

Gestion intelligente de capteurs et fusion multisensorielle pour la détection et le suivi d'obstacles sur route

Smart Sensing and Multisensorial Data Fusion for Road Obstacle Detection and Tracking

par Laurent TRASSOUDAIN, Paul CHECCHIN, Joseph ALIZON, François COLLANGE et Jean GALLICE

Laboratoire des Sciences et Matériaux pour l'Electronique et d'Automatique (LASMEA), U.R.A. 1793 du C.N.R.S.,
Université Blaise Pascal de ClermontFerrand, F-63177 Aubière cedex, France e-mail trassou@lasmea.univ-bpclermont.fr

résumé et mots clés

Dans cet article, nous présentons une solution multisensorielle temps réel pour la détection et le suivi d'obstacles sur route. Cette solution est basée sur l'utilisation d'un capteur mixte caméra vidéo/capteur de profondeur placé à l'avant d'un véhicule expérimental.

Le capteur multisensoriel est décrit. Le calibrage permet l'alignement des données hétérogènes. Deux facultés du capteur sont développées : la perception dirigée permet l'acquisition d'une image de profondeur dans une zone définie dans l'image de luminance; l'asservissement visuel réalise la focalisation du faisceau laser sur un point de l'image de luminance. De façon générale, ces facultés permettent un contrôle par rétroaction sur le mode d'acquisition du capteur en fonction de la situation dans laquelle se trouve le système de perception.

La stratégie de perception est basée sur la sélection du capteur adéquat pour un objectif donné. La détection d'obstacle repose sur la segmentation et l'interprétation des données de profondeur qui sont d'une grande pertinence dans ce contexte. En revanche, la cadence d'acquisition de ces données n'est pas suffisante si l'on souhaite dériver les caractéristiques cinématiques des obstacles. En conséquence, le suivi des obstacles combine un traitement de l'image de luminance rapide avec un traitement de l'information 3D. Le premier permet de réactualiser la position de l'obstacle afin d'asservir le faisceau laser sur celui-ci et le second assure la connaissance de la taille du modèle de l'obstacle à chercher dans l'image. Cet algorithme de fusion de données hétérogènes accompagné d'un filtrage de Kalman permet d'inférer les caractéristiques cinématiques des obstacles dont la connaissance est indispensable pour aborder ceux-ci dans de bonnes conditions.

Ces recherches sont menées dans le cadre du projet européen PROMETHEUS et sont validées en situation réelle à bord du véhicule expérimental Prolab.

Perception de l'environnement, Fusion de données, Capteur actif, Capteur intelligent, Détection d'obstacles.

abstract and key words

In this article, we present a multisensorial solution for road obstacle detection and tracking. This solution is based on a mixed camera/3D sensor mounted on the front of an experimental vehicle.

The multisensor is described. The calibration step enables the matching of the heterogeneous data. Two capabilities of the sensor have been developed : the controlled perception making possible the acquisition of depth data in an area defined in the intensity image; the visual servoing carrying out the focusing of the laser beam on a moving target detected in the intensity image. These two capabilities allow a feedback control on the acquisition mode of the sensor according to the environment.

The perception strategy is based on the selection of the best sensor for a given goal. The obstacle detection is based on the segmentation and interpretation of depth data which are well suited in this context. However, the rate of acquisition of these data is too slow in order to extract the kinematic state of the obstacle. So, the tracking process is based on the collaboration between intensity image processing which ensures the tracking itself and a 3D process which returns the obstacle model size to search in the image. This algorithm of heterogeneous data fusion, associated with a Kalman filtering, permits to compute the state of obstacles. This work fits into the european project PROMETHEUS. Experimental results have been validated in real situation on the Prolab vehicle.

Perception, Data fusion, Active vision, Smart sensing, Obstacle detection.

1. introduction

Les problèmes complexes posés par la perception et l'interprétation de scènes d'environnements naturels constituent un point de focalisation des activités de la communauté « Vision par Ordinateur »; et plus particulièrement dans le domaine applicatif des véhicules autonomes pour lequel la perception extéroceptive de l'environnement représente le point le plus difficile à résoudre. C'est le cas du projet européen Eurêka-Prometheus, dont la problématique générale est l'assistance à la conduite sur route (copilote électronique).

Dans le cadre de ce projet, cet article présente les principaux résultats que nous avons obtenus à partir d'une stratégie de perception multisensorielle pour la détection et le suivi d'obstacles. Nous montrons que l'utilisation conjointe de données hétérogènes accompagnée d'une commande adéquate du capteur permet de résoudre le délicat problème de perception par une méthode robuste. De plus nous pouvons calculer l'état cinématique de l'obstacle avec une précision acceptable. Nous attachons une importance particulière à ce dernier point, en effet, il apparaît plus délicat de déterminer l'état cinématique des obstacles par des méthodes basées sur la vision seule. C'est le cas des travaux de Dickmanns [TDD94] qui font référence en matière de perception dans les véhicules routiers intelligents.

Au niveau de la perception de l'environnement, un système de fusion de données complet mis en œuvre dans un véhicule expérimental est décrit dans [RM95]. A un plus haut niveau dans la chaîne de perception, [DER94] et [H95] décrivent des méthodes de fusion de cartes partielles de l'environnement en vue d'obtenir une carte globale des obstacles autour du véhicule expérimental. Concernant l'application détection d'obstacles, cet article propose une application de la fusion de données telle qu'elle est définie dans [R94]. Sans apporter de théorie nouvelle, nous présentons comment nous avons abordé les quatre fonctions d'un système de fusion afin de mélanger les données issues de capteurs différents pour prendre une décision fiable. Ces fonctions sont : le capteur, le recalage des données, la fusion et le contrôle adaptatif de ces trois dernières fonctions.

Le capteur, maillon essentiel de la chaîne de perception, doit être choisi de façon optimale pour l'application visée. Même s'il n'intervient pas directement dans la combinaison des informations, il est susceptible d'être piloté par un module de supervision et ce afin de fournir les données les plus pertinentes en fonction du contexte. Ainsi, les notions de fenêtre d'intérêt, de mode veille et mode actif sont utilisées.

Avant de combiner les informations, il est nécessaire de les ramener dans un référentiel commun, c'est le rôle du recalage. Dans notre cas, il s'agit d'un recalage géométrique qui doit être pris comme une étape préliminaire à la fusion.

La combinaison d'informations fait ici appel à une heuristique dédiée à l'application où deux processus coopèrent afin de tirer le meilleur parti des données disponibles. Un filtre de Kalman prend

en compte l'aspect dynamique du processus.

Le contrôle adaptatif n'apparaît pas formellement dans l'article mais on verra comment le système gère des données autorisant une décision « précoce » permettant ensuite de le reconfigurer pour prendre une décision plus fiable.

Après avoir décrit le capteur multisensoriel constitué d'un capteur d'image de luminance et d'un capteur de profondeur, nous présentons la procédure de calibrage assurant la mise en correspondance d'un point de l'espace 3D avec le point associé dans l'image 2D.

Dans la seconde partie nous décrivons les deux modes de contrôle du capteur, perception dirigée et asservissement visuel, qui confèrent une certaine forme d'intelligence au capteur.

La troisième partie présente les applications de détection et de suivi d'obstacles mise en œuvre en temps réel dans le véhicule expérimental Prolab. Elles permettent le calcul des paramètres cinématiques du véhicule suivi.

2. le capteur multisensoriel

2.1. description

Une caméra CCD standard rigidement couplée à un capteur 3D forme le capteur multisensoriel. Le capteur 3D est construit sur la base d'un télémètre laser à temps de vol dont le faisceau est dirigé par deux miroirs plans.

La complémentarité entre les données hétérogènes issues de ce capteur est un aspect important car elle permet d'allier la rapidité d'acquisition d'une image vidéo et la précision de localisation des données de profondeur.

2.2. calibrage du capteur multisensoriel

Le calibrage est l'étape préalable indispensable à la fusion de données. Il permet l'alignement de l'ensemble des données dans le même référentiel.

Dans notre cas précis, le but du calibrage du capteur multisensoriel est de mettre en correspondance, c'est-à-dire recalcr, une image de distance, ou image 3D, et une image de luminance. Ainsi, à la résolution près, pour chaque point de l'espace, dans le champ de vision du capteur, nous connaissons la luminance ainsi que la position 3D.

Dans un premier temps, il faut calibrer le capteur 3D. Il s'agit d'accéder aux coordonnées cartésiennes d'un point de l'espace à partir des données issues du capteur 3D. De plus, ces coordonnées doivent être exprimées dans un repère quelconque, car le repère lié au capteur n'est pas, a priori, le repère adéquat. Le calibrage

de la caméra intervient dans un second temps. Il permet de lier la position d'un point dans l'espace et la position de sa projection dans l'image de luminance.

Nous utiliserons les repères suivants :

- $\mathcal{R}_3(O_3, x_3, y_3, z_3)$ est le repère 3D lié au capteur 3D.
- $\mathcal{R}_a(O_a, x_a, y_a, z_a)$ est le repère 3D absolu.
- $\mathcal{R}_c(O, u, v)$ est le repère 2D discret lié à la caméra vidéo.

Les matrices de passages sont représentées figure 1. M_{3a} est la matrice de changement de repère entre \mathcal{R}_3 et \mathcal{R}_a . M_{ac} permet le passage entre \mathcal{R}_a et \mathcal{R}_c . M_{3c} est la combinaison de ces deux changements de repère. Elle permet, le cas échéant, de s'affranchir du repère absolu.

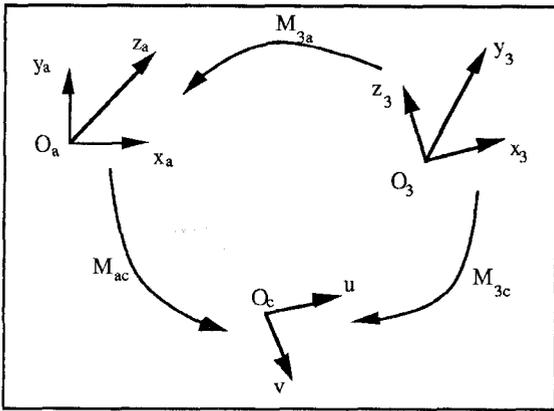


Figure 1. – Les matrices de changement de repères.

2.2.1. calibrage du capteur 3D

Nous proposons ici une méthode générale pour calibrer un capteur 3D constitué par un télémètre laser. Calibrer un tel capteur revient à être capable de déterminer les coordonnées cartésiennes (x_a, y_a, z_a) d'un point de l'espace dans \mathcal{R}_a à partir des données $Y = (y_1, \dots, y_d)$ fournies par le capteur. La transformation fait intervenir la cascade de deux transformations :

- le passage des mesures aux coordonnées cartésiennes du point dans \mathcal{R}_3 ,
- le changement de repère entre \mathcal{R}_3 et \mathcal{R}_a .

La première transformation englobe le n -uplet A de paramètres non linéaires et le n -uplet B de paramètres linéaires. M_{3a} réalise la seconde transformation. On exprime, tout d'abord, les coordonnées cartésiennes (x_3, y_3, z_3) de ce point dans le repère du capteur \mathcal{R}_3 en fonction des paramètres linéaires $B = (b_1, \dots, b_m)$ et non linéaires $A = (a_1, \dots, a_n)$ du capteur et des données Y par

les relations suivantes :

$$\begin{pmatrix} x_3 = \sum_{i=1}^r b_i \cdot f_i(A, Y) \\ y_3 = \sum_{i=r+1}^{r+s} b_i \cdot f_i(A, Y) \\ z_3 = \sum_{i=r+s+1}^m b_i \cdot f_i(A, Y) \end{pmatrix} \quad (1)$$

où les $f_i(A, Y)$ sont des fonctions non linéaires. m fonctions linéaires permettent le calcul des coordonnées cartésiennes (x_3, y_3, z_3) dont r pour x_3 et s pour y_3 .

Le système (1) peut être mis sous une forme matricielle :

$$X_3 = D \cdot F(A, Y) \quad (2)$$

avec

$$X_3 = (x_3 \quad y_3 \quad z_3)^t,$$

$$D = \begin{pmatrix} b_1 \dots b_r & 0 & 0 \\ 0 & b_{r+1} \dots b_{r+s} & 0 \\ 0 & 0 & b_{r+s+1} \dots b_m \end{pmatrix},$$

et

$$F(A, Y) = (f_1(A, Y) \quad \dots \quad f_m(A, Y))^t.$$

La relation (2) exprime donc les coordonnées cartésiennes d'un point de l'espace dans le repère du capteur \mathcal{R}_3 en fonction des données fournies par le capteur et des paramètres intrinsèques du capteur.

Il s'agit maintenant d'effectuer le changement de repère entre \mathcal{R}_3 et \mathcal{R}_a . Une telle transformation est réalisée, classiquement, par trois rotations $R_{\alpha\beta\gamma}$ et trois translations T_{xyz} :

$$X_a = R_{\alpha\beta\gamma} \cdot X_3 + T_{xyz}$$

ou simplement en utilisant les coordonnées homogènes

$$X_a = M_{3a} \cdot \bar{X}_3 \quad (3)$$

avec

$$\bar{X}_3 = (X_3^t \quad 1)^t$$

et

$$M_{3a} = (R_{\alpha\beta\gamma} \quad T_{xyz})$$

Le changement de repère pourra donc être effectué en homogénéisant la relation (2) de la façon suivante :

$$\bar{X}_3 = \bar{D} \cdot \bar{F}(A, Y) \quad (4)$$

avec

$$\bar{D} = \begin{pmatrix} D & 0 \\ 0 & 1 \end{pmatrix}$$

et

$$\bar{F}(A, Y) = (F^t(A, Y) \quad 1)^t$$

Grâce aux relations (3) et (4), on obtient finalement :

$$X_a = M_{3a} \cdot \bar{D} \cdot \bar{F}(A, Y) \quad (5)$$

Cette équation exprime la position d'un point de l'espace dans le repère absolu \mathcal{R}_a en fonction des paramètres du capteur et des mesures qu'il effectue. Il faut donc estimer les coefficients du produit de matrice $M_{3a} \cdot \bar{D}$ puis les paramètres non linéaires. Les premiers le sont par la méthode des moindres carrés, les seconds par la méthode itérative de Newton. L'ensemble des paramètres, ainsi déterminé une fois pour toute, permet donc de calculer les coordonnées cartésiennes (x_a, y_a, z_a) à partir des mesures Y issues du capteur.

Dans le cas particulier de notre capteur 3D, les deux organes principaux sont un télémètre laser et une tête de déflexion à deux miroirs plans. Les données issues du capteur sont trois tensions $Y = (V_\rho, V_{\theta_x}, V_{\theta_y})$ respectivement proportionnelles à la distance ρ mesurée par le télémètre et aux déviations (θ_x, θ_y) des miroirs. Les relations (6) relient les grandeurs électriques et les grandeurs géométriques.

$$\begin{pmatrix} \rho = K_\rho \cdot V_\rho \\ \theta_x = K_{\theta_x} \cdot V_{\theta_x} \\ \theta_y = K_{\theta_y} \cdot V_{\theta_y} \end{pmatrix} \quad (6)$$

où les K_i sont les paramètres intrinsèques électriques du capteur.

La figure 2 schématise la géométrie du capteur 3D à ceci près que l'origine du repère est en fait au centre du miroir Y. Les paramètres intrinsèques géométriques du capteur 3D sont la distance e qui sépare les deux miroirs et la distance d entre la référence du télémètre et la position du faisceau sur le miroir X. Le faisceau laser, émis à une distance d du miroir X, vient frapper ce miroir en a . Il est dévié, avec un angle θ_x , en direction du miroir Y qu'il atteint en b . Le miroir Y dévie à son tour le faisceau avec un angle θ_y avant que le faisceau atteigne le point $X_3(x_3, y_3, z_3)$ de l'espace. Compte tenu de cette géométrie, la relation (7) permet le passage entre les coordonnées pseudo-sphériques $(\rho, \theta_x, \theta_y)$ et les coordonnées cartésiennes (x_3, y_3, z_3) dans le repère du capteur \mathcal{R}_3 :

$$\begin{pmatrix} x_3 = (\rho - d) \cdot \sin \theta_x \\ y_3 = (\rho - d) \cdot \cos \theta_x \cdot \sin \theta_y - e \cdot \sin \theta_y \\ z_3 = (\rho - d) \cdot \cos \theta_x \cdot \cos \theta_y - e \cdot \cos \theta_y \end{pmatrix} \quad (7)$$

Compte tenu des relations (6) et (7), on effectue l'identification du modèle en posant :

$$F(A, Y) = \begin{pmatrix} V_\rho \cdot \sin(K_{\theta_x} \cdot V_{\theta_x}) \\ \sin(K_{\theta_x} \cdot V_{\theta_x}) \\ V_\rho \cdot \cos(K_{\theta_x} \cdot V_{\theta_x}) \cdot \sin(K_{\theta_y} \cdot V_{\theta_y}) \\ \cos(K_{\theta_x} \cdot V_{\theta_x}) \cdot \sin(K_{\theta_y} \cdot V_{\theta_y}) \\ \sin(K_{\theta_y} \cdot V_{\theta_y}) \\ V_\rho \cdot \cos(K_{\theta_x} \cdot V_{\theta_x}) \cdot \cos(K_{\theta_y} \cdot V_{\theta_y}) \\ \cos(K_{\theta_x} \cdot V_{\theta_x}) \cdot \cos(K_{\theta_y} \cdot V_{\theta_y}) \\ \cos(K_{\theta_y} \cdot V_{\theta_y}) \end{pmatrix}$$

avec

- $A = (K_{\theta_x}, K_{\theta_y})$
- $B = (K_\rho, -d, K_\rho, -d, -e, K_\rho, -d, -e)$
- $r = 2; s = 3; m = 8$

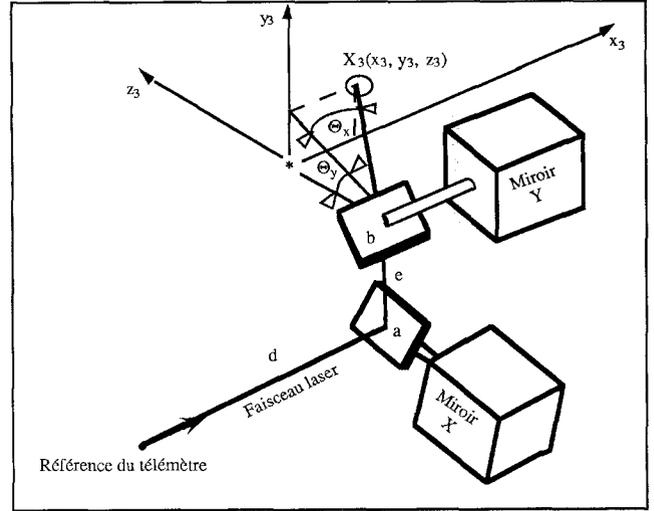


Figure 2. - La géométrie du capteur 3D.

2.2.2. calibration de la caméra

Le calibrage de la caméra est un problème bien connu en vision. Ce problème peut se poser en ces termes : étant donné un point $X_a(x_a, y_a, z_a)$ dans un repère absolu \mathcal{R}_a , quelle transformation permet de connaître la position discrète (u, v) de sa projection dans le repère \mathcal{R}_c de l'image? Nous avons utilisé la méthode décrite par Toscani [FT87]. Elle utilise un modèle sténopé de la caméra (voir figure 3) et décompose la transformation en trois phases : le changement de repère, la projection perspective et la numérisation.

Changement de repère : la première opération consiste à exprimer les données dans le repère lié à la caméra. La transformation qui permet de passer des coordonnées (x_a, y_a, z_a) d'un point exprimées dans le repère absolu \mathcal{R}_a aux coordonnées (x_v, y_v, z_v) dans le référentiel \mathcal{R}_v lié à la caméra est une transformation rigide correspondant à trois rotations $R_{\alpha\beta\gamma}$ et trois translations T_{xyz} . Cette opération s'écrit sous la forme suivante :

$$\bar{X}_v = T_1 \cdot \bar{X}_a \text{ avec } T_1 = \begin{pmatrix} R_{\alpha\beta\gamma} & T_{xyz} \\ 0 & 1 \end{pmatrix} \quad (8)$$

T_1 étant une matrice 4x4, $R_{\alpha\beta\gamma}$ une matrice 3x3, T_{xyz} un vecteur 3x1, $\bar{X}_v = (x_v \ y_v \ z_v \ 1)^t$ et $\bar{X}_a = (x_a \ y_a \ z_a \ 1)^t$.

Projection perspective : soit $\bar{X}_v = (x_v \ y_v \ z_v \ 1)^t$ les coordonnées d'un point de l'espace exprimées dans le repère de la caméra \mathcal{R}_v et (x, y) les coordonnées de sa projection dans le

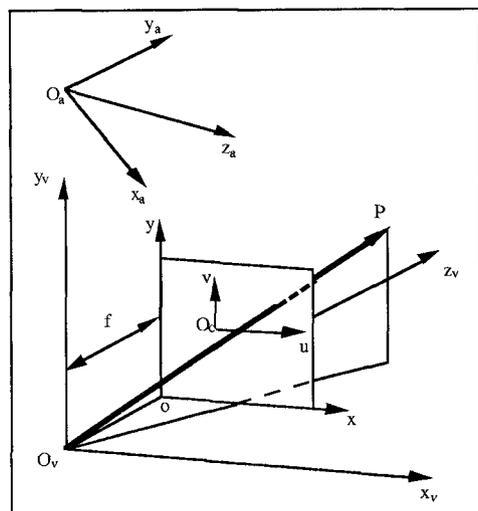


Figure 3. – Le modèle de la caméra.

plan image. La transformation perspective qui permet de calculer les coordonnées x et y est définie par les formules suivantes :

$$\frac{x}{f} = \frac{x_v}{z_v} \text{ et } \frac{y}{f} = \frac{y_v}{z_v}$$

où f est la distance focale de la caméra. Cette relation n'est pas linéaire mais elle peut s'exprimer sous forme matricielle en utilisant des coordonnées homogènes. Soient $(w.x, w.y, w)$ les coordonnées homogènes de la projection de ce point et T_2 la matrice de projection perspective :

$$\begin{pmatrix} w.x \\ w.y \\ w \end{pmatrix} = T_2 \cdot \bar{X}_v \text{ avec } T_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 0 \end{pmatrix} \quad (9)$$

Numérisation : cette dernière opération réalise le passage du système de coordonnées homogènes défini en (9) et lié au repère de la caméra, au système de coordonnées homogènes lié à l'image dont l'unité est le pixel. Ce modèle tient compte du fait que le centre de la matrice ne correspond pas à la position exacte de la projection du centre optique et de la discrétisation de la surface photosensible.

$$\begin{pmatrix} w.u \\ w.v \\ w \end{pmatrix} = T_3 \cdot \begin{pmatrix} w.x \\ w.y \\ w \end{pmatrix} \text{ avec } T_3 = \begin{pmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (10)$$

où

- (u_0, v_0) sont les coordonnées en pixels de l'intersection de l'axe optique et du plan image.
- k_u et k_v sont les facteurs d'échelle respectivement sur les axes u et v .

La transformation globale s'écrit alors comme la combinaison des opérations élémentaires :

$$\begin{pmatrix} w.u \\ w.v \\ w \end{pmatrix} = T_3 \cdot T_2 \cdot T_1 \cdot \begin{pmatrix} x_a \\ y_a \\ z_a \\ 1 \end{pmatrix} = M_{ac} \cdot \begin{pmatrix} x_a \\ y_a \\ z_a \\ 1 \end{pmatrix} \quad (11)$$

Les paramètres de la caméra sont donc divisés en deux catégories :

- Les paramètres intrinsèques (k_u, k_v, u_0, v_0, f) qui apparaissent dans les matrices T_2 et T_3 sont des coefficients qui dépendent uniquement de la caméra utilisée et du mode de numérisation.
- Les six paramètres extrinsèques de la matrice T_1 qui dépendent de la position du repère dans lequel les données sont exprimées.

Calibrer la caméra revient à estimer les 12 coefficients m_{acij} de la matrice M_{ac} . Comme les coefficients sont calculés à un facteur près à cause du facteur de projection w , on choisit classiquement $m_{ac34} = 1$. Les onze autres coefficients de la matrice M_{ac} sont calculés par une méthode du type des moindres carrés à l'aide d'au moins six points non coplanaires.

Les relations (3) et (11) permettent d'effectuer la même transformation à partir du repère \mathcal{R}_3 lié au capteur en écrivant :

$$\begin{pmatrix} w.u \\ w.v \\ w \end{pmatrix} = M_{3c} \cdot \begin{pmatrix} x_3 \\ y_3 \\ z_3 \\ 1 \end{pmatrix} \quad (12)$$

avec

$$M_{3c} = M_{ac} \cdot \bar{M}_{3a} \text{ où } \bar{M}_{3a} = \begin{pmatrix} M_{3a} & 0 \\ 0 & 1 \end{pmatrix}$$

3. un capteur intelligent

Généralement, on dit d'un capteur qu'il est intelligent si sa manière de percevoir l'information est capable d'évoluer en fonction de la situation dans laquelle il se trouve. Dans cette optique, nous avons doté le capteur multisensoriel de deux facultés qui font de lui un capteur intelligent. Elles permettent une coopération étroite entre les mondes de luminance et 3D. En particulier, les conclusions issues des traitements de l'image de luminance vont influencer le lieu d'acquisition des données 3D. La première faculté est la perception dirigée, la seconde est un asservissement visuel.

3.1. la perception dirigée

Parfois, il est très délicat d'obtenir un résultat escompté grâce au traitement de l'image de luminance. La détection des obstacles routiers est un exemple de problème qui n'est pas complètement

résolu en traitement d'images. Cependant, une approche peut consister à déterminer des zones d'intérêt dans lesquelles la présomption de présence d'un obstacle est importante. A partir de là, il peut être très utile, voire déterminant de pouvoir faire l'acquisition d'une image 3D dans une zone de l'espace définie dans l'image de luminance afin de confirmer ou d'infirmer la présence de l'obstacle dans cette zone.

Il s'agit donc de diriger la perception de la partie capteur 3D du capteur multisensoriel dans une zone définie dans l'image de luminance. Le problème qui se pose est donc le problème inverse. En effet, le problème direct est résolu par la chaîne complète de calibrage qui permet, comme on vient de le voir, de déterminer la position (u, v) de la projection dans l'image du faisceau laser à partir des données $(\rho, \theta_x, \theta_y)$ issues du capteur. Le problème inverse consiste donc à déterminer la position des miroirs (θ_x, θ_y) pour que la projection dans l'image de l'impact laser se situe en (u, v) .

A partir de l'équation (12), on peut écrire :

$$\begin{cases} u = \frac{m_{3c11}.x_3 + m_{3c12}.y_3 + m_{3c13}.z_3 + m_{3c14}}{m_{3c31}.x_3 + m_{3c32}.y_3 + m_{3c33}.z_3 + 1} \\ v = \frac{m_{3c21}.x_3 + m_{3c22}.y_3 + m_{3c23}.z_3 + m_{3c24}}{m_{3c31}.x_3 + m_{3c32}.y_3 + m_{3c33}.z_3 + 1} \end{cases} \quad (13)$$

En introduisant deux nouvelles variables $t_1 = x_3/z_3$ et $t_2 = y_3/z_3$ on peut réécrire (13) comme suit (z_3 n'est jamais nul dans notre configuration expérimentale) :

$$\begin{cases} u = \frac{m_{3c11}.t_1 + m_{3c12}.t_2 + m_{3c13} + m_{3c14}/z_3}{m_{3c31}.t_1 + m_{3c32}.t_2 + m_{3c33} + 1/z_3} \\ v = \frac{m_{3c21}.t_1 + m_{3c22}.t_2 + m_{3c23} + m_{3c24}/z_3}{m_{3c31}.t_1 + m_{3c32}.t_2 + m_{3c33} + 1/z_3} \end{cases} \quad (14)$$

d'où l'on tire le système :

$$\begin{cases} (m_{3c31}.u - m_{3c11}).t_1 + (m_{3c32}.u - m_{3c12}).t_2 \\ \quad = (m_{3c13} - m_{3c33}.u) + \frac{m_{3c14} - u}{z_3} \\ (m_{3c31}.v - m_{3c21}).t_1 + (m_{3c32}.v - m_{3c22}).t_2 \\ \quad = (m_{3c23} - m_{3c33}.v) + \frac{m_{3c24} - v}{z_3} \end{cases} \quad (15)$$

En définissant les matrices suivantes, avec $z_3 = \rho \cdot \cos \theta_x \cdot \cos \theta_y$:

$$t = \begin{pmatrix} t_1 & t_2 \end{pmatrix}^t$$

$$L = \begin{pmatrix} m_{3c31}.u - m_{3c11} & m_{3c32}.u - m_{3c12} \\ m_{3c31}.v - m_{3c21} & m_{3c32}.v - m_{3c22} \end{pmatrix}$$

$$S = \begin{pmatrix} m_{3c13} - m_{3c33}.u \\ m_{3c23} - m_{3c33}.v \end{pmatrix}$$

$$N = \begin{pmatrix} \frac{m_{3c14} - u}{\cos \theta_x \cdot \cos \theta_y} \\ \frac{m_{3c24} - v}{\cos \theta_x \cdot \cos \theta_y} \end{pmatrix}$$

Le système (15) peut être écrit sous une forme matricielle :

$$L.t = S + \rho^{-1}.N$$

et finalement

$$t = L^{-1}.S + \rho^{-1}.L^{-1}.N \quad (16)$$

Cette équation fait apparaître deux termes : le premier est indépendant de ρ , alors que le second est proportionnel à ρ^{-1} . On peut, en première approximation, négliger le second terme et ainsi déterminer (θ_x, θ_y) en résolvant le système suivant :

$$\begin{cases} \theta_x = \arctan(t_1 \cdot \cos \theta_y) \\ \theta_y = \arctan(t_2) \\ t = L^{-1}.S \end{cases}$$

La valeur de ρ peut être réinjectée de manière itérative dans la relation (16) pour affiner les valeurs de θ_x et θ_y .

Ainsi, la perception dirigée permet l'acquisition d'une image 3D dans une zone définie dans l'image de luminance.

3.2. l'asservissement visuel

La seconde faculté du capteur multisensoriel est un asservissement visuel du faisceau laser sur un point mobile de l'image 2D. Dans ce cas, le capteur 3D n'est pas utilisé comme un capteur d'image mais il est asservi en position dans une direction unique. Cette faculté permet, dans une procédure de suivi d'obstacle par exemple, de connaître en permanence la position spatiale de cet obstacle.

Le point clef de cette approche est la détermination du jacobien J du capteur. En d'autres termes, il s'agit d'exprimer les variations de la projection d'un point dans l'image (\dot{u}, \dot{v}) en fonction des variations des déviations des miroirs $(\dot{\theta}_x, \dot{\theta}_y)$ en considérant que la distance ρ de la cible à suivre est constante entre deux itérations :

$$\begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix} = J \cdot \begin{pmatrix} \dot{\theta}_x \\ \dot{\theta}_y \end{pmatrix} \quad (17)$$

Le jacobien J peut être obtenu à partir des relations suivantes :

$$\begin{pmatrix} w.u \\ w.v \\ w \end{pmatrix} = M_{3c} \cdot \begin{pmatrix} x_3 \\ y_3 \\ z_3 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} x_3 = \rho \cdot \sin \theta_x \\ y_3 = \rho \cdot \cos \theta_x \cdot \sin \theta_y \\ z_3 = \rho \cdot \cos \theta_x \cdot \cos \theta_y \end{pmatrix} \quad (18)$$

En effet, à partir de la relation (12) on peut écrire :

$$\begin{cases} \dot{u} = \frac{\partial u}{\partial x_3} \cdot \dot{x}_3 + \frac{\partial u}{\partial y_3} \cdot \dot{y}_3 + \frac{\partial u}{\partial z_3} \cdot \dot{z}_3 \\ \dot{v} = \frac{\partial v}{\partial x_3} \cdot \dot{x}_3 + \frac{\partial v}{\partial y_3} \cdot \dot{y}_3 + \frac{\partial v}{\partial z_3} \cdot \dot{z}_3 \end{cases}$$

avec

$$\begin{aligned} \frac{\partial u}{\partial x_3} &= \frac{1}{w^2} \cdot \left(\frac{\partial(w \cdot u)}{\partial x_3} \cdot w - (w \cdot u) \cdot \frac{\partial w}{\partial x_3} \right) = \frac{m_{3c11}}{w} - \frac{m_{3c31} \cdot (w \cdot u)}{w^2} \\ \frac{\partial u}{\partial y_3} &= \frac{1}{w^2} \cdot \left(\frac{\partial(w \cdot u)}{\partial y_3} \cdot w - (w \cdot u) \cdot \frac{\partial w}{\partial y_3} \right) = \frac{m_{3c12}}{w} - \frac{m_{3c32} \cdot (w \cdot u)}{w^2} \\ \frac{\partial u}{\partial z_3} &= \frac{1}{w^2} \cdot \left(\frac{\partial(w \cdot u)}{\partial z_3} \cdot w - (w \cdot u) \cdot \frac{\partial w}{\partial z_3} \right) = \frac{m_{3c13}}{w} - \frac{m_{3c33} \cdot (w \cdot u)}{w^2} \\ \frac{\partial v}{\partial x_3} &= \frac{1}{w^2} \cdot \left(\frac{\partial(w \cdot v)}{\partial x_3} \cdot w - (w \cdot v) \cdot \frac{\partial w}{\partial x_3} \right) = \frac{m_{3c21}}{w} - \frac{m_{3c31} \cdot (w \cdot v)}{w^2} \\ \frac{\partial v}{\partial y_3} &= \frac{1}{w^2} \cdot \left(\frac{\partial(w \cdot v)}{\partial y_3} \cdot w - (w \cdot v) \cdot \frac{\partial w}{\partial y_3} \right) = \frac{m_{3c22}}{w} - \frac{m_{3c32} \cdot (w \cdot v)}{w^2} \\ \frac{\partial v}{\partial z_3} &= \frac{1}{w^2} \cdot \left(\frac{\partial(w \cdot v)}{\partial z_3} \cdot w - (w \cdot v) \cdot \frac{\partial w}{\partial z_3} \right) = \frac{m_{3c23}}{w} - \frac{m_{3c33} \cdot (w \cdot v)}{w^2} \end{aligned}$$

Puis, à partir de la relation (18), si l'on suppose que ρ est constant entre deux itérations, il vient :

$$\begin{cases} \dot{x}_3 = \frac{\partial x_3}{\partial \theta_x} \cdot \dot{\theta}_x + \frac{\partial x_3}{\partial \theta_y} \cdot \dot{\theta}_y \\ \dot{y}_3 = \frac{\partial y_3}{\partial \theta_x} \cdot \dot{\theta}_x + \frac{\partial y_3}{\partial \theta_y} \cdot \dot{\theta}_y \\ \dot{z}_3 = \frac{\partial z_3}{\partial \theta_x} \cdot \dot{\theta}_x + \frac{\partial z_3}{\partial \theta_y} \cdot \dot{\theta}_y \end{cases}$$

avec

$$\begin{aligned} \frac{\partial x_3}{\partial \theta_x} &= x_3 \cdot \cot \theta_x & \frac{\partial x_3}{\partial \theta_y} &= 0 \\ \frac{\partial y_3}{\partial \theta_x} &= -y_3 \cdot \tan \theta_x & \frac{\partial y_3}{\partial \theta_y} &= z_3 \\ \frac{\partial z_3}{\partial \theta_x} &= -z_3 \cdot \tan \theta_x & \frac{\partial z_3}{\partial \theta_y} &= -y_3 \end{aligned}$$

ce qui permet d'obtenir finalement :

$$J = \begin{pmatrix} \frac{m_{3c11}}{w} - \frac{m_{3c31} \cdot (w \cdot u)}{w^2} & \frac{m_{3c12}}{w} - \frac{m_{3c32} \cdot (w \cdot u)}{w^2} & \frac{m_{3c13}}{w} - \frac{m_{3c33} \cdot (w \cdot u)}{w^2} \\ \frac{m_{3c21}}{w} - \frac{m_{3c31} \cdot (w \cdot v)}{w^2} & \frac{m_{3c22}}{w} - \frac{m_{3c32} \cdot (w \cdot v)}{w^2} & \frac{m_{3c23}}{w} - \frac{m_{3c33} \cdot (w \cdot v)}{w^2} \end{pmatrix} \cdot \begin{pmatrix} x_3 \cdot \cot \theta_x & 0 \\ -y_3 \cdot \tan \theta_x & z_3 \\ -z_3 \cdot \tan \theta_x & -y_3 \end{pmatrix}$$

En inversant la relation (17), on peut exprimer les variations de déviation à appliquer aux miroirs ($\dot{\theta}_x, \dot{\theta}_y$) en fonction des variations de position (\dot{u}, \dot{v}) de la projection du faisceau laser dans l'image :

$$\begin{pmatrix} \dot{\theta}_x \\ \dot{\theta}_y \end{pmatrix} = J^{-1} \cdot \begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix} \quad (19)$$

Grâce à cette relation, on peut réaliser un asservissement visuel du faisceau laser où la consigne est la position d'un point dans l'image dont on connaît la distance à un instant initial. La figure 4 précise le principe de l'asservissement. Le capteur 3D délivre la position (ρ, θ_x, θ_y) d'un point de l'espace, le calibrage du capteur 3D suivi de celui de la caméra fournissent la position (u, v) de la projection de ce point dans l'image vidéo. Cette position comparée avec la consigne (u_c, v_c) donnée dans le plan image constitue l'erreur de l'asservissement. La relation (19) permet de déterminer la correction de déviation à apporter aux miroirs afin d'asservir le faisceau laser sur la cible.

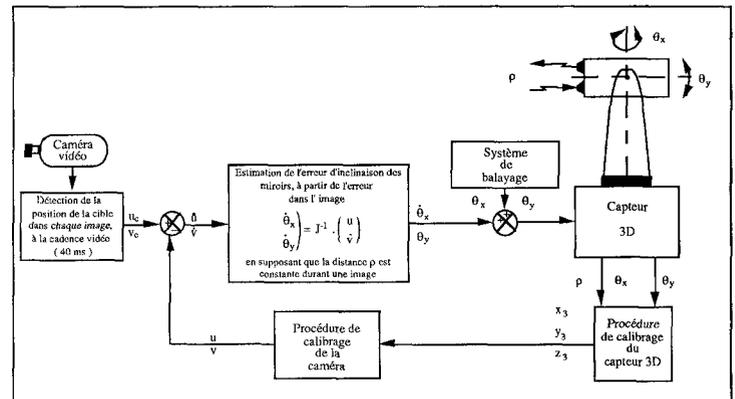


Figure 4. – Synoptique de l'asservissement visuel.

4. détection et suivi d'obstacles par fusion multisensorielle

4.1. introduction

Nous proposons ici une méthode de détection et de suivi des obstacles. En effet, s'il est absolument nécessaire de détecter les obstacles présents sur la chaussée afin de les éviter, il est également indispensable de connaître leur comportement cinématique pour effectuer correctement leur évitement.

L'analyse du comportement cinématique passe obligatoirement par une analyse temporelle plus ou moins longue d'où un suivi temporel. Cette approche est déduite du comportement du conducteur humain : dans une phase pré-attentive, il scrute l'ensemble

de la scène à la recherche d'éventuels obstacles, puis, lorsqu'un véhicule apparaît, il focalise son attention sur lui pour analyser son comportement cinématique plus en détails et ainsi l'aborder en toute sécurité ou bien engager la manoeuvre appropriée.

4.2. la détection d'obstacles

La détection d'obstacles correspond à la phase pré-attentive. Nous avons envisagé deux approches différentes. Dans cette approche que nous ne détaillons pas dans cet article, la détection d'obstacles est faite par une analyse préliminaire de l'image de luminance qui permet de déduire des zones d'intérêt dans lesquelles la présence d'obstacles est fort probable. Les obstacles sont ensuite éventuellement validés par l'acquisition d'une image 3D grâce à la perception dirigée. Les fenêtres d'analyse sont issues de la recherche de segments d'intérêt en utilisant le même principe que l'algorithme de suivi que nous proposerons dans une prochaine section (pour plus de détails consulter [XTAT+93]).

Dans la seconde approche que nous reprenons ici, l'ensemble de la scène est scrutée par un balayage du faisceau laser pour obtenir une image 3D. Le balayage se poursuit tant qu'un obstacle n'est pas apparu. La détection est exclusivement basée sur l'exploitation de l'image 3D. L'algorithme de détection se déroule en deux phases successives qui sont la segmentation en régions de l'image 3D puis la reconnaissance des obstacles parmi ces régions.

4.2.1. segmentation en régions de l'image 3D

Le but de la segmentation en régions est d'obtenir des zones formées d'impacts laser spatialement proches. Cela signifie que deux impacts appartiennent à une même région si, d'une part, ils se situent à la même distance z_a de notre véhicule avec une tolérance σ_z , et si, d'autre part, ils sont connexes c'est-à-dire s'ils sont reliés par des voisins appartenant à la même région. Le critère de segmentation est la distance z_a car, dans l'hypothèse d'une route plane, les obstacles routiers peuvent être modélisés par des portions de plans orthogonaux à la route constitués de points situés à une distance z_a constante.

Cette façon de segmenter l'image 3D est efficace car la route n'est pas incluse dans l'image grâce à l'implantation particulière du capteur dans le véhicule, les obstacles apparaissent donc de façon discrète dans la scène.

Pour segmenter l'image, nous avons retenu un algorithme de croissance de régions. Les régions sont construites à partir d'un impact et croissent en incorporant, de proche en proche, les impacts situés à une distance z_a similaire.

Le résultat de la segmentation est donc un ensemble de régions R_i comportant N_i impacts. Parmi ces régions, certaines peuvent être des obstacles, d'autres pas. Il est donc nécessaire de mettre en œuvre une étape d'identification pour distinguer les véhicules parmi l'ensemble des régions.

4.2.2. l'identification des obstacles

L'identification consiste donc à extraire les obstacles de l'ensemble des régions. Nous nous bornons dans un premier temps à la détection des obstacles routiers de type voiture qui sont évidemment les plus courants donc ceux qui présentent un danger quasi permanent. Ce sont également les plus simples à détecter.

Le principe de l'identification est simple. Le modèle 3D d'un véhicule vu de derrière est un rectangle. Il s'agit de comparer les dimensions de chaque région à celle du modèle afin de décider si la région est un obstacle. La position de la région dans l'espace est également prise en compte.

Ainsi, chaque région R_i issue de la segmentation comportant N_i impacts est paramétrée par le vecteur :

$$R_i = (X_{min,i}, X_{max,i}, Y_{min,i}, Y_{max,i}, Z_i)$$

où $(X_{min,i}, X_{max,i})$ sont les valeurs minimale et maximale des impacts dans la direction x_a , $(Y_{min,i}, Y_{max,i})$ les valeurs minimale et maximale dans la direction y_a , et Z_i la distance de la région à notre véhicule.

Les cinq paramètres peuvent être calculés de la façon suivante :

$$\begin{cases} X_{min,i} = \min \{ x_{a,i,j}, & \forall j \in (1, \dots, N_i) \} \\ X_{max,i} = \max \{ x_{a,i,j}, & \forall j \in (1, \dots, N_i) \} \\ Y_{min,i} = \min \{ y_{a,i,j}, & \forall j \in (1, \dots, N_i) \} \\ Y_{max,i} = \max \{ y_{a,i,j}, & \forall j \in (1, \dots, N_i) \} \\ Z_i = \left(\sum_{j=1}^{N_i} z_{a,i,j} \right) / N_i \end{cases}$$

Ces paramètres permettent à leur tour de calculer trois caractéristiques des régions qui seront déterminantes pour l'identification des obstacles :

- la hauteur h_i : $h_i = Y_{max,i} - Y_{min,i}$.
- la largeur l_i : $l_i = X_{max,i} - X_{min,i}$.
- la position du centre $C_i(X_{m,i}, Y_{m,i}, Z_i)$:

$$\begin{cases} X_{m,i} = (X_{min,i} + X_{max,i}) / 2 \\ Y_{m,i} = (Y_{min,i} + Y_{max,i}) / 2 \end{cases}$$

Ces trois caractéristiques permettent de savoir si pour une région donnée, on est en présence d'un obstacle en les comparant aux dimensions des modèles de véhicules moyennant certaines tolérances.

Soit \mathcal{O} l'ensemble des N obstacles routiers. Un obstacle est représenté par le vecteur $O_j = (L_j, H_j)$ où L_j est la largeur de l'obstacle et H_j sa hauteur. Pour chaque région R_i , on peut écrire :

$$R_i \in \mathcal{O} \text{ si } \left\{ \begin{array}{l} h_i = H_j \pm \sigma_{H,j} \\ l_i = L_j \pm \sigma_{L,j} \\ Y_{m,i} = (H_j \pm \sigma_{H,j})/2 \end{array} \right\} \forall j \in (1, \dots, N).$$

où $\sigma_{H,j}$ et $\sigma_{L,j}$ sont les tolérances sur les dimensions de l'obstacle O_j . Elles tiennent compte des variations de tailles possibles pour

une même catégorie de véhicules mais aussi de l'incertitude sur les mesures délivrées par le capteur 3D.

Lorsque l'on bénéficie conjointement d'un algorithme de localisation de la route [C91, J93], on ajoute la contrainte :

$$X_{m,i} \in [X_{rg}(Z_i), X_{rd}(Z_i)]$$

où $X_{rg}(Z_i)$ et $X_{rd}(Z_i)$ sont les abscisses des bords gauche et droit de la chaussée dans le repère absolu \mathcal{R}_a à la distance Z_i de notre véhicule. Cette contrainte impose à la région de se trouver sur la route pour être un obstacle.

Nous avons ainsi détecté les obstacles, il est maintenant nécessaire de mettre en œuvre leur suivi.

4.3. le suivi de l'obstacle

4.3.1. méthode et initialisation

Le but du suivi est de localiser précisément les obstacles pendant une période suffisamment longue afin d'obtenir des caractéristiques dynamiques précises de l'obstacle, en particulier sa vitesse longitudinale relative. Nous avons choisi d'utiliser une méthode multisensorielle pour allier la rapidité du signal vidéo et la précision de la localisation du capteur 3D. En effet, l'obtention de caractéristiques dynamiques est délicate par une approche basée sur la détection d'obstacles dans l'image 3D car l'acquisition d'une image n'est pas suffisamment rapide (200 ms pour une image 10 x 80 points de mesure). Par ailleurs, la détection dans l'ensemble de l'image de luminance, qui est un problème délicat, ne permettrait pas une localisation précise dans l'espace, encore moins un calcul précis de sa vitesse relative. Cette coopération entre les mondes 3D et de luminance utilise la seconde faculté du capteur multisensoriel : l'asservissement visuel (voir section (3.2)). C'est donc la fusion de données multisensorielles qui va permettre de suivre l'obstacle routier.

La détection d'obstacles décrite précédemment constitue la phase d'initialisation de l'algorithme de suivi. Le suivi consiste en la recherche d'un modèle 2D dans une fenêtre d'intérêt en utilisant la connaissance de la distance de l'obstacle à notre véhicule. Inversement, sa nouvelle position dans l'image permet l'asservissement du faisceau laser.

L'initialisation de la fenêtre d'analyse est possible grâce à la relation de calibrage (11) qui permet de calculer la position $c_i(u_{m,i}, v_{m,i})$ du centre de la région obstacle dans le repère \mathcal{R}_c de l'image de luminance en fonction de sa position $C_i(X_{m,i}, Y_{m,i}, Z_i)$ dans le repère absolu \mathcal{R}_a que l'on a déterminée lors de la détection.

La recherche de la nouvelle position de l'obstacle dans l'image s'effectue donc dans une fenêtre d'intérêt F_i centrée sur c_i .

Soit donc F_i la fenêtre d'analyse pour le suivi de l'obstacle R_i , elle est définie par :

$$F_i = (u_{w,min,i}, u_{w,max,i}, v_{w,min,i}, v_{w,max,i})$$

où $(u_{w,min,i}, u_{w,max,i})$ sont les valeurs minimale et maximale de la fenêtre d'intérêt dans la direction u et $(v_{w,min,i}, v_{w,max,i})$ les valeurs minimale et maximale de la fenêtre dans la direction v . Si l'on note $proj$, la transformation réalisée par la relation de calibrage (11), les paramètres de F_i sont calculés comme suit :

$$\begin{cases} (u_{w,min,i} \quad v_{w,min,i}) = proj \begin{pmatrix} X_{m,i} - g_{wl} \cdot l_i / 2 \\ Y_{m,i} - g_{wh} \cdot h_i / 2 \\ Z_i \end{pmatrix} \\ (u_{w,max,i} \quad v_{w,max,i}) = proj \begin{pmatrix} X_{m,i} + g_{wl} \cdot l_i / 2 \\ Y_{m,i} + g_{wh} \cdot h_i / 2 \\ Z_i \end{pmatrix} \end{cases}$$

où g_{wl} et g_{wh} sont les gains horizontal et vertical des dimensions de la fenêtre F_i . Ils correspondent à l'accroissement de la taille de la fenêtre d'intérêt dans l'image de luminance par rapport aux dimensions de l'obstacle routier dans cette même image. Plus ces valeurs sont importantes, plus les déplacements de l'obstacle peuvent être importants. En revanche la quantité d'information à traiter est également plus grande.

Le suivi de l'obstacle va donc consister en une détection dans une fenêtre d'intérêt de l'image de luminance. Cette fenêtre sera doublement active : sa position va évoluer avec celle de l'obstacle et sa taille va dépendre de la distance qui est connue en permanence grâce à l'asservissement visuel.

La détection d'obstacle dans la fenêtre d'analyse est une chaîne complète de traitement d'images dont l'ultime étape est la recherche d'un modèle 2D. Le modèle 2D que nous avons choisi découle d'une constatation simple : une voiture vue de derrière présente des segments horizontaux en grand nombre (voir figure 5).

Le modèle est donc un *rectangle* englobant de tels segments. Compte tenu de ce modèle, la chaîne de traitement se décompose comme suit : détection des points de contour, chaînage des points de contour, approximation polygonale et recherche du modèle de véhicule.

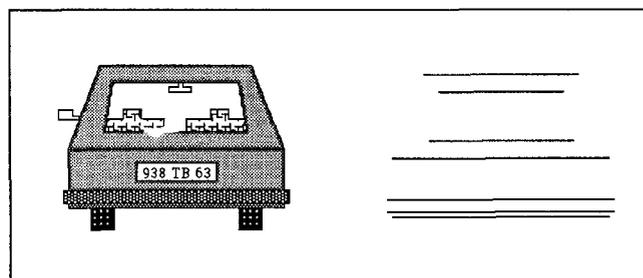


Figure 5. – Modèle de véhicule vue de derrière.

4.3.2. recherche de la nouvelle position de l'obstacle

A l'issue d'une approximation polygonale des points de contour, l'image peut être décrite par l'ensemble \mathcal{S} des N_s segments :

$$\mathcal{S} = \{(p_{1i}, p_{2i}), \forall i \in (1, \dots, N_s)\}$$

où les p_{1i} et p_{2i} sont les points extrémités du segment s_i définis par leurs coordonnées :

$$\begin{cases} p_{1i} = (u_{1i}, v_{1i}) \\ p_{2i} = (u_{2i}, v_{2i}) \end{cases}$$

A partir de la connaissance des segments, la recherche de l'obstacle dans la fenêtre d'intérêt se déroule en trois phases successives :

- la sélection des segments d'intérêt,
- la construction des différents rectangles,
- la recherche du meilleur modèle.

Afin de limiter la dimension du problème et de manipuler exclusivement des caractéristiques importantes de l'image, nous introduisons la notion de segments d'intérêt.

La sélection de ces segments s'effectue selon deux critères : le premier est bien entendu l'horizontalité, le second est la longueur. En effet, un segment doit être à peu près horizontal et avoir une longueur cohérente avec la largeur w_i de l'obstacle dans l'image 2D. Ainsi, on définit le sous-ensemble \mathcal{SI} , des N_{SI} segments d'intérêt, de \mathcal{S} . Pour tout segment $s_i \in \mathcal{S}$,

$$s_i \in \mathcal{SI} \text{ si } \begin{cases} |a_i| \leq \sigma_p \\ (w_i - \sigma_w) < |u_{1i} - u_{2i}| < w_i \end{cases}$$

où σ_p est la tolérance sur la pente des segments, w_i la largeur de l'obstacle dans l'image et σ_w la tolérance sur cette largeur. Cette tolérance traduit le fait que la longueur de tous les segments n'est pas égale à la largeur de l'obstacle dans l'image. Ainsi, on ne retient que les segments dont la longueur est inférieure à celle du segment le plus long de la voiture vue de derrière.

Comme l'on connaît en permanence la position $(X_{m,i}, Y_{m,i}, Z_i)$ de la région obstacle dans le repère absolu grâce à l'asservissement visuel, la largeur w_i est calculée pour chacune des positions de l'obstacle dans l'image par :

$$w_i = \left\| \text{proj} \begin{pmatrix} X_{m,i} + l_i/2 \\ Y_{m,i} \\ Z_i \end{pmatrix} - \text{proj} \begin{pmatrix} X_{m,i} - l_i/2 \\ Y_{m,i} \\ Z_i \end{pmatrix} \right\|$$

On construit ensuite l'ensemble \mathcal{E}_r des N_r rectangles englobant n_{si} segments d'intérêt. On note que :

$$N_r = C_{N_{si}}^{n_{si}}$$

Les rectangles r_i sont définis par :

$$r_i = (u_{r,min,i}, u_{r,max,i}, v_{r,min,i}, v_{r,max,i})$$

où $(u_{r,min,i}, u_{r,max,i})$ sont les valeurs minimale et maximale du rectangle dans la direction u et $(v_{r,min,i}, v_{r,max,i})$ les valeurs minimale et maximale du rectangle dans la direction v .

Etant donné une sélection de n_{si} segments d'intérêt s_i , le rectangle correspondant est calculé comme suit :

$$\begin{cases} u_{r,min,i} = \min \{u_{1i} \text{ et } u_{2i}, \forall i \in (1, \dots, n_{si})\} \\ u_{r,max,i} = \max \{u_{1i} \text{ et } u_{2i}, \forall i \in (1, \dots, n_{si})\} \\ v_{r,min,i} = \min \{v_{1i} \text{ et } v_{2i}, \forall i \in (1, \dots, n_{si})\} \\ v_{r,max,i} = \max \{v_{1i} \text{ et } v_{2i}, \forall i \in (1, \dots, n_{si})\} \end{cases}$$

Nous recherchons enfin le rectangle dont les dimensions correspondent au mieux au véhicule que nous suivons. Pour cela, nous cherchons à minimiser un critère K_i qui tient compte des dimensions du rectangle par l'intermédiaire de la largeur w_i ainsi que d'un facteur de forme $f_{f,i}$ tel que :

$$f_{f,i} = \frac{4 \cdot h_i}{3 \cdot l_i}$$

Le facteur de forme étant défini par les dimensions 3D de l'obstacle, il faut tenir compte du rapport 4/3 de la matrice de la caméra. On considère ici que les proportions sont conservées quelle que soit la position de l'obstacle dans l'image.

Ce critère est défini par :

$$K_i = |w_i - (u_{r,max,i} - u_{r,min,i})| + |f_{f,i} - \frac{v_{r,max,i} - v_{r,min,i}}{u_{r,max,i} - u_{r,min,i}}|$$

Ainsi, la nouvelle position de l'obstacle o_i est définie par :

$$o_i \in \mathcal{E}_r \text{ et tel que } K_i < K_j \quad \forall j \neq i \in (1, \dots, N_r)$$

La connaissance de o_i donne les nouvelles consignes (u_c, v_c) de l'asservissement que l'on place au centre de l'obstacle :

$$\begin{cases} u_c = \frac{u_{r,min,i} + u_{r,max,i}}{2} \\ v_c = \frac{v_{r,min,i} + v_{r,max,i}}{2} \end{cases}$$

Le processus est ainsi réitéré jusqu'à ce qu'il soit nécessaire d'effectuer un nouveau balayage pour analyser l'ensemble de la scène.

Lors de chaque itération, la fenêtre d'analyse F_i est gérée activement grâce au traitement de l'image ainsi qu'à l'asservissement visuel. En effet, la fenêtre est centrée sur la position (u_c, v_c) du centre de l'obstacle dans l'image et ses dimensions m_i et n_i sont fixées par la connaissance de la position 3D par :

$$\begin{cases} m_i = g_{wl} \cdot w_i \\ n_i = \left\| \text{proj} \begin{pmatrix} X_{m,i} \\ Y_{m,i} + g_{wh} \cdot h_i/2 \\ Z_i \end{pmatrix} - \text{proj} \begin{pmatrix} X_{m,i} \\ Y_{m,i} - g_{wh} \cdot h_i/2 \\ Z_i \end{pmatrix} \right\| \end{cases}$$

les nouveaux paramètres de la fenêtre F_i deviennent ainsi :

$$\begin{cases} u_{w,min,i} = u_c - m_i / 2 \\ u_{w,max,i} = u_c + m_i / 2 \\ v_{w,min,i} = v_c - n_i / 2 \\ v_{w,max,i} = v_c + n_i / 2 \end{cases}$$

Pour accélérer considérablement l'exécution des traitements d'images, les plus coûteux en temps de calcul, nous avons implanté l'algorithmique sur une machine parallèle de vision développée au laboratoire : la machine Transvision à base de processeurs du type Transputer [D91].

4.4. calcul de l'état cinématique de l'obstacle

L'algorithme que l'on vient de présenter permet donc de suivre un obstacle et ainsi d'asservir le faisceau laser dans sa direction. Nous mettons à profit la connaissance permanente de la position relative de l'obstacle dans le repère absolu lié à notre véhicule pour en déduire l'évolution de son état dynamique. Nous nous bornons dans un premier temps à calculer la vitesse relative longitudinale et l'accélération relative longitudinale de l'obstacle.

Nous avons choisi de modéliser la trajectoire de l'obstacle par un modèle à accélération constante. La trajectoire de l'obstacle selon l'axe longitudinal s'écrit donc :

$$\begin{aligned} V &= V_0 + \gamma \cdot t \\ z_a &= z_{a0} + V_0 \cdot t + \frac{1}{2} \cdot \gamma \cdot t^2 \end{aligned}$$

où

- z_{a0} est la distance initiale de l'obstacle,
- V_0 la vitesse relative initiale,
- V la vitesse instantanée,
- γ l'accélération relative constante.

Les paramètres sont donc la vitesse relative V et l'accélération γ . L'évolution de ces deux paramètres est estimée par un filtre de Kalman [KB60].

Si l'on écrit les équations d'état et de mesures du filtre respectivement comme suit :

$$\begin{cases} x(k+1) &= M(k+1, k) \cdot x(k) + G(k) \cdot u(k) \\ y(k) &= H(k) \cdot x(k) + v(k) \end{cases}$$

où

- x est le vecteur de paramètres du modèle,
- M la matrice d'état,
- u le processus générateur régissant l'évolution de x ,
- G la matrice de transfert de u ,
- y le vecteur d'observation,
- H la matrice de mesure,
- v le vecteur de bruit de mesure.

Compte tenu du modèle que l'on a choisi, on pose :

$$\begin{aligned} x(k) &= \begin{pmatrix} z_a \\ V \\ \gamma \end{pmatrix} \\ M(k+1, k) &= \begin{pmatrix} 1 & \Delta t & \frac{\Delta t^2}{2} \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{pmatrix} \\ G(k) &= \begin{pmatrix} \frac{\Delta t^2}{2} \\ \Delta t \\ 1 \end{pmatrix} \\ H(k) &= (1 \ 0 \ 0) \end{aligned}$$

où Δt est la période d'échantillonnage.

La mesure $y(k)$ se borne ainsi exclusivement à la coordonnée cartésienne z_a .

5. résultats

Nous présentons quelques résultats concernant la détection et le suivi des obstacles.

La figure 6 représente le résultat de la détection d'obstacles dans une situation réelle de circulation autoroutière. Nous avons choisi les paramètres suivants :

$$\begin{aligned} N &= 1 \text{ (un seul modèle d'obstacle routier)} \\ H_1 &= 1,4 \text{ m} & L_1 &= 1,5 \text{ m} \\ \sigma_{H1} &= 20 \text{ cm} & \sigma_{L1} &= 20 \text{ cm} \end{aligned}$$

Dans une scène relativement simple, on peut voir que les deux voitures présentes ont effectivement été détectées. Par contre, le rail de sécurité n'a pas été pris en compte car la largeur l_i de la région qu'il constitue est quasiment nulle du fait que ce rail est dans l'axe longitudinal de notre véhicule. Les critères de sélection permettent également d'éliminer les ponts ou bien les panneaux indicateurs.

Les figures 7, 8 et 9 illustrent le suivi d'obstacles. Pour une image donnée, la figure 7 présente les traitements successifs. En réalité, les traitements sont faits exclusivement dans la fenêtre d'intérêt, symbolisée par le rectangle, pour laquelle on a choisi des gains $g_{wl} = g_{wh} = 2$. Les contours (b) sont obtenus avec un seuil $\varepsilon_{ch} = 20$. Les segments (c) sont une approximation polygonale des chaînes de contour ne tolérant pas une erreur supérieure à $\varepsilon_s = 2$. Les paramètres d'extraction des segments d'intérêts (d) sont $\sigma_p = 0,1$ et $\sigma_w = 50\% \cdot w_i$. L'ensemble \mathcal{E}_r des rectangles qui regroupent $n_{si} = 3$ segments d'intérêt est représenté en (e). Enfin, (f) montre la superposition du rectangle dont les dimensions correspondent le mieux à celles de l'obstacle. On peut ainsi juger de la qualité de notre approche multisensorielle.

La figure 8 illustre le suivi de cet obstacle dans une séquence de plusieurs images. Le grand rectangle est la fenêtre d'analyse dans laquelle on a détecté la nouvelle position de l'obstacle (petit rectangle).

La figure 9 présente l'évolution des variables lors d'une séquence test et permet aussi de juger de la qualité de l'asservissement visuel. Dans le scénario réel, le véhicule obstacle passe d'un côté à l'autre de la chaussée alors que le véhicule expérimental circule en droite ligne au milieu de la route. C'est la composante horizontale X_m du centre de l'obstacle qui illustre le mieux le scénario. L'ordonnée Y_m n'évolue quasiment pas ce qui est fort logique du fait que l'on circule sur une route horizontale. On notera en particulier que les coordonnées (u_m, v_m) de la projection du faisceau laser

Détection et suivi d'obstacle sur route

dans l'image sont correctement asservies sur la consigne (u_c, v_c) (se reporter également au synoptique de l'asservissement de la figure 4). La démonstration est plus flagrante pour la coordonnée horizontale u_m pour laquelle les variations sont plus importantes. Nous avons également précisé l'évolution des angles de déflexion θ_x et θ_y .

L'ensemble des algorithmes de détection, suivi et calcul de l'état cinématique de l'obstacle ont été mis en œuvre en temps réel sur le véhicule démonstrateur Prolab évoluant dans des conditions réelles de circulation (jusqu'à 130 km/h). Une expérimentation sur cible fixe nous a permis de déterminer l'erreur commise sur les valeurs estimées en vitesse et position par rapport aux valeurs réelles.

- erreur relative pour la distance de 1% sur une plage de mesure allant de 20 à 60 m.
- erreur relative pour le calcul de la vitesse de 5% sur une plage de mesure allant de 30 à 70 km/h.
- Le temps de convergence du filtre est de l'ordre 200 ms.

Ainsi, 200 ms après la détection d'un obstacle, nous pouvons calculer sa vitesse relative à 5% près afin de prédire l'intention de cet obstacle pour alerter le conducteur d'un danger éventuel. Cet aspect de supervision n'est pas traité dans cet article car il fait l'objet d'autres travaux au sein du groupe [HLLM94,HVG94].

6. conclusion

Nous avons décrit une approche multisensorielle pour la perception et l'interprétation de scènes d'environnements naturels par fusion de données vidéo-téléométriques. Nous avons appliqué cette méthode à la détection et au suivi d'obstacles sur scènes routières. De nombreux essais en circulation réelle ont été effectués dans le cadre du projet Prometheus. Ce module de détection est un des éléments essentiels de la chaîne de perception du véhicule démonstrateur Prolab. Cette méthode peut s'adresser à de nombreuses applications de perception.

Les travaux présentés dans cet article donnent des résultats encourageants. Des améliorations sont envisagées, en particulier pour prendre en compte l'aspect prédictif du filtrage de Kalman afin d'inférer les paramètres de la nouvelle zone de recherche à partir des données filtrées. De plus, l'estimation de la distance relative du véhicule suivi peut être étendue à l'estimation de la position 3D. Les développements les plus prometteurs portent sur la gestion des apparitions, disparitions et occlusions d'obstacles.

BIBLIOGRAPHIE

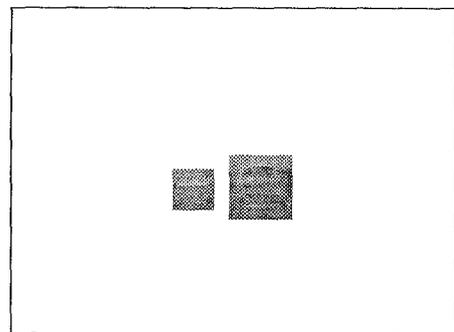
[C91] R. Chapuis : *Suivi de primitives images - Application à la conduite automatique sur route*", Thèse de doctorat d'université, Clermont-Ferrand, 1991.



(a)



(b)



(c)

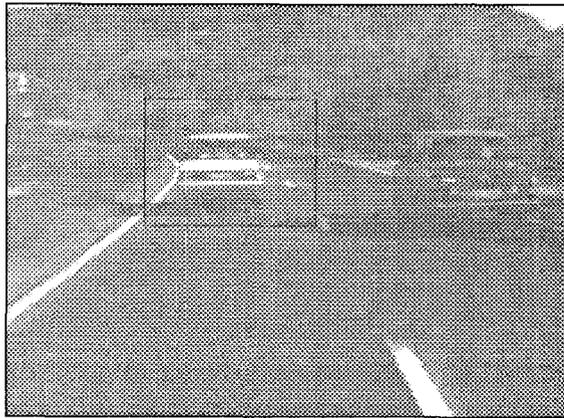
Figure 6. – Détection d'obstacles en situation autoroutière : a) la scène autoroutière, b) l'image 3D (les impacts sont d'autant plus gros, qu'il sont proches) recalée avec l'image de luminance, c) résultat de la détection.

[D91] J.P. Derutin : *A parallel vision machine : TRANSVISION*, Computer Architecture for Machine Perception, pp. 241-251, 1991.

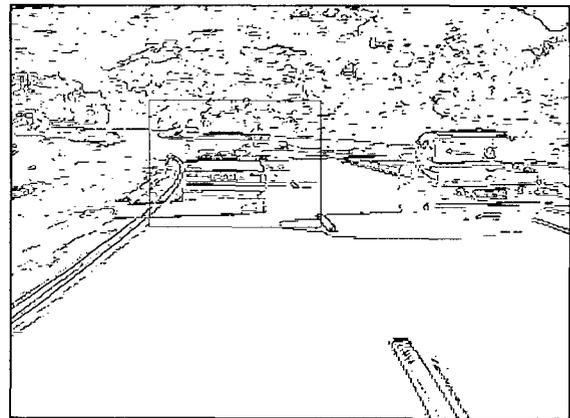
[FT87] O. D. Faugeras, G. Toscani : *Camera Calibration for 3D Computer Vision*, Proc. of International Workshop on Machine Vision and Machine Intelligence, Tokyo, February 87.

[DER94] V. Der Gharapetian, B. El-Eter, M. Rombaut : *Reliable system for dynamic fusion of high level data in the Prolab II demonstrator*, Intelligent Vehicles '94, Paris, 24-26 octobre, pp. 473-478, 1994.

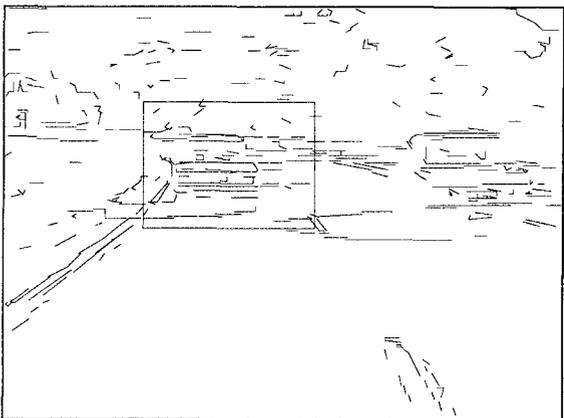
[HLLM94] M. Hassoun, L. Laugier, N. LeFort, D. Meizel : *An assistance system for diagnosis and monitoring of driving manoeuvres*, IEEE-SMC IMACS international symposium on signal processing, robotics and neural networks, Lille, pp.1-6, 1994.



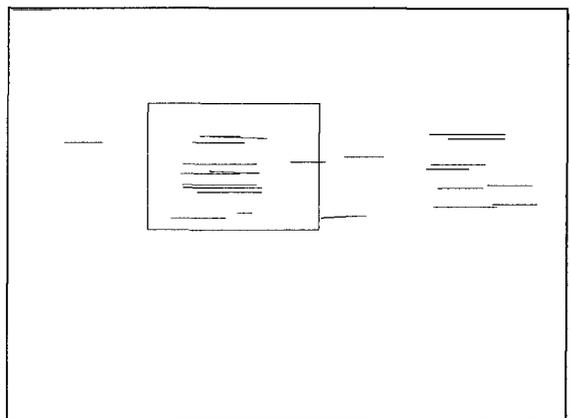
(a)



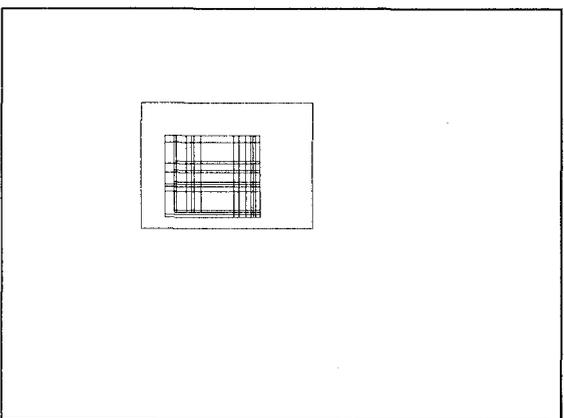
(b)



(c)



(d)



(e)



(f)

Figure 7. – Suivi d'obstacle : a) la sirène routière, b) l'image des contours, c) les segments, d) les segments d'intérêt, e) l'ensemble des rectangles, f) la nouvelle position de l'obstacle.

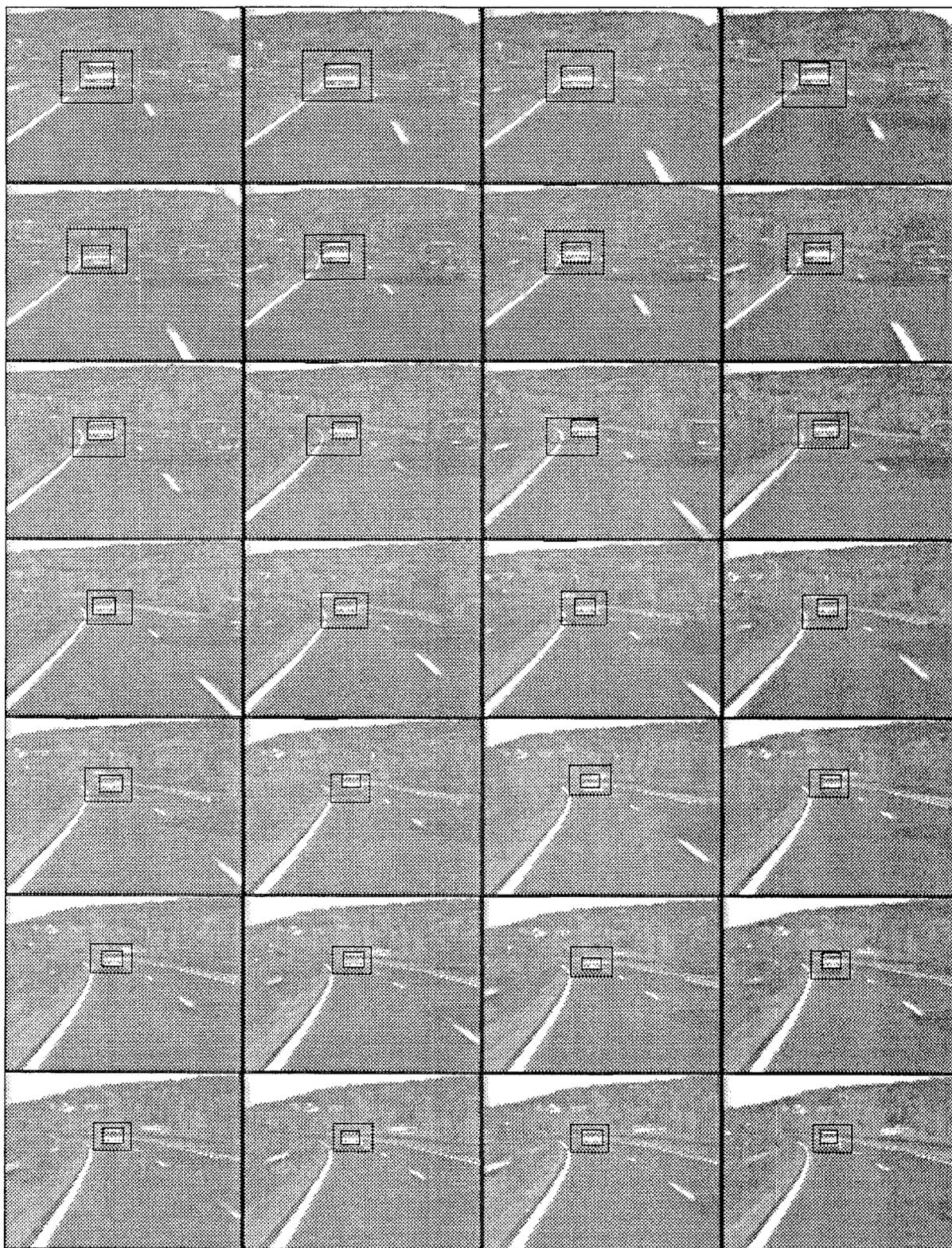


Figure 8. – Suivi d'obstacle sur route.

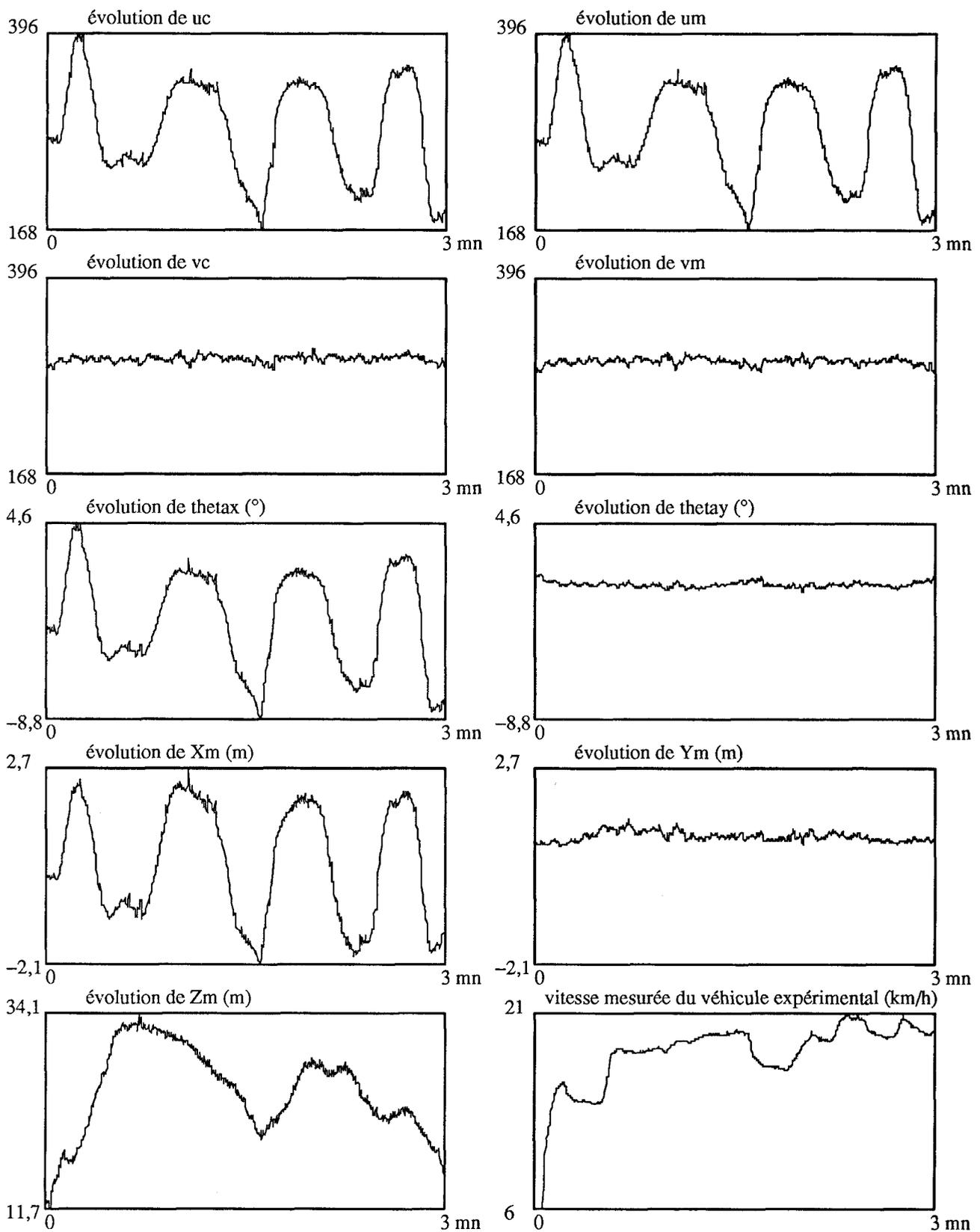


Figure 9. – Evolution des variables lors d'une séquence routière.

Détection et suivi d'obstacle sur route

- [HVG94] D. Hutber, T. Vieville, G. Giraudon : *Data fusion for reliable detection and tracking of multiple obstacles in a road environment. An asynchronous approach*, IEEE-SMC IMACS international symposium on signal processing, robotics and neural networks, Lille, pp.15-18, 1994.
- [H95] D. Hutber : *Suivi multi-capteurs de cibles multiples en vision par ordinateur, appliqué à un véhicule dans un environnement routier*, thèse de l'université de Nice - Sophia Antipolis, novembre 1995.
- [J93] F. Jurie : *Perception visuelle par propagation d'informations dans un réseau d'hypothèses : application à l'interprétation de scènes routières*, Thèse de doctorat d'université, Clermont-Ferrand, 1993.
- [KB60] R. E. Kalman, R.S. Bucy : *New results in linear filtering and prediction theory*, Trans. ASME, Series D journal of Basic Eng., vol. 38, pp 95-101, 1960.
- [R94] R. Reynaud : *La fusion de données, du capteur au raisonnement*, Editorial, Traitement du signal, vol. 11, n 6, pp 431-434, 1994.
- [RM95] R. Reynaud, T. Maurin : *Le concept ProChip : la détection des situations dangereuses - fusion multisensorielle*, L'intégration des systèmes électroniques dans l'automobile du XXI^e siècle, Chapitre 2, pp 90-103, Cépadués éditions, ISBN 2.85428.386.4, 1995.
- [TDD94] F. Thomanek, E.D. Dickmanns, D. Dickmanns : *Multiple object recognition and scene interpretation for autonomous road vehicle guidance*, Intelligent Vehicles '94, Paris, 24-26 octobre, pp. 231-236, 1994.
- [XTAT+93] M. Xie, L. Trassoudaine, J. Alizon, M. Thonnat, J. Gallice : *Active and Intelligent Sensing of Road Obstacles : Application to the European Eureka-Prometheus Project*, 4th Int. Conference on Computer Vision, Berlin, 11-13 mai, pp. 616-623, 1993.

Manuscrit reçu le 6 septembre 1994.

LES AUTEURS

Laurent TRASSOUDAINE



Est titulaire d'une thèse de doctorat de l'université Blaise Pascal de Clermont Ferrand depuis 1993. Il est maître de conférences à l'IUP Génie des Systèmes Industriels de cette même université. Il exerce ses activités de recherche au Laboratoire des Sciences et Matériaux pour l'Electronique et d'Automatique (URA CNRA 1793). Il s'intéresse aux systèmes de perception dans les machines intelligentes et plus particulièrement à la fusion de données multisensorielles et à la segmentation d'images de profondeur.

Paul CHECCHIN



Est ingénieur du CUST de Clermont-Ferrand (section Génie Electrique) et titulaire d'un DEA Electronique et Systèmes de l'Université Blaise Pascal option Vision pour la Robotique. Il poursuit actuellement une thèse de doctorat au LASMEA. Ses travaux de recherches concernent la segmentation d'images de profondeur.

Joseph ALIZON



Est chargé de recherche au CNRS et responsable de l'opération scientifique "Capteurs Intelligents et Fusion de Données" au Laboratoire des Sciences et Matériaux pour l'Electronique, et d'Automatique. Ses centres d'intérêt portent sur la conception et la mise en oeuvre de capteurs extéroceptifs pour robots mobiles évoluant en milieu naturel extérieur.

François COLLANGE



Est maître de conférences à l'IUT de Clermont-Ferrand. Après une thèse de doctorat en Electronique en 1986, il a rejoint en 1987 le Centre de Morphologie Mathématique de Fontainebleau. Depuis 1990, il travaille au Laboratoire des Sciences et Matériaux pour l'Electronique, et d'Automatique (LASMEA). Ses recherches portent sur les méthodes de fusion de données multicapteurs.

Jean GALLICE



Est professeur à l'Université Blaise Pascal. Il enseigne, principalement dans le département Génie Electrique du CUST (école d'ingénieurs), le traitement numérique des signaux. Il est co-responsable au LASMEA, URA 1793 du CNRS, du groupe de recherche GRAVIR (Automatique : Vision et Robotique). Il s'intéresse plus particulièrement à la vision active et aux asservissements visuels.