Performance du capteur vision pour le positionnement 3D dans le cadre de la détection d'obstacles routiers

Sébastien Glaser, Didier Aubert et Jean-Phillippe Tarel
LIVIC - INRETS/LCPC
13, route de la minière
78000 Versailles - Satory
glaser@lcpc.fr

Introduction

Les accidents de la route présente, en France, un bilan lourd en terme à la fois de morts (7242 en 2002) et de blessés graves (24091 personnes ont nécessité une hospitalisation de plus de 6 jours). Les efforts conjoints des gestionnaires de l'infrastructure et des pouvoirs publics permettent de faire baisser d'années en années ces chiffres. De leur côté, les constructeurs et les équipementiers proposent des voitures de plus en plus sures, intégrant des dispositifs de sécurité passive et active performants, parmi ceux-ci, nous pouvons citer:

- *Sécurité Passive :* caisse dont la déformation est programmée, prétenseur de ceinture, airbag.
- *Sécurité Active :* A.B.S., contrôle du mouvement de lacet (E.S.P.)

Mais ces systèmes de sécurité passive ne font que limiter la gravité et la mortalité des accidents. Les systèmes actifs, quant à eux, ne font que repousser les limites de la contrôlabilité du véhicule, les limites physiques restent, elles, inchangées. De plus, le conducteur est souvent à l'origine de l'accident du fait d'un défaut d'attention ou d'une mauvaise estimation de la dangerosité de la situation, [11,12] montre que l'on peut relier 75% des accidents à ces problèmes.

Aussi, les systèmes dits de sécurité prédictive présente un grand intérêt. Ces systèmes doivent être capable de percevoir leur environnement, de prédire l'évolution du véhicule et de donner une alerte (au conducteur ou à un automate) lorsqu'une situation présente un caractère accidentogène. Ainsi, l'arrêt sur obstacle permettrait de diminuer fortement la gravité et la mortalité de certains accidents. En effet, comme le montre la figure 1 du L.A.B. (Laboratoire d'Accidentologie et de Biomécanique, GIE PSA/Renault), la mortalité et la gravité d'un accident pour le passager d'un véhicule léger (ici avec ceinture et sans AirBag) peut être directement relié à la différence de vitesse avant et après le choc (Equivalent Energetic Speed).

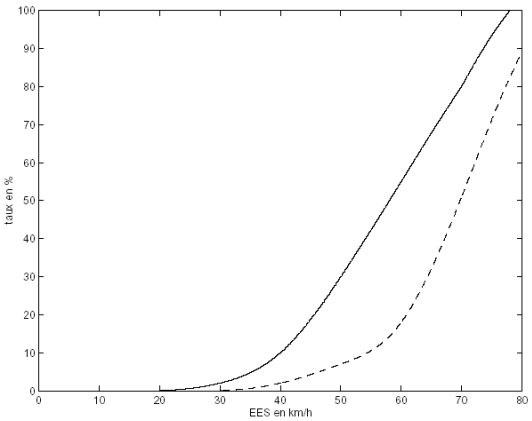


Fig. 1 : Mortalité (trait pointillé) et gravité (trait plein) d'un accident (source : L.A.B.)

Un système automatique détectant un obstacle dangereux et freinant fortement permettrait, sinon d'arrêter le véhicule, de le freiner pour diminuer sa vitesse au moment de l'impact.

Si nous détectons un obstacle dangereux à une distance *D*, en roulant à une vitesse V_0 et que le système a un temps de réaction T et a une décélération γ ($\gamma > 0$), nous pouvons distinguer trois vitesses:

- La vitesse d'arrêt: si la vitesse du véhicule est inférieur à V_A, le véhicule s'arrête avant l'obstacle.

$$V_A = -T\gamma + \sqrt{T^2\gamma^2 + 2D\gamma} \quad (1)$$

- La vitesse maximale: lorsque la vitesse du véhicule est supérieur à $V_{\max}$, le système n'a pas le temps de réagir.

$$V_{max} = \frac{D}{T} \quad (2)$$

- Entre ces deux vitesses, le système permet de réduire la vitesse du véhicule. Au moment de l'impact, sa vitesse est:

$$V_{impact} = \sqrt{V_0^2 + 2\gamma\,(V_0 T - D)} \quad (3)$$

  Le système a permis une diminution de la vitesse de:

$$\Delta V = V_0 \left(1 - \sqrt{1 + 2\frac{\gamma}{V_0^2}\,(V_0 T - D)}\right) \quad (4)$$

Aussi, un tel système doit percevoir de manière précise l'environnement du véhicule. Pour cela, des capteurs comme les radars ou les télémètres lasers semblent tout indiqués. En effet,
ils ont l'avantage de fournir une information précise sur la distance à un objet avec une fréquence élevée. Mais, un système fondé uniquement sur un capteur de ce type présenterait de nombreux problèmes:

- Du fait des vibrations de la caisse, et principalement du mouvement de tangage, la route peut être détectée comme étant un obstacle proche.
- Les objets détectés ne sont pas forcément des obstacles dangereux, un radar ou un télémètre laser ne fournissent pas la position de l'objet sur la route. Aussi, une pile de pont, ou une barrière de sécurité peuvent être pris pour un obstacle dangereux.
- Des objets présents sur la route, mais sur les autres voies, ne sont pas forcement des obstacles dangereux.

L'intégration d'un capteur fondé sur la vision, monocaméra ou stéréo, permet de résoudre ces problèmes: en faisant une détection uniquement fondé sur ce capteur, ou en fusionnant les données provenant des différents capteurs. Dans ces deux cas, l'information de la précision des données provenant du capteur vision est importante, or celle-ci n'est pas ou peu étudiée dans la littérature. Dans cet article, nous proposons d'étudier la précision de deux types de capteur fondés sur la vision. Le premier consiste en une caméra seule placé au niveau du rétroviseur intérieur, le second en un capteur stéréovision dont les caméras sont placées en haut du pare-brise avec un écartement maximal (compte-tenu de la géométrie du véhicule).

Dans la première partie, nous allons rappeler au lecteur les transformations utilisées pour obtenir, à partir du monde en trois dimension, une image en deux dimensions. Dans la partie suivante, nous traiterons le cas inverse qui est de reconstruire le monde en trois dimensions à partir d'une succession temporelle et/ou spatiale d'image. Finalement, nous

comparerons les précisions des deux capteurs.

Le capteur vision

Le but de cette section est double. Tout d'abord, nous allons présenter le capteur vision dans le cadre général, c'est à dire hors du contexte routier. Ceci nous permettra de poser clairement une partie des équations ainsi que les limites de l'étude. Ensuite, ce capteur sera intégré dans le véhicule, et les équations et variables utilisées seront développées.

Présentation du capteur vision

Lorsqu'une seule caméra est utilisée, nous parlerons de vision monocaméra. Lorsque deux caméras sont utilisées conjointement, c'est alors de la stéréovision. Développons tout d'abord les équations nous permettant de connaître les coordonnées d'un point sur une image, en fonction de ses coordonnées dans le monde réel, ainsi que les variables utilisées.

Cas monocaméra

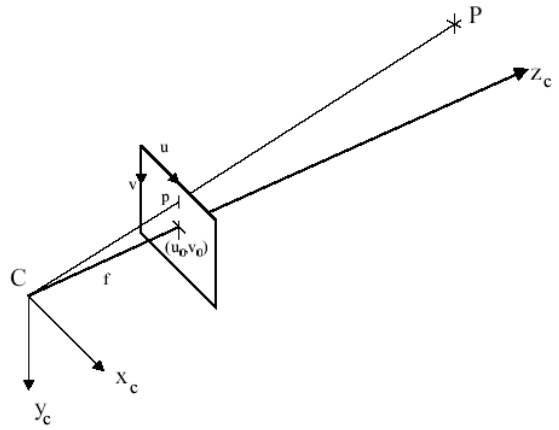


Fig. 2 : Représentation de la caméra

Soit une caméra représentée par son centre optique *C* et son plan image, repéré par le couple $[\vec{u},\vec{v}]^T$ (A^T représente la transposée de *A*). Ce plan image se trouve à la distance *f*, appelée longueur de la focale de la caméra. L'axe optique est l'axe reliant le centre de la caméra et sa projection sur le plan image. Cette projection est représentée sur la figure 2, et les coordonnées de sa projection seront notées $c_0 = [u_0, v_0]^T$.

Un premier repère, le repère image, est formé à partir de $R_i = (c_0, \vec{u}, \vec{v})$. Ensuite, le repère lié à la caméra, $R_c = (C, \vec{X}_c, \vec{Y}_c, \vec{Z}_c)$, est créé à partir du centre de la caméra et de l'axe optique. Les deux autres vecteurs de ce repère sont choisis de sorte qu'ils soient parallèles à ceux du plan image. La projection de points du repère caméra dans le repère image s'exprime alors de la façon suivante:

$$\begin{cases} u = \alpha_u \frac{x}{z} + u_0 \\ v = \alpha_v \frac{y}{z} + v_0 \end{cases} \quad (5)$$

Où $\alpha_u = f / t_u$, $\alpha_v = f / t_v$, t_u et t_v sont les tailles des pixels selon *u* et *v* respectivement. Compte-tenu des caméras actuellement utilisées, nous pouvons faire l'approximation suivante : $\alpha_u \approx \alpha_v \approx \alpha$. L'expression matricielle de cette projection est, en coordonnée homogène:

$$M_{proj} = \begin{bmatrix} \alpha_u & 0 & u_o & 0 \\ 0 & \alpha_v & v_o & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (6)$$

Cas Stéréovision

Pour un capteur fondé sur la stéréovision, le problème est similaire. En effet, en utilisant un repère intermédiaire $R_H = (H, \vec{x}, \vec{y}, \vec{z})$ (figure 3), nous pouvons nous ramener au problème précédent en considérant une transformation supplémentaire permettant de passer du repère R_H au repère R_{C_1}, repère lié à la première caméra, et, de même R_{C_2}, pour la

seconde. Notons ces transformations respectives M_{HC_1} et M_{HC_2}.

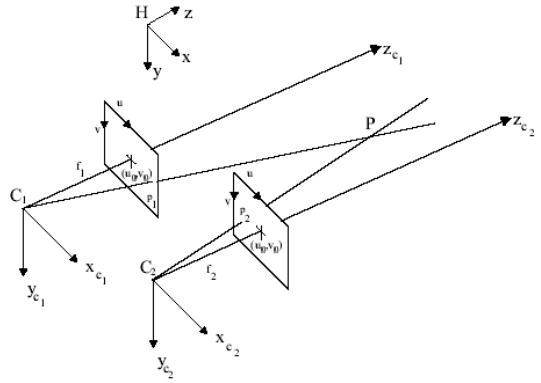


Fig. 3 : Représentation du capteur stéréovision

Dans le cadre de notre application, les cameras sont positionnées de sorte qu'il suffit d'effectuer une translation du repère R_{C_1} suivant l'axe $C_1X_{C_1}$ de L (écartement entre les caméras) pour obtenir le repère lié à la seconde caméra. En prenant le repère R_H au milieu de cet axe, les transformations M_{HC_1} et M_{HC_2} s'écrivent:

$$M_{HC_1} = \begin{bmatrix} 1 & 0 & 0 & L/2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
$$M_{HC_2} = \begin{bmatrix} 1 & 0 & 0 & -L/2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7)$$

Les matrices de projection d'un point du repère R_H dans le repère image des cameras 1 et 2 s'écrivent respectivement $M_{proj_1} M_{HC_1}$ et $M_{proj_2} M_{HC_2}$. Les matrices de projection sont fonction des caractéristiques des caméras.

Limite de l'étude

Pour pouvoir comparer les capteurs vision monocaméra et stéréovision, les seuls obstacles considérés seront les obstacles fixes. En effet, pour des objets en mouvement, la vision monocaméra ne peut pas donner d'informations sur la position sans contrainte forte sur l'objet à détecter (par exemple un modèle, ou la vitesse de déplacement...).

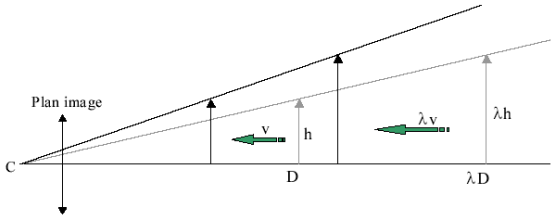


Fig. 4 : Problème de la vitesse inconnue

La figure 4 montre que des indéterminations peuvent subsister: si deux objets ont, à un instant *t*, la même projection sur le plan image de la caméra, que ces objets ont pour vitesse, par rapport au centre de camera *C*, *v* et λv et qu'ils sont éloignés de *C* respectivement de *D* et de λD, alors à un autre instant *t'* ces deux objets auront la même projection.

Les deux objets sont alors temporellement indiscernables.

Néanmoins, nous pouvons obtenir un temps à collision [1]. Ce temps T_C est défini comme étant le rapport :

$$T_c = \frac{D}{v} = \frac{\lambda D}{\lambda v} \quad (8)$$

Il ne souffre donc pas de l'indétermination du facteur λ. Un système de localisation par stéréovision n'a pas ce type de problème, car l'information est obtenue sans ambiguïté par triangulation. Pour pouvoir comparer ces deux capteurs, nous nous limiterons donc à l'étude sur obstacle fixe.

Intégration du capteur vision dans le véhicule

Cette partie sera dédiée à l'étude des transformations nécessaires pour obtenir les coordonnées d'un point dans le plan image à partir de ses coordonnées dans le

repère absolu de la scène. La transformation de projection a été étudiée dans la partie précédente.

Position des caméras dans le véhicule

La figure 5 nous montre le positionnement des caméras dans le véhicule, ainsi que les différents repères que nous avons utilisés.

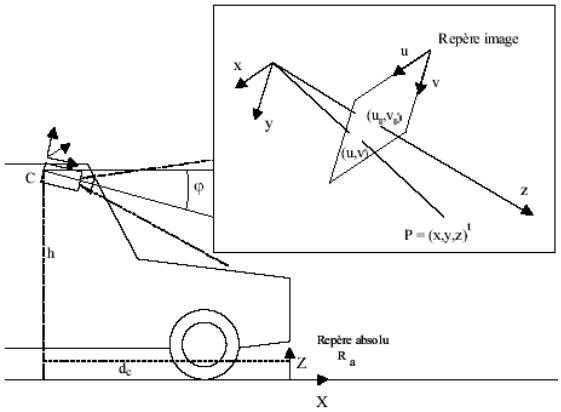


Fig. 5 : Intégration des caméras dans le véhicule. Repères utiles

Pour la suite, nous avons besoin de décrire 4 repères d'intérêt. Le premier est le repère absolu, noté R_a, dans lequel nous recherchons les coordonnées des objets observés, R_v est le repère lié au véhicule, initialement confondus avec le repère absolu. Le troisième correspond au repère caméra, noté R_c, ce repère a pour origine le centre de caméra et est orienté suivant l'axe de la caméra. Le dernier repère correspond au repère image,
noté(s) R_i. Les deux derniers repères ont été présentés dans la partie précédente. Les autres paramètres introduits sont :

- φ : l'angle entre la direction de l'axe optique des caméras et l'horizontale,
- h : la hauteur des caméras,
- d_c : la distance entre l'avant du véhicule et la projection du centre de caméra sur la route.
- et dans le cas d'un capteur stéréovision, L est l'écartement des deux caméras.

Dans la suite du document, nous adoptons la notation suivante : l'expression des coordonnées d'un point dans le repère lié au véhicule sera en majuscule, dans le repère lié à la caméra, les coordonnées seront en minuscule. Si une ambiguïté persiste, nous indiquerons en indice le repère dans lequel les coordonnées sont exprimées.

Passage du repère véhicule au repère caméra

Le passage du repère véhicule au repère caméra se fait par la composition d'une translation de vecteur $\vec{t} = d_c\vec{X} - h\vec{Z}$ et d'une rotation autour de $\vec{Y}$ d'angle φ. Dans le cas d'une perception par un système de stéréovision, nous nous sommes ramenés à ce moment là au repère R_H. Les différentes matrices de passage en coordonnées homogènes sont :

$$T_{\vec{t}} = \begin{bmatrix} 1 & 0 & 0 & d_c \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -h \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (9)$$

$$R_{\vec{Y}} = \begin{bmatrix} \cos\varphi & 0 & -\sin\varphi & 0 \\ 0 & 1 & 0 & 0 \\ \sin\varphi & 0 & \cos\varphi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (10)$$

Pour exprimer totalement les coordonnées des points dans le repère lié à la caméra, il s'agit de réaliser une permutation des indices (et ainsi obtenir comme axe de profondeur *z* et *(x,y)* un plan parallèle au plan image), ainsi l'expression de la projection sera plus simple. Soit M_{perm}, qui nous permet d'avoir les coordonnées du point à projeter dans le repère caméra.

$$M_{perm} = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (11)$$

La matrice de passage entre le repère véhicule et le repère caméra (pour le cas monocaméra), ou le repère R_H (pour la stéréovision) est $D = M_{perm} R_Y^p T_l^p$:

$$D = \begin{bmatrix} 0 & -1 & 0 & 0 \\ -\sin\varphi & 0 & -\cos\varphi & d_c \sin\varphi - h\cos\varphi \\ \cos\varphi & 0 & -\sin\varphi & d_c \cos\varphi + h\sin\varphi \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (12)$$

Dans le cas de la stéréovison, pour obtenir les coordonnées dans le repère lié à chaque caméra, il faut multiplier par les matrices de translation, décrites par l'équation (7). Finalement, les coordonnées dans le repère image sont obtenues en utilisant la matrice de projection (6). L'expression générale du passage du repère absolue au repère lié à l'image est donc:

$$M_i = M_{proj} M_{H_i} D \quad (13)$$

Dans cette expression, *i* représente la camera:

- dans le cas monocaméra, *i=mono* et M_{H_i} est la matrice identité.
- dans le cas de la stéréovision, *i* peut prendre la valeur *gauche* ou *droite* et les matrices M_{H_i} associées sont respectivement M_{HC_1} ou M_{HC_2}.

Passage du repère absolue au repère lié au véhicule

Soit maintenant Q_t, la matrice de passage entre le repère lié à la route et le repère lié au véhicule à l'instant *t*. Les coefficients de cette matrice proviennent des capteurs proprioceptifs du véhicule (accéléromètre, topomètre, GPS...). Nous obtenons la matrice de passage entre R_a et le plan image.

$$M_{i_t} = M_i Q_t = (m_{ijk}^t)_{j=1..3,k=1..4} \quad (14)$$

Soit *P* un point de coordonnée $[X,Y,Z,1]_{R_a}^T$ exprimé dans R_a, alors ses coordonnées dans R_c seront :

$$p = M_{i_t} P = [x,y,z]_{R_i}^T \quad (15)$$

Les coordonnées de *p* dans l'image seront, d'après l'équation (5) :

$$\begin{cases} u_p^t = \frac{x}{z} = \frac{m_{i11}^t X + m_{i12}^t Y + m_{i13}^t Z + m_{i14}^t}{m_{i31}^t X + m_{i32}^t Y + m_{i33}^t Z + m_{i34}^t} \\ v_p^t = \frac{y}{z} = \frac{m_{i21}^t X + m_{i22}^t Y + m_{i23}^t Z + m_{i24}^t}{m_{i31}^t X + m_{i32}^t Y + m_{i33}^t Z + m_{i34}^t} \end{cases} \quad (16)$$

Problème inverse : la rétroprojection

Connaissant les transformations qu'il faut effectuer pour passer des coordonnées d'un point dans le repère absolu au coordonnées du point dans l'image, nous allons maintenant effectuer le travail inverse, en utilisant uniquement la connaissance de la position du point dans l'image et des transformations à effectuer.

Différences entre monovision et stéréovision

Pour localiser un point fixe dans l'espace, nous avons besoin d'obtenir trois informations : ses coordonnées *X*, *Y* et *Z*.

Dans le cas d'une vision par un système monocaméra, nous obtenons, à chaque instant, une image *i* qui peut nous fournir deux informations, les coordonnées u_i et v_i du point suivi. Il nous faut donc au minimum deux images (donc des images à des instants différents), ainsi que la connaissance du déplacement de la caméra entre ces deux images. Nous parlerons de rétroprojection temporelle.

Pour un système fondé sur la stéréovision, à chaque instant nous obtenons deux

images (noté *id* et *ig*), donc 4 informations (u_{id}, v_{id}, u_{ig} et v_{ig}). Cela nous permet d'obtenir, en théorie, à chaque instant la position du point dans l'espace. C'est dans ce cas de la rétroprojection spatiale. Néanmoins, avec un système stéréovision, nous pouvons combiner ces deux méthodes pour suivre à la fois temporellement et spatialement un point.
La reconnaissance du point suivi entre les images fait appel à des techniques d'appariement qui ne seront pas développées dans cet article. Une contrainte forte commune à ces deux types de rétroprojection est la connaissance précise de la transformation entre deux images.

$2D \Rightarrow 3D$ *: la rétroprojection*

Comme le montre la figure 6, pour reconstruire un point *P* en 3 dimensions provenant d'une séquence d'images (au minimum 2), il suffit de connaître les positions C^t du centre optique ainsi que les positions des points p^t dans les différentes images. En cherchant l'intersection des droites $C^t p^t$, nous pourrions trouver *P*.

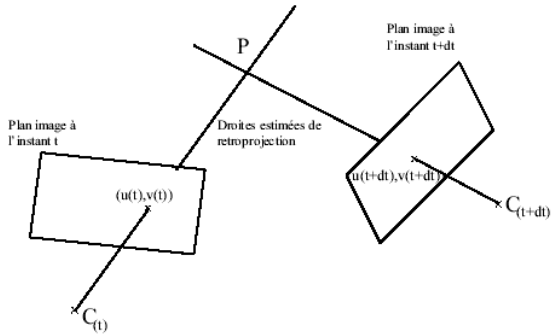


Fig. 6 : Rétroprojection théorique dans le cas monocaméra

Cependant, le pixel d'une image ayant une taille non-nulle, il existe une infinité de droites passant par ce pixel (fig. 7). En supposant une répartition équiprobable de ces droites, nous utilisons une méthode statistique pour trouver une position moyenne du point *P* ([6,7]). Ce type de méthode nous permet d'obtenir, en même temps, une estimation de l'erreur commise sur le positionnement du point dans l'espace.

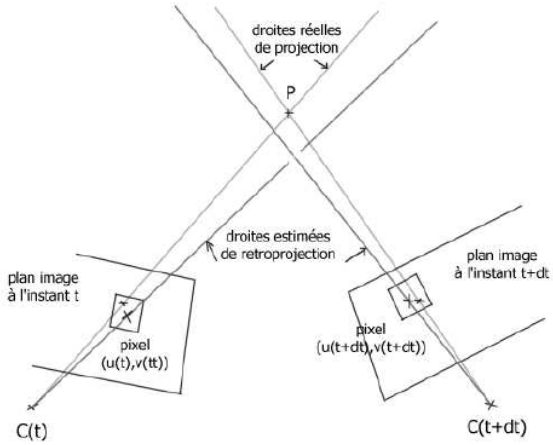


Fig. 7 : Erreur due à la taille du pixel

Estimation de la position

Soit un point *P* dont nous recherchons les coordonnées $[X,Y,Z]^T$ dans R_a. En supposant que nous connaissons ses projections $[u_p^t, v_p^t]_{t=1..n}$ dans des images différentes. Nous avons, pour chaque image (d'après l'équation (16)):

$$\begin{cases} (u_p^t m_{i31}^t - m_{i11}^t)X + (u_p^t m_{i32}^t - m_{i12}^t)Y + \\ \quad (u_p^t m_{i33}^t - m_{i13}^t)Z = m_{i14}^t - u_p^t m_{i34}^t \\ (v_p^t m_{i31}^t - m_{i21}^t)X + (u_p^t m_{i32}^t - m_{i22}^t)Y + \\ \quad (u_p^t m_{i33}^t - m_{i23}^t)Z = m_{i24}^t - v_p^t m_{i34}^t \end{cases} \quad (17)$$

Nous obtenons un système de la forme :

$$AP = b \quad (18)$$

A est de dimension *2t*3*, dans le cas monovision, et *4t*3* pour la stéréovision. Nous avons pour le cas monovision :

$$A_m = \begin{pmatrix} \vdots & & \\ u_p^t m_{31}^t - m_{11}^t & u_p^t m_{32}^t - m_{12}^t & u_p^t m_{33}^t - m_{13}^t \\ v_p^t m_{31}^t - m_{21}^t & u_p^t m_{32}^t - m_{22}^t & u_p^t m_{33}^t - m_{23}^t \\ \vdots & & \end{pmatrix} \quad (19)$$

$$b_m = \begin{pmatrix} \vdots \\ m_{14}^t - u_p^t m_{34}^t \\ m_{24}^t - v_p^t m_{34}^t \\ \vdots \end{pmatrix} \quad (20)$$

La matrice dans le cas de la stéréovision contient les informations provenant des images gauche et droite à chaque instant *t*. La résolution par la méthode des moindres carrés est possible à partir de deux images. L'équation (18) devient :

$$A^T AP = A^T b$$

Donc, si $A^T A$ est inversible (ce qui implique un mouvement non-nul entre au moins deux images dans le cas de la monovision) :

$$P = (A^T A)^{-1} A^T b \quad (21)$$

Estimation de l'erreur

L'erreur inhérente au système est due à la taille du pixel. Ce qui donne un cône tridimensionnel (fig. 7) de positions probables du point *P* à partir d'un pixel au lieu d'une droite (si cette taille est nulle). L'erreur commise à la traversée d'un pixel est de $\max(t_u/2, t_v/2)$.

La répartition des droites de rétroprojection au travers d'un pixel est équiprobable. Ce qui nous amène à faire l'hypothèse d'une erreur du type $X = \overline{X} + \varepsilon_X$. Par la suite, nous négligerons les termes en ε_X d'ordre supérieur à 1. Notre modélisation restera donc valable pour des erreurs faibles. Nous considérons une erreur de même type sur les autres paramètres. Donc $P = \overline{P} + \varepsilon_P$, $A = \overline{A} + \varepsilon_A$ et $b = \overline{b} + \varepsilon_b$. D'après l'équation (18), nous pouvons écrire :

$$\bar{A}\bar{P} = \bar{b} \quad (22)$$

Nous obtenons, en développant (18) et en négligeant les erreurs d'ordre 2 :

$$\bar{A}\varepsilon_P = \varepsilon_b + \varepsilon_A \bar{P} \quad (23)$$

Soit, en posant $D = (A^T A)^{-1} A^T$, il vient :

$$\varepsilon_P = D(\varepsilon_b + \varepsilon_A \bar{P}) \quad (23)$$

Les erreurs commises sur les axes principaux sont obtenues en prenant les éléments de la diagonale de la matrice de covariance sur *P*, notée Cov_P :

$$Cov_P = \varepsilon_P \varepsilon_P^T = D(\varepsilon_b + \varepsilon_A \bar{P})(\varepsilon_b + \varepsilon_A \bar{P})^T D^T \quad (24)$$

La méthode exposée ci-dessus est simple, néanmoins, elle permet d'estimer à la fois la position d'une cible et l'erreur commise lors de cette estimation. Comme nous le verrons par la suite, pour améliorer la précision de l'estimation de la position, il suffit d'augmenter le nombre d'images utilisées. Ceci ayant un impact faible puisque la méthode consiste à multiplier des matrice, opération relativement rapide, et à inverser la matrice $A^T A$ qui est une matrice *3*3*. Cette méthode peut être amélioré en considérant uniquement l'apport d'une nouvelle information (une image en monocaméra ou deux en stéréovision) sur la localisation d'un point et sur la précision de cette locatisation. [8] montre que ce type de considération amène à un moindre carré itératif, demandant moins de calcul. Finalement cette méthode peut être rapidement transformé en un filtre de kalman.

Comparaison du capteur mono et stéréo

Nous allons, dans cette partie, étudier les performances du système de rétroprojection vis-à-vis des variations de certains paramètres. Finalement, nous présenterons quelques résultats expérimentaux utilisant la retroprojection.

Dans le cas général, nous prendrons un point fixe à rétroprojeter se situant à une altitude de *0,5m*, et ayant un écart latéral de *1,5m*. Les paramètres relatifs à la caméra et à son positionnement sont :

$t_u = t_v = 8{,}3e^{-6}\ m$	Taille des pixels
$f = 8{,}5\ mm$	Focale de la caméra
$h = 1\ m$	Hauteur de la caméra
$d_c = 1\ m$	Distance à l'avant du véhicule
$l = 1\ m$	Espacement entre les deux caméras
$\alpha = 5{,}4°$	Inclinaison des caméras
$v = 14\ m/s$	Vitesse du véhicule
$n = 10$	Nombre d'images utilisées
$dt = 1/25\ s$	Durée entre 2 images

La configuration utilisée ici est celle du LIVIC. Ainsi, pour étendre notre étude à d'autres cas, nous étudierons les effets des variations de ces paramètres. Pour pouvoir percevoir le plus loin possible, nous avons considéré la distance maximale qu'il est possible d'obtenir pour un tel dispositif lorsqu'il est placé derrière le pare-brise d'un véhicule.

Approximation de l'erreur

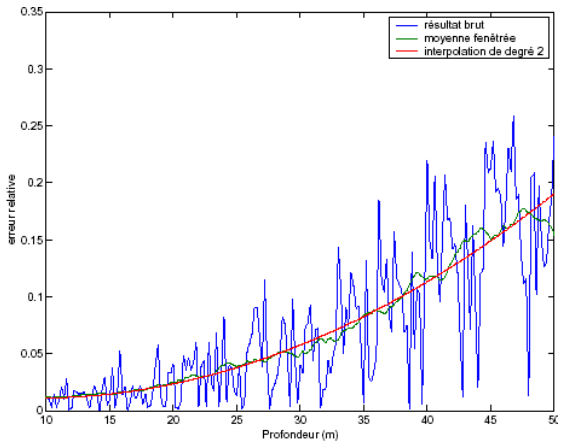


Fig. 8 : Comparaison entre le résultat brut, une moyenne glissante et l'interpolation de degré 2, pour une séquence de 10 images

Dans la figure 8, nous avons caractérisé l'erreur relative de l'estimation de la profondeur de la cible dans la scène, en fonction de cette profondeur (distance à l'obstacle). L'erreur est ici uniquement due à la taille du pixel sur la matrice CCD. Nous pouvons dégager un comportement global de l'évolution de l'erreur. Cette figure présente la moyenne glissante ainsi que l'interpolation polynomiale de degré 2. Ces courbes suivent bien le comportement de l'erreur. Par la suite, nous associerons la courbe d'erreur à son interpolation polynomiale de degré 2.

Variation du nombre d'images

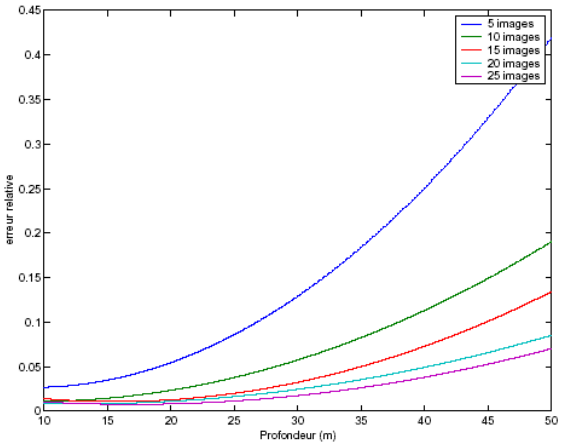


Fig. 9 : Variation de l'erreur sur la profondeur en monocaméra pour 5,10,15,20 et 25 images

Comme représentée sur la figure 9, l'erreur sur l'estimation de la profondeur diminue avec le nombre d'images, mais, compte-tenu de l'application détection d'obstacles, il n'est pas possible d'attendre 25 images pour détecter précisément un obstacle. Pour une rétroprojection utilisant 10 ou 15 images, le rapport entre précision et rapidité de détection est convenable (l'erreur est inférieure à *10%* à *25m* en utilisant 10 images, elle reste inférieure à ce seuil à *35m* en utilisant 15 images). Pour les courbes utilisant 15, 20 et 25 images, nous observons un minimum de l'erreur entre *15* et *20m* de profondeur (fig. 9). Or, compte-tenu de la vitesse du véhicule et des dimensions de la matrice CCD, le point physique ne peut pas être visible dans toutes les images de la séquence. L'appariement n'est donc pas possible. La partie inférieur à ce minimum n'ayant pas de signification physique, il faut en fait considérer uniquement la partie supérieure à ce minimum (minimum particulièrement visible sur la figure 10).

Dans le cadre d'une détection par un système de stéréovision, on ne peut pas dégager de comportement global des courbes, du fait de la faible erreur. La figure 9 montre le comportement moyen de l'erreur de rétroprojection. Il faut rapprocher ces courbes de la figure 8, qui montre les oscillations de l'erreur autour de sa position moyenne. Néanmoins, l'erreur maximale reste inférieure à *5%* à *90m*.

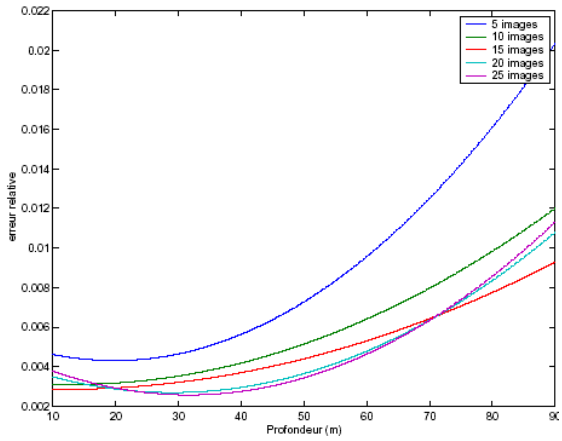


Fig. 10 : Variation de l'erreur sur la profondeur en stéréovision pour 5,10,15,20 et 25 images

Ainsi, l'erreur de rétroprojection est beaucoup plus faible dans le cas stéréo que dans le cas monocaméra pour deux raisons. D'une part, un système stéréo a deux fois plus d'informations qu'un système monocaméra, pour un même nombre d'échantillons. Néanmoins, cette raison n'est pas suffisante. En effet, sur la figure 10, l'erreur à *50m* pour 5 images est de l'ordre de *0,007%*, contre presque *0,2%* pour 10 images à *50m* dans le cas monocaméra (or nous avons le même nombre d'informations que dans le cas stéréo). D'autre part, dans le cadre de la détection d'obstacles routiers, généralement le véhicule se déplace en direction de ces obstacles.

Or, comme le montre la figure 11, dans le cas de la vision monocaméra, la zone des positions probables du point cible est bien plus grande que dans le cas de la stéréovision. Ceci est dû au déplacement des plans images qui n'est que longitudinal pour la monovision (contre un déplacement latéral et longitudinal pour le plan image dans le cas de la stéréovision).

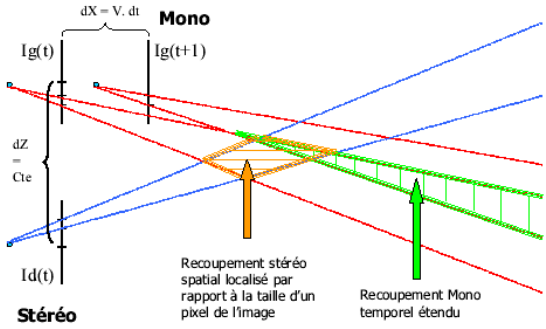


Fig. 11 : Zone des positions probables d'un point P, dans le cas d'une rétroprojection spatiale et temporelle

La figure 12 montre la précision d'une rétroprojection spatiale par rapport à une rétroprojection spatiale et temporelle. Il faut remarquer que cette figure nous présente le comportement moyen de l'erreur. Néanmoins, l'erreur maximale de détection est inférieure à *8%* (sans suivi temporel).

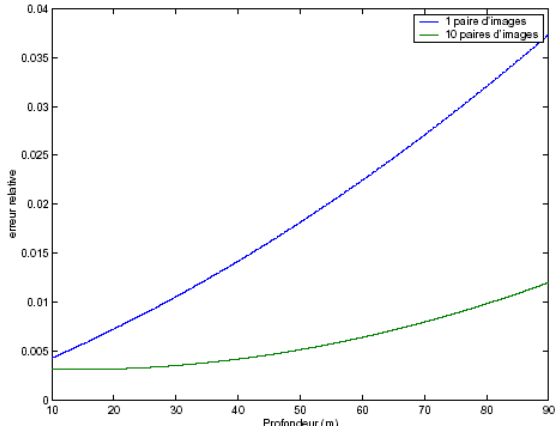


Fig. 12 : Variation de l'erreur sur la profondeur pour 1 et 10 images stéréo

L'utilisation seule de la rétroprojection spatiale donne des résultats satisfaisants.

Dépendance à la position du point

Pour l'instant, nous avons présenté des résultats en fixant une position pour le point visé. Les figures 13 et 14, montrent une cartographie de l'erreur relative sur l'estimation de la position du point à rétroprojeter en fonction de sa position (l'altitude reste fixe et est égale à *0,5m*). La figure 13 présente deux zones distinctes. La première, où l'erreur est supérieure à *10%*, correspond à l'axe optique de la caméra et à son voisinage proche (une dizaine de pixels autour du centre optique). La seconde zone (erreur inférieure à *10%*) correspond à des points projetés qui auront, dans l'image, un déplacement plus important au travers de la séquence. Nous pouvons retrouver ces zones, mais plus atténuées, dans la figure 14. Les zones de fortes erreurs sont sur et autour de l'axe optique de chaque caméra.

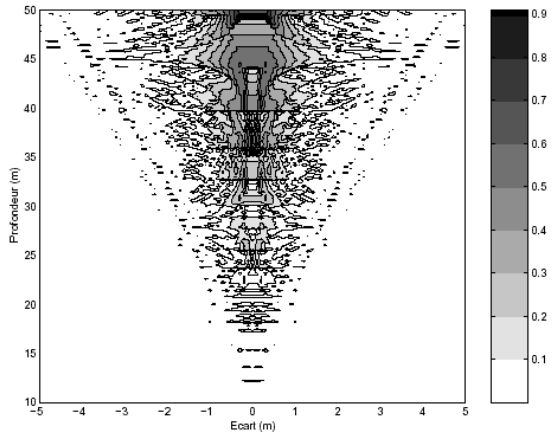


Fig. 13 : Variation de l'erreur sur la profondeur pour 10 images mono

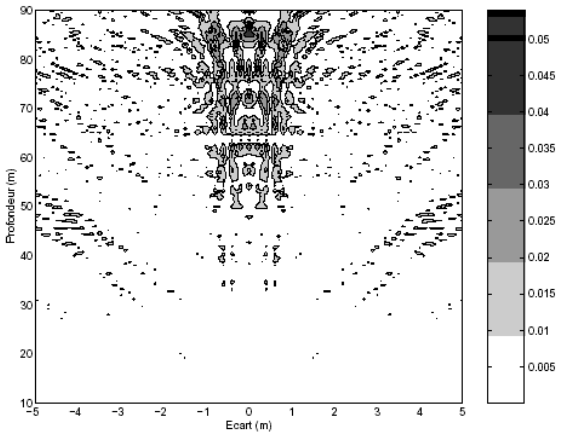


Fig. 14 : Variation de l'erreur sur la profondeur pour 10 images stéréo

Ces résultats sont valides dans le cadre d'une détection parfaite. Or, les différents détecteurs de points particuliers ([2,3,4]) sont généralement précis au pixel près. Cette erreur est introduite dans la section suivante.

Prise en compte de l'erreur de détection

Pour simuler cette erreur de détection, nous introduisons un bruit blanc centré. Ainsi, chacun des 8 pixels entourant le point de projection réel peut être détecté comme point particulier.

Dans le cas monoculaire, l'erreur moyenne relative en fonction de la profondeur est multipliée par un facteur 10 (fig 15). Néanmoins, dans le cadre d'une application de détection d'obstacles à faible distance, inférieure à *20m*, l'erreur maximale est de l'ordre de *10%*. Les résultats restent exploitables.

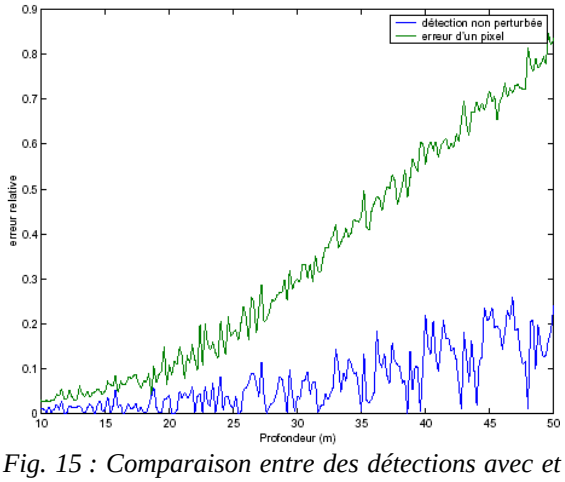


Fig. 15 : Comparaison entre des détections avec et sans erreur, cas mono

Dans le cadre d'une application en stéréovision, une erreur d'un pixel lors de la détection multiplie l'erreur relative moyenne en fonction de la profondeur par un facteur 6 (fig. 16). Mais elle reste inférieure à *7%* à *90m* (avec un suivi temporel). La robustesse d'un système stéréo vis-à-vis de l'erreur de détection est donc bien plus grande que celle d'un système monocaméra. En effet, nous

prenons dans le cas stereo plus d'informations que pour le cas monocaméra. Or le bruit est centré, donc la moyenne des mesures est plus proche de la valeur réelle.

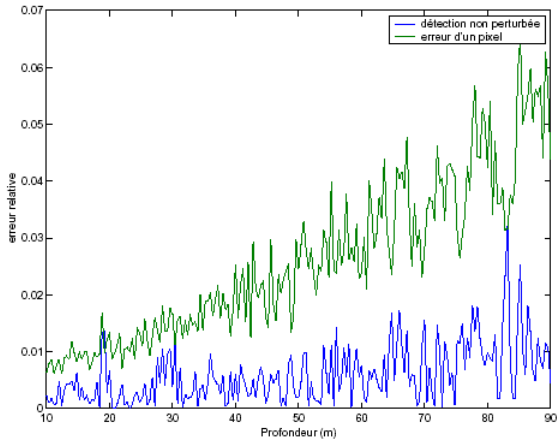


Fig. 16 : Comparaison entre des détections avec et sans erreur, cas stéréo

Variation de la précision de reconstruction en fonction des paramètres intrinsèques

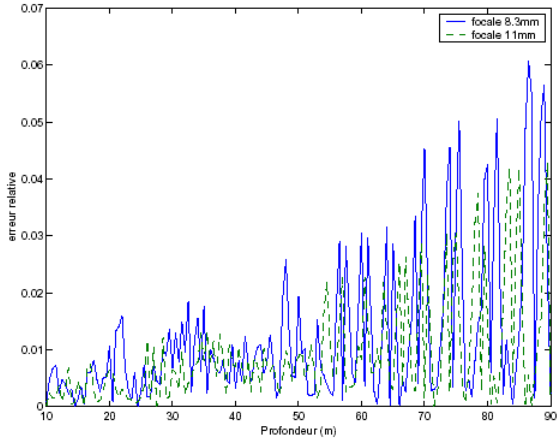


Fig. 17 : Variation de la focale (cas stéréo)

D'autres paramètres interviennent sur la précision de la reconstruction. Le premier est celui de la vitesse du véhicule : plus elle est grande, plus la reconstruction est précise. Cela provient du fait que le déplacement entre deux images de l'objet à rétroprojeter est plus grande, donc l'intersection des cônes de rétroprojection est plus petite. Par contre, la focale intervient peu sur la précision. En la faisant varier de 7mm à 11mm, les erreurs relatives restent les mêmes (voir fig. 17).

Par contre, la taille du pixel est très importante. En divisant par deux les dimensions du pixel, l'erreur moyenne relative en fonction de la profondeur est divisée par un facteur 3 (voir fig. 17). D'où l'intérêt des détecteurs de points subpixeliques ([5]) et/ou des capteurs vidéo de grandes résolutions.

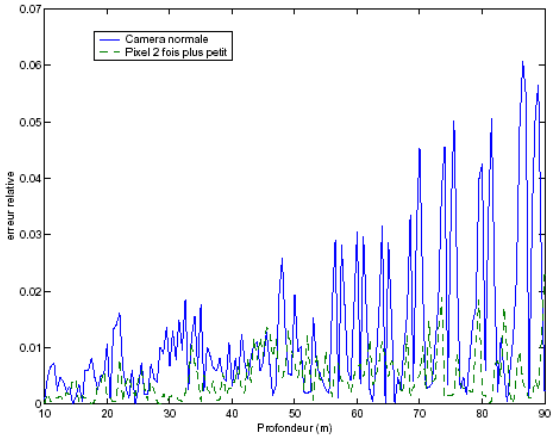
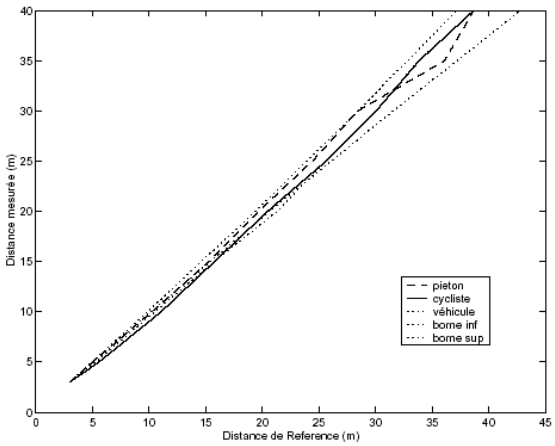


Fig. 18 : Importance de la taille du pixel (cas stéréo)

Confrontation aux résultats réels

Le LIVIC développe des systèmes d'aides à la conduite. Pour la localisation des obstacles, c'est une nouvelle technique de stéréovision qui a été mise en oeuvre [9].



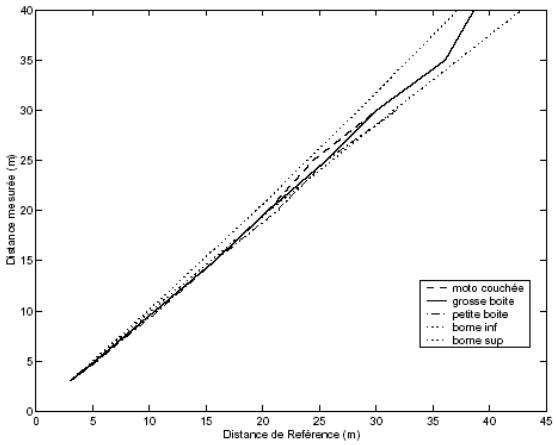


Fig. 19 : Précision de l'estimation de la distance d'un objet, pour différentes cibles et différentes distances [10]

Dans la figure 19, la distance calculée par le système est comparée à la distance réelle pour différents objets et différentes distances. Le cône d'incertitude est présenté en pointillé sur ces figures. Le calcul est effectué sur une paire d'image, la rétroprojection est ici uniquement spatiale. De plus la définition de l'image n'est que du *1/4* de PAL pour assurer le calcul en temps réel de la distance. Ainsi, la précision est, dans le pire des cas, de *0,7%* à *3m* et de *14%* à *40m*. Ces calculs de précisions sont à rapprocher des figures 18 et 12. En effet, pour nos simulations, le format utilisé est le format PAL complet, donc la précision est ici 3 fois plus faible. D'après les simulations, la précision à *40m* est en moyenne de *1,5%* (fig. 12), en prenant le cas le plus défavorable, nous arrivons à *3%* (fig. 8). Compte tenu de la résolution choisie, l'erreur devrait être de l'ordre *9%*, si la détection était sans erreur. La figure 16 nous montre qu'une détection avec une erreur de 1 pixel, multiplie l'erreur de reconstruction par un facteur variant entre 2 et 3. En prenant en compte ce paramètre, l'erreur commise à la reconstruction devrait varier entre *18%* et *27%*. La méthode de reconstruction est donc très précise.

Conclusion

Pour réaliser une assistance à la conduite de type "sécurité prédictive", il est nécessaire de spécifier son domaine de fonctionnement et les informations dont elle a besoin. Ce cahier des charges étant posé, il convient alors de choisir le ou les capteurs extéroceptifs adéquat(s). Pour cela, il est indispensable de connaître le domaine de fonctionnement et les précisions à attendre de chaque capteur. Si les performances en terme de précision des capteurs télémétriques sont bien identifiées, celles du capteur vision ne sont que partiellement décrites dans la littérature.

Pour pallier cette méconnaissance, nous avons mené une étude qui a permis de quantifier la précision de positionnement de l'objet détecté. Nous avons mis en parallèle les performances théoriques atteignables en mono-vision et en stéréovision.

Comme attendu les mesures fournies par la stéréovision sont beaucoup plus précises que celles obtenues par mono-vision. A titre de comparaison, l'objet sera localisé avec une précision de *10%* à *90m* en stéréovision et suivi sur 10 images, alors qu'il faut descendre à *20m* en mono-vision pour conserver la même qualité. La stéréovision présente aussi un autre avantage, la connaissance de la profondeur par triangulation. Ainsi, une seule paire d'images permet de localiser un objet avec cette même précision jusqu'à *35m*.

Compte tenu des niveaux de fiabilité exigés dans le domaine de la sécurité de la conduite, il semble nécessaire de fusionner les informations issues de différents types de capteur. C'est une des voies entreprises au LIVIC qui exploite un capteur stéréoscopique et la télémétrie LASER. Or dans le cadre des techniques de fusion la précision de l'information est une donnée nécessaire au fonctionnement des

algorithmes. Notre étude trouve donc aussi toute sa signification dans ce contexte.

Références

[1] J. L. Barron, S. S. Beauchemin, and D. J. Fleet. *On optical flow*. 6th Int. Conf. on Artificial Intelligence and Information- Control Systems of Robots (AIICSR), pages 3-14, 1994. Sept. 12-16, 1994, Smolenice Castle, Slovakia.

[2] Achard-Rouquet C., Bigorgne E., and Devars J. *Un détecteur de points caractéristiques sur les images multispectrales. extension vers un détecteur sub-pixellique*. ICPR, 2000.

[3] Harris C. and Stephens M. *A combined corner and edge detector*. Proceedings of the 4th Avley Vision Conference, pages 147-151, 1988.

[4] Schmid C., Mohr R., and Bauckage C. *Comparing and evaluating interest points*. ICCV, pages 230-235, 1998.

[5] Devernay F. *A non-maxima suppression method for edge detection with sub-pixel accuracy*. Rapport technique, INRIA, 1995.

[6] Ramparani F. *Perception multisensorielle de la structure géométrique d'une scène*. Thèse, Institut National Polytechnique de Grenoble - spécialité : informatique, 1984.

[7] Olague G. *Planification du placement de caméras pour des mesures 3D de précision*. Thèse, Institut National Polytechnique de Grenoble - Laboratoire GRAVIR-IMAG-INRIA, spécialité : Imagerie, vision, robotique, 1998.

[8] Legras J. *Algorithmes et programmes d'optimisation non lineaire avec contraintes*. Masson, 1980.

[9] Labayrade R. and Aubert D. *Robust and fast stereovision based road obstacles detection for driving safety assistance*. Machine Vision and Application 2002 (MVA 2002), Nara, Japan, 11-13 December 2002.

[10] Labayrade R., Aubert D., and J. P. Tarel. *Real time obstacle detection on non flat road geometry through v-disparity representation*. IEEE Intelligent Vehicle Symposium, Versailles, June 2002.

[11] C. Thomas, A. Damville, T. Perron, C. Mautuit, and J.Y. Le Coz. *Comportement humain en accidents corporels*. In 2ème Journée Sécurité Automobile, Rouen, 2000.

[12] C. Thomas, T. Perron, J.Y. Le Coz, and V. Aguade. *What happens on the road before fatal car crashes?* In?,?