdite à la recherche, c'est-à-dire zone de vision monoculaire.

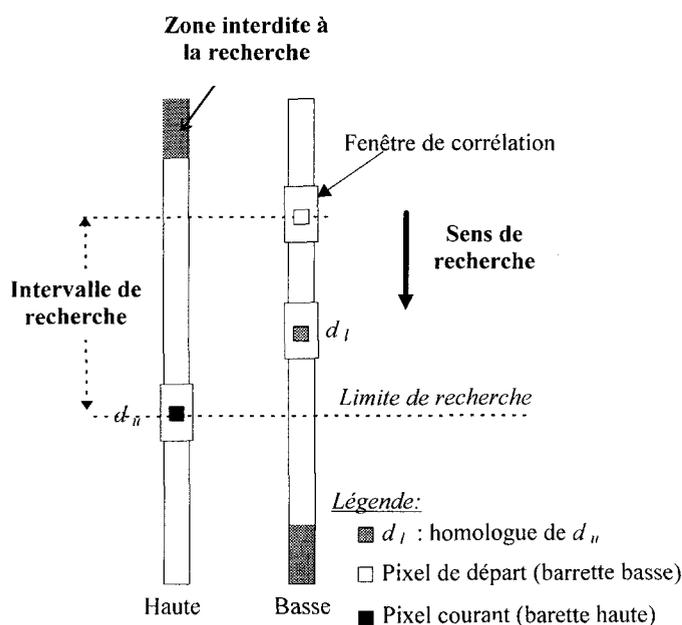


Figure 10. – Mise en correspondance, paramètres de recherche des pixels homologues.

4.1. sens et intervalle de recherche

La distance d'un point objet est inversement proportionnelle à la disparité de ses images (équation 1) or, dans un contexte de vision d'alerte, il faut favoriser la vitesse de détection des objets les plus proches. Il est donc préférable de déplacer la fenêtre d'analyse depuis la position correspondant à la disparité la plus grande [figure 10], vers la position correspondant à une disparité nulle (cas d'un point objet théoriquement à l'infini, en fait à la limite des possibilités de résolution du système). En pratique, la distance utile de perception est limitée par la perte de précision [figures 8 et 9]. A une distance utile maximale Z_{max} correspond une disparité minimale D_{min} , non nulle, qui peut encore réduire la plage de recherche. Le sens de recherche étant ainsi imposé, il convient de délimiter l'intervalle de recherche, donc déterminer la position initiale de la fenêtre d'analyse.

Nous considérerons, pour la commodité de l'exposé, que les origines des coordonnées pixels y_b et y_h se situent au milieu des barrettes de longueur L [figure 11]. On peut alors caractériser trois

zones du plan de vision par les signes des coordonnées des pixels image, et par les stratégies de recherche employées :

- La zone 1 correspond à y_h positive, la recherche de l'homologue s'effectue sur la barrette basse, depuis $y_b = L/2$, jusqu'à $y_b = y_h$.
- La zone 2 est la symétrique de la zone 1, y_b est négative, l'analyse fait évoluer y_h de $-L/2$ à y_b .
- La zone 3, située entre les axes optiques, correspond au complément des deux précédentes. Dans ce cas, le choix de la barrette de recherche est indifférent, et l'évolution doit se faire systématiquement sur une demi-barrette, de l'extrémité à l'origine, avec des coordonnées de signe contraire à la coordonnée du pixel de référence.

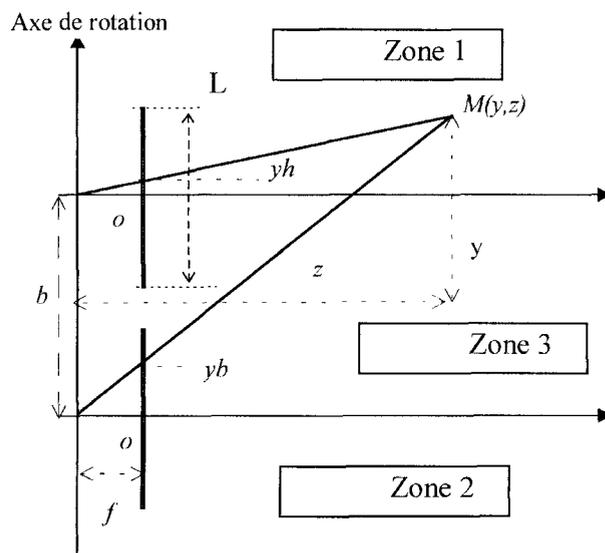


Figure 11. – Localisation du point objet $M(y, z)$.

Ces simples considérations géométriques permettent de limiter le nombre de pixels à traiter dans l'étape de mise en correspondance.

4.2. zone interdite à la recherche

La zone interdite à la recherche (ZIR) est une région d'une barrette, qui perçoit une région de l'espace invisible à l'autre barrette [figure 12]. Etant donné la symétrie du système, la ZIR est la même pour les deux capteurs, il suffit de la déterminer pour la barrette haute. Le but visé consiste à éviter la recherche de l'homologue d'une primitive dans une région où il est strictement impossible de trouver un homologue réel. Cette restriction donne des résultats plus sûrs en évitant les fausses correspondances.

Définissons, pour la barrette haute, trois régions du plan, explicitées par la figure 12 :

- **Région stéréo** : Cette région est vue par les deux capteurs,

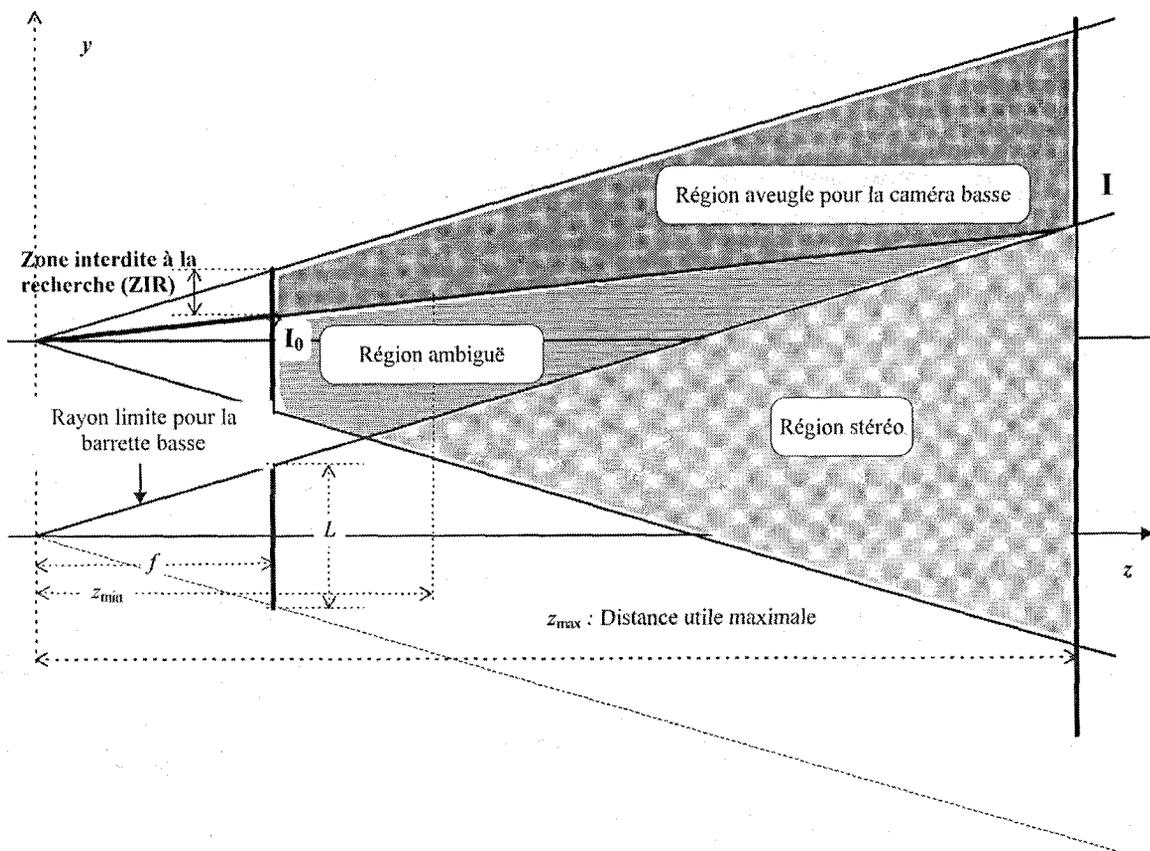


Figure 12. – Définition des régions de visibilité.

- **Région aveugle pour la barrette basse** : Cette région n'est pas visible par la barrette basse,
- **Région ambiguë** : La projection d'un point dans cette zone peut être confondue avec la projection d'un autre point dans la région stéréo (cette ambiguïté a lieu également pour des systèmes traditionnels).

Nous avons vu précédemment que la limite de perception du système, ou distance utile maximale Z_{max} , est fixée par la précision exigée. La limite de la zone aveugle pour la barrette basse à cette distance (point I) se projette en I_0 , borne de la ZIR . Ce qui implique que ZIR représente la disparité du point I , en limite du champ de vision stéréo, situé à la distance Z_{max} :

$$ZIR = \frac{fb}{z_{max}}$$

On constate que la ZIR dépend des trois paramètres, la distance focale f , la distance utile maximale z_{max} , et la base du système b . Sa valeur doit être très petite par rapport à la taille de la barrette L . Dans la stratégie d'appariement présentée précédemment, ZIR intervient dans les zones 1 et 2 [figure 11], pour inhiber le processus de recherche lorsque, soit y_h , soit y_b se trouvent dans la zone interdite.

Nous avons tracé les variations de la ZIR en fonction des paramètres géométriques du système [figure 13], ce qui apporte une information complémentaire sur l'adéquation des paramètres du système de vision.

4.3. méthode de mise en correspondance

Partant de deux colonnes image I_h et I_b , nous devons localiser les homologues des primitives de I_h dans I_b issues d'un même objet. C'est parmi les tâches les plus difficiles en vision artificielle[6]. Il suffit, en effet, que les deux images aient une radiométrie différente ou que des objets soient occultés d'une prise de vue à une autre, pour qu'on soit confronté à des problèmes algorithmiques complexes. Plusieurs méthodes sont proposées dans la littérature[12]. Leurs différences se manifestent dans les trois points suivants :

- Choix des primitives (jonctions, segments, pixels).
- Démarche suivie pour appairer les primitives choisies.
- Temps de calcul.

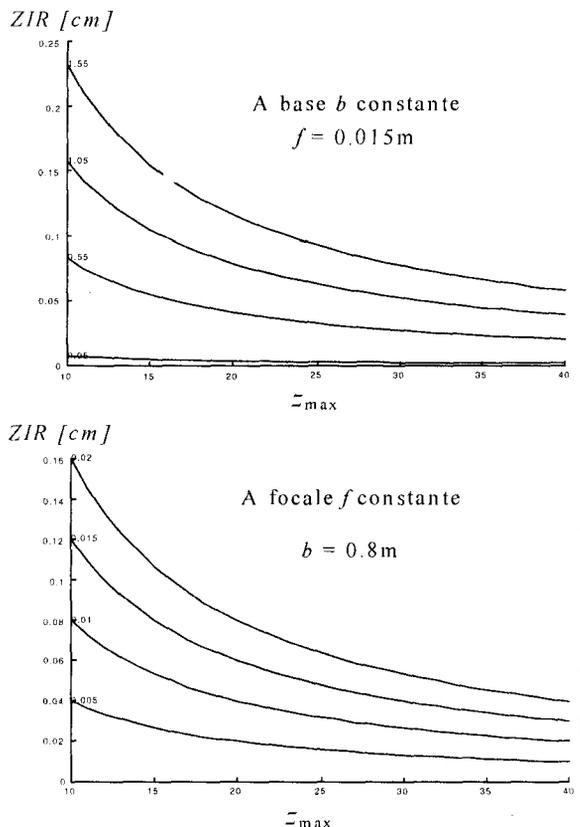


Figure 13. – Variation de ZIR en fonction de Z_{max} .

La primitive qu'on doit utiliser dans notre cas est imposée par le système. C'est le pixel. On cherche les contours pertinents (maximums locaux des gradients de G_c) afin d'obtenir une image ternaire I_t . Celle-ci représente le signe des gradients significatifs, c'est à dire dont le module est supérieur à une valeur minimale G_{mini} paramétrable en fonction du rapport signal sur bruit :

$$\begin{aligned} \|G_c\| < G_{mini} &\Rightarrow I_t = 0 \\ \|G_c\| \geq G_{mini} &\Rightarrow I_t = \text{signe}(G_c) \end{aligned}$$

À ce stade, on dispose de deux images ternaires contenant les valeurs $(-1, 0, 1)$. Il faut essayer d'apparier les pixels de la colonne haute et les pixels de la colonne basse. C'est ainsi qu'on a défini une mesure de ressemblance donnant les coordonnées des points **candidats homologues** à l'appariement.

Mesure de ressemblance

On définit une fonction $c(P)$ permettant de caractériser les propriétés intrinsèques d'un voisinage centré sur le pixel P à traiter. On définit par la suite la fonction R de mesure de ressemblance de la façon suivante : pour tout pixel P de la colonne image haute, on associe le pixel P^i de la colonne image basse appelé candidat homologue si seulement si :

$$R(P^i) = \|c(P) - c(P^i)\| < s$$

avec s un seuil, homogène au module du vecteur des différences, en-dessous duquel on suppose que les deux pixels sont bien des points homologues. On forme par la suite un ensemble avec tous les candidats et on garde le P^i qui présente la plus forte mesure de ressemblance. c retourne un vecteur M de dimension 4, défini de la façon suivante :

$$\begin{aligned} M(1) &= \text{somme}(V_t < 0) ; \quad M(2) = \text{somme}(V_t == 0) ; \\ M(3) &= \text{somme}(V_t > 0) ; \quad \text{et } M(4) = \text{moyenne}(V). \end{aligned}$$

V et V_t sont respectivement le voisinages en niveau de gris et le voisinage ternaire d'un pixel quelconque. La fonction *somme(...)* fournit le cardinal de l'ensemble des éléments répondant à la condition entre parenthèses et la fonction *moyenne(...)* calcule la moyenne des niveaux de gris du voisinage V .

Il est évident que le calcul de R ne se fera que sur les gradients non nuls, de même signe en excluant la ZIR . Une étude comparative de cette méthode avec les méthodes classiques est en cours.

5. conclusion

Les performances obtenues par ce nouveau système de vision sont résumées dans le tableau comparatif ci-dessous [tableau 1]. Son intérêt est évident pour toutes les applications où la contrainte de temps réel est impérative. En effet, la réduction drastique du nombre et de la complexité des calculs permet d'envisager une architecture flot de données, traitant les deux images au rythme vidéo, et fournissant à la fois l'information radiométrique, la segmentation, mais aussi pour les points caractéristiques la position dans le repère cylindrique lié au capteur.

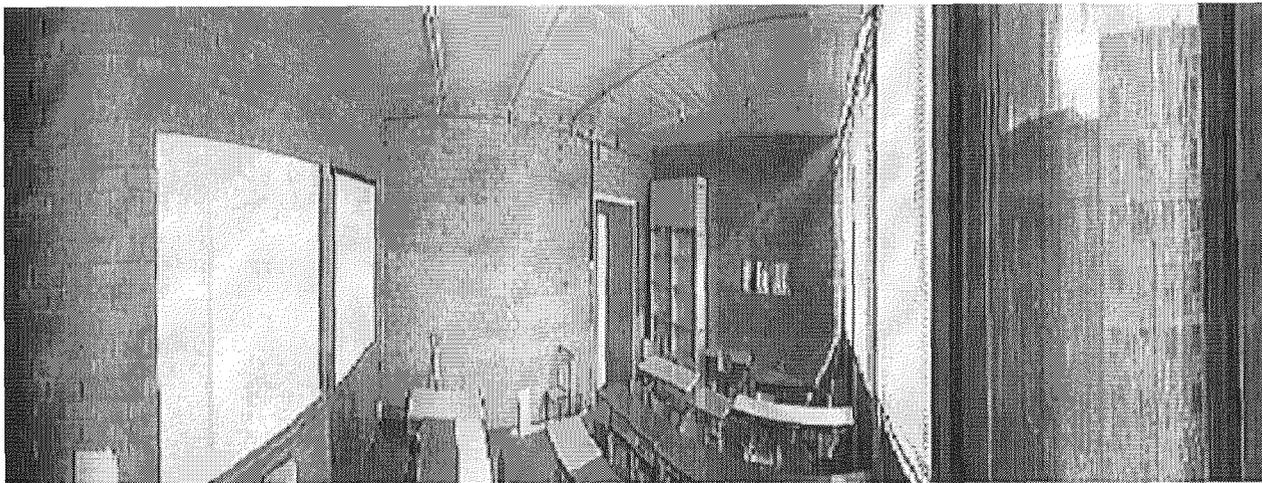
Par ailleurs, la simplicité des propriétés géométriques de ce système autorise une méthode efficace de réduction du temps de calcul dans le procédé de recherche des points homologues.

Le contexte d'utilisation est dans des scènes pour lesquelles l'environnement est constitué d'objets essentiellement polyédriques. Les déformations géométriques des segments rectilignes permettent une analyse qualitative des distances, même dans un contexte monocapteur, c'est-à-dire dans les régions aveugles à l'une des barrettes du système. Ceci augmente le champ de la vision d'alerte.

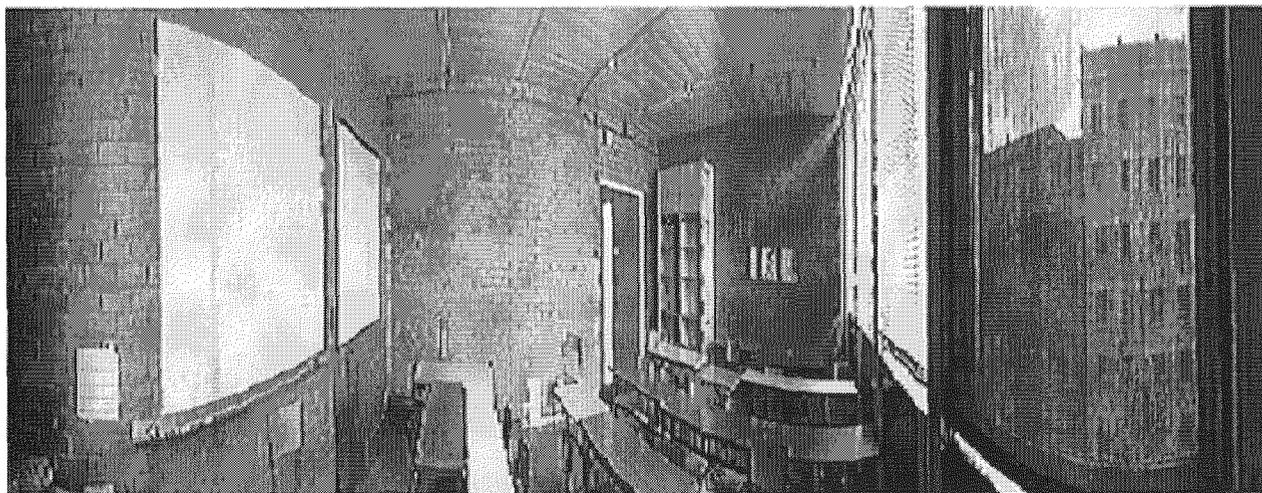
La réalisation d'un prototype, et l'évaluation de ses performances est en cours, dans le cadre d'un projet commun au groupe « Perception » du laboratoire PARC, qui assure la définition de la partie algorithmique, et au LEAM, qui se charge de la réalisation électronique du système.

Tableau 1. – Comparaison des caractéristiques des deux systèmes

	Système de vision classique	Système de vision proposé
Perspective directe	Calcul de la matrice de calibrage (12 × 11 minimum par point de vue).	Réglage des composants du système à l'assemblage
Perspective inverse	Calcul des droites épipolaires	Epipolarité implicite.
Complexité	<ul style="list-style-type: none"> • Recherche d'homologues 2D, • Inversions de matrices 4 × 4, • Minimisation d'erreurs 	<ul style="list-style-type: none"> • Recherche d'homologues 1D, • Opérations scalaires élémentaires (comparaisons, additions, etc.)
Temps de calcul	<ul style="list-style-type: none"> • Acceptable pour simple version binoculaires • Très lent pour vision périphérique 	<ul style="list-style-type: none"> • Système temps réel



a) Image haute



b) Image basse

Figure 14. – Couple d'images périphériques 360° d'une scène d'intérieur.

BIBLIOGRAPHIE

- [1] N. Ayache, « Construction et fusion de représentations visuelles 3D », Thèse Paris XI, Orsay 1988.
- [2] D.H. Ballard, C.M. Brown, *Computer Vision*, Prentice Hall, London, 1982.
- [3] A. Basu, « Active Calibration of Cameras : Theory and Implementation », *IEEE Trans. on Systems Man and Cybernetics*, Vol. 25, N°2, February 1995, pp. 256–265.
- [4] J.E. Besançon, *Vision par ordinateur en deux et trois dimensions*, Eyrolles, Paris 1988.
- [5] J. Devars, S. Guetari, T. Manière, « Système temps réel de vision périphérique pour l'analyse de scènes 3D dynamiques », *Journées d'études SEE – GDR 134 sur le mouvement dans les images et les signaux*, Paris juin 1995, pp. 28–29.
- [6] O. Faugeras, *Three-Dimensional Computer Vision, A Geometric Viewpoint*, MIT Press, London, 1993.
- [7] S. Guetari, J. Devars, « Méthode Statistique pour la Détection du Mouvement dans une Séquence d'Images », *Canadian Conference on Electrical and*

- Computer Engineering*, IEEE Canada, Montréal, September, 1995, pp. 459–463.
- [8] S. Guetari, J. Devars, « Taking Into Account System Properties for Real Time Stereo Vision, A Geometric Approach », *IASTED, Signal and Image Processing SIP-95*, November 1995, pp. 1021–1025.
- [9] R. Horaud, O. Monga, *Vision par Ordinateur; outils fondamentaux*, Hermès, Paris, 1993.
- [10] T.S. Huang, *Topics in Applied Physics : Picture Processing and Digital Filtering*, Second Edition, Vol. 6, Springer-Verlag, Berlin, 1979.
- [11] H. Ishiguro, M. Yamaoto, S. Tsuji, « Omni-Directional Stereo », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 14, N°2, February 1992, pp. 257–269.
- [12] Y. Shirai, *Three-Dimensional Computer Vision*, Springer-Verlag, Berlin, 1987.
- [13] G. Toscani, « Système de calibration et perception du mouvement en vision artificielle », Thèse Paris XI, Orsay, 1987.

Manuscrit reçu le 18 septembre 1995.

LES AUTEURS

Skander GUETARI



Skander Guetari est ingénieur Génie Electrique de l'Ecole Nationale d'Ingénieurs de Tunis (ENIT) en 1992. Il a obtenu son DEA de robotique à l'Université Pierre et Marie Curie (UPMC) en 1993. Il est actuellement en dernière année de thèse à l'UPMC et a travaillé sur divers projets en génie informatique à dominante algorithmique et électronique. Le thème actuel de sa recherche est l'étude algorithmique du projet.

Jean DEVARs



Jean Devars, Docteur Es Sciences en 1984 est professeur depuis 1985. Il anime depuis 1992 l'équipe vision du laboratoire Perception Automatique et Réseaux Connexionistes. Il est à l'origine de la conception du système de vision.

Ryad BENOSMAN



Ryad Benosman obtient sa maîtrise d'informatique à l'université Paris Nord en 1993, suivie d'un DEA en Robotique en 1994 à l'Université Pierre et Marie Curie. Il est actuellement en thèse et travaille depuis deux ans sur la vision robotique plus spécialement sur le calibrage et la reconstruction d'environnements perçus par le capteur omnidirectionnel.

Thierry MANIERE



Thierry Manière, diplômé du DEA d'électronique option hyperfréquences de l'Université Pierre et Marie Curie (UPMC) en 1993, termine une thèse au laboratoire d'Electronique Analogique et Micro-ondes de l'UPMC, sur le thème de la réalisation électronique du système de vision.