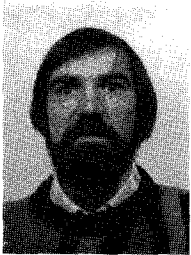


Estimation récursive de primitives 3 D au moyen d'une caméra mobile

Recursive estimation of 3 D features

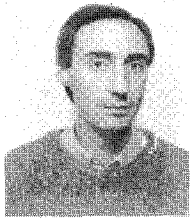
using a moving camera



Patrick RIVES

IRISA/INRIA, Centre de Rennes, Campus de Beaulieu, 35042 Rennes Cedex.

Docteur 3^e cycle en Conception Assistée de Système. Responsable de la partie Vision Robotique du Projet Image au Centre INRIA de Rennes. Activités de Recherches centrées sur l'exploitation du mouvement en vision, la modélisation de primitives 3D, la reconstruction de scène 3D, le développement d'outils de simulation : Application aux robots mobiles évoluant dans des univers inconnus *a priori*.



Bernard ESPIAU

IRISA/INRIA, Centre de Rennes, Campus de Beaulieu, 35042 RENNES CEDEX.

Ingénieur ENSM, Docteur-Ingénieur en Automatique, Docteur ès Sciences. Chef de Projet Robotique au Centre INRIA de Rennes. Activités de Recherches centrées sur l'Analyse des boucles perception/commande en robotique : aspects théoriques, architectures temps réel associées, CAO : Applications à la robotique de milieu hostile et à la téléopération.

RÉSUMÉ

Cet article présente un schéma d'estimation de primitives géométriques 3D à partir d'informations contenues dans une séquence d'images fournie par une caméra mobile dont le mouvement est connu. Nous rappelons tout d'abord les modèles de base utilisés : points et segments de droites, et nous établissons comment leur projection perspective dans l'image varie avec le mouvement de la caméra. Des techniques de filtrage récursif au sein de la séquence d'images sont utilisées pour reconstruire, de façon incrémentale, la scène 3D, au fur et à mesure du déplacement de la caméra. Des résultats expérimentaux dans le domaine de la robotique sont présentés.

MOTS CLÉS

Mouvement, séquence d'images, Optic Flow, filtrage récursif, boucle fermée, robotique.

SUMMARY

This paper presents a scheme allowing to estimate parameters which describe geometrical structures in a 3D scene by only using informations issued from a sequence of images provided by a mobile vision sensor with known motion. We first recall the basic used models: points and lines, and then we relate their perspective projection in the image plane to the camera motion. Some techniques of recursive filtering are used into the sequence of images to incrementally build the 3D scene all along the displacement of the camera. Some experimental results in the field of robotics are given.

KEY WORDS

Motion, time varying analysis, Optic Flow, recursive filtering, closed loop control, robotics.

1. Introduction

Un récent développement en vision par ordinateur et analyse de scène concerne l'utilisation explicite du mouvement pour estimer la structure géométrique et cinématique des primitives 3D contenues dans la scène. Ce mouvement peut être introduit dans les algorithmes tant dans le domaine de l'image (champ des vitesses apparentes dans l'image, connu sous le nom d'*Optic Flow*, résultant de la projection du champ des vitesses dans la scène), que dans le domaine de la scène (mouvements dus à la caméra et/ou aux objets mobiles dans la scène).

L'estimation du mouvement dans l'image ou, plus exactement dans la séquence d'images, peut être utilisée dans des buts de segmentation ou pour accéder à des informations 3D au niveau de la scène. Dans ce dernier cas des contraintes supplémentaires sur celle-ci doivent être imposées. Pour estimer le champ des vitesses dans l'image, deux classes d'approches ont été mises en œuvre : des techniques de mises en correspondances d'éléments caractéristiques dans la suite d'images (*features based methods*) comme dans [1] à [4] et des techniques différentielles (*gradient based methods*) comme décrites dans [5] à [9]. Pour exploiter cette estimation du mouvement dans l'identification des structures géométriques dans la scène, il est nécessaire d'établir un modèle des primitives utilisées (points, courbes, surfaces...) non seulement géométrique mais aussi cinématique et d'étudier les propriétés de la projection de ce modèle dans la séquence d'images [10, 11, 12].

L'analyse du mouvement peut se révéler très utile dans de nombreuses applications, principalement en robotique : stéréovision [13, 4], identification de trajectoire dans le cas d'un robot mobile ou d'une cible [10, 14], contrôle adaptatif de robots [15] et certaines tâches spécifiques de robotique comme, par exemple, la navigation et l'évitement d'obstacles [16].

Cet article présente un schéma général d'estimation de primitives 3D basé sur l'utilisation explicite du mouvement à la fois dans la séquence d'images et dans la scène. Ce schéma est appliqué à l'estimation de la profondeur relative caméra-objets (*carte des*

profondeurs) puis étendu à des primitives 3D de type segments. Nous restreindrons l'étude au cas où le capteur est en mouvement dans un univers statique.

L'idée de base de ce schéma est d'utiliser des mesures instantanées de paramètres 2D dans la séquence d'images pour raffiner l'estimation sur les primitives 3D au fur et à mesure du déplacement du capteur dans la scène. De plus, comme nous contrôlons ce déplacement, cela nous permettra de l'optimiser en fonction de la trajectoire nominale désirée dans la tâche principale (par exemple, trajectoire de référence) et du mouvement entraînant une meilleure estimation de la primitive 3D. C'est une approche de commande dans laquelle la fermeture de la boucle permet de réaliser un compromis entre ces deux objectifs.

L'article est organisé comme suit : la Section II est consacrée à l'établissement des modèles cinématiques reliant l'aspect géométrique des primitives 3D aux mouvements de la caméra. La Section III est dédiée à l'extraction des paramètres dynamiques dans la séquence d'images. Dans la Section IV, nous décrivons l'aspect estimation récursive de la primitive 3D tout au long de la séquence d'images. La Section V est consacrée à l'élaboration de la commande en boucle fermée du robot vis-à-vis de son environnement. Enfin, la Section VI présentera les résultats obtenus en simulation dans le cadre d'une tâche d'évitement d'obstacles. Par ailleurs, l'Annexe A donne une liste des notations employées.

2. Modélisation des primitives

2.1. HYPOTHÈSES DE BASES

Nous considérons dans un espace 3D, un ensemble d'objets statiques. Un repère de référence R_s est associé à la scène 3D, un repère R_c est lié à la caméra mobile dont le modèle est du type projection centrale et la distance focale choisie égale à un (*fig. 1*).

Les coordonnées du point P dans R_s sont :

$$(1) \quad X_s^T = (X_s \ Y_s \ Z_s)$$

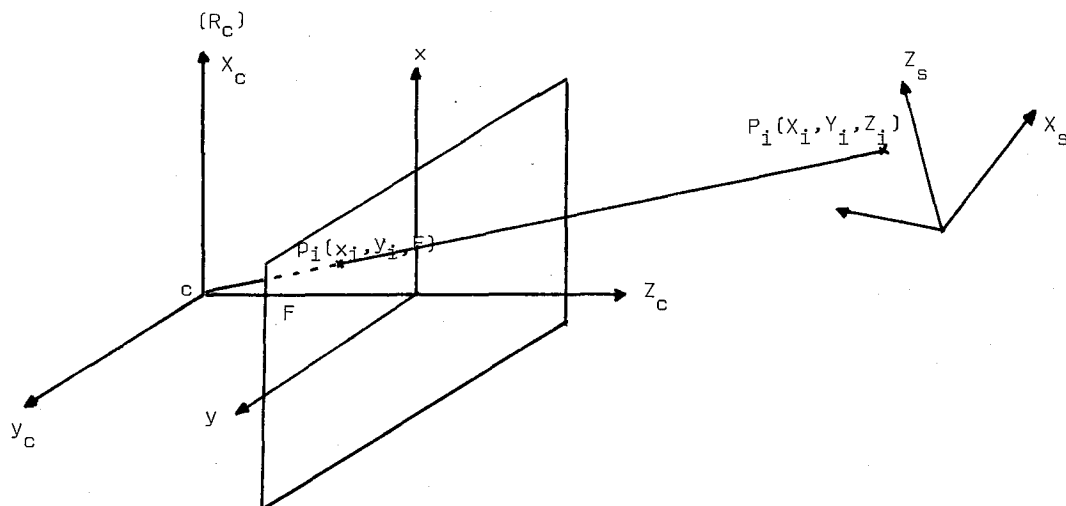


Fig. 1. — Modèle de la caméra.

et, dans R_c :

$$X^T = (X, Y, Z), \quad Z > 1$$

Sa projection dans le plan image est, en utilisant les coordonnées homogènes :

$$(2) \quad \begin{pmatrix} x' \\ y' \\ z' \\ \omega \end{pmatrix} = \begin{pmatrix} & & & 0 \\ & I_3 & & 0 \\ & & & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

avec

$$(3) \quad \mathbf{x}^P = \frac{1}{\omega} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$$

nous obtenons donc les coordonnées images (avec $F=1$)

$$(4) \quad \begin{cases} x = \frac{X}{Z} \\ y = \frac{Y}{Z} \end{cases}$$

Le mouvement de la caméra est complètement décrit par le torseur cinématique associé τ , qui sera exprimé dans le repère R_c et noté :

$$(5) \quad \tau = \begin{pmatrix} -\mathbf{V} \\ -\boldsymbol{\Omega} \end{pmatrix}$$

avec

$$(7) \quad \mathbf{V}^T = (V_x \ V_y \ V_z)$$

$$(8) \quad \boldsymbol{\Omega}^T = (\Omega_x \ \Omega_y \ \Omega_z)$$

ce qui permet d'exprimer la vitesse relative de n'importe quel point P dans R_c par :

$$(9) \quad \frac{d\mathbf{X}}{dt} = \dot{\mathbf{X}} = \mathbf{V} + \boldsymbol{\Omega} \times \mathbf{X}$$

où \times représente le produit vectoriel.

2.2. RELATIONS ENTRE LES MODÈLES 3D ET LES PARAMÈTRES DYNAMIQUES DANS LA SÉQUENCE D'IMAGES

L'estimation de la primitive 3D étant réalisée à partir de mesures dans la séquence d'images, il est nécessaire d'établir les équations reliant les paramètres de la primitive à des paramètres (2D+t) extraits de la séquence d'images. Plus précisément, nous allons établir les équations différentielles donnant l'évolution des paramètres représentant la primitive au cours du mouvement de la caméra, ainsi que les expressions des variations correspondantes de la projection de la primitive dans le repère image. Nous montrerons alors que mesurant ces variations dans l'image et connaissant le mouvement du capteur, nous pouvons remonter aux valeurs des paramètres de la primitive 3D. Ce principe, d'application simple dans le cas de primitives de type « points », sera également employé pour les primitives de type « segments ».

2.2.1. Cas d'une primitive de type « point »

Les paramètres de la primitive 3D de type « point », sont les coordonnées 3D du point dans le repère R_c . Les équations reliant ces coordonnées à des mesures dans le repère image sont obtenues en dérivant les équations (4) et (5) et en utilisant l'équation (9) soit :

$$(10) \quad v = \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} : \begin{cases} \dot{x} = u_T + u_r \\ \dot{y} = v_T + v_r \end{cases}$$

avec

$$(12) \quad \begin{cases} u_T = \frac{1}{Z} (-V_x + x V_z) \\ v_T = \frac{1}{Z} (-V_y + y V_z) \end{cases}$$

$$(13) \quad \begin{cases} u_r = y \Omega_z + x y \Omega_x - (1 + x^2) \Omega_y \\ v_r = -x \Omega_z - x y \Omega_y + (1 + y^2) \Omega_x \end{cases}$$

A partir de ces expressions, il apparaît que la connaissance de τ peut permettre de calculer les coordonnées 3D (X, Y, Z) d'un point P dans le repère R_c , si on peut mesurer dans l'image une forme linéaire :

$$(16) \quad v^* = \lambda_1 \dot{x} + \lambda_2 \dot{y}$$

où λ_1 et λ_2 sont connus.

Une condition supplémentaire, due aux équations (12) et (13) est naturellement :

$$(17) \quad \|\mathbf{V}\| \neq 0$$

Dans le paragraphe 3.1.1 nous montrerons comment estimer v^* qui représente le vecteur des vitesses apparentes dans l'image (*Optic Flow*).

2.2.2. Cas d'une primitive de type « segment »

Considérons un segment de droite \mathcal{D} dans R_c , défini par son équation paramétrique :

$$(18) \quad \mathbf{X} = \mathbf{X}_0 + \xi \mathbf{u}, \quad \forall X \in \mathcal{D}$$

ou, de façon équivalente, par :

$$(19) \quad (\mathbf{X} - \mathbf{X}_0) \times \mathbf{u} = 0$$

La représentation de \mathcal{D} par \mathbf{X}_0 et le vecteur directeur \mathbf{u} est redondante. Pour établir le modèle cinématique d'un segment en mouvement, nous choisissons de le représenter par les coordonnées pluckeriennes (\mathbf{u}, \mathbf{h}) , où :

$$(20) \quad \mathbf{h} = \mathbf{X}_0 \times \mathbf{u}$$

avec

$$(21) \quad \mathbf{u}^T = (\alpha \ \beta \ \gamma), \quad \|\mathbf{u}\| = 1$$

Les composantes de \mathbf{h} (h_1, h_2, h_3) sont également liées par l'équation :

$$(22) \quad \langle \mathbf{h}, \mathbf{u} \rangle = 0$$

Il est alors possible d'établir le modèle cinématique dans R_c en différentiant l'équation (19) par rapport

au temps ce qui conduit à la relation :

$$(23) \quad \dot{\mathbf{X}} \times \mathbf{u} - \dot{\mathbf{X}}_0 \times \mathbf{u} - \dot{\mathbf{u}} \times (\mathbf{X} - \mathbf{X}_0) = 0$$

en utilisant les équations (9) et (18), nous obtenons :

$$(24) \quad (\mathbf{V} - \dot{\mathbf{X}}_0 + \boldsymbol{\Omega} \times \mathbf{X}_0) \times \mathbf{u} + \xi (\boldsymbol{\Omega} \times \mathbf{u} - \dot{\mathbf{u}}) \times \mathbf{u} = 0$$

Comme u est un vecteur unitaire, nous obtenons finalement :

$$(25) \quad \begin{cases} \dot{\mathbf{u}} = \boldsymbol{\Omega} \times \mathbf{u} \\ (\mathbf{V} + \boldsymbol{\Omega} \times \mathbf{X}_0 - \dot{\mathbf{X}}_0) \times \mathbf{u} = 0 \end{cases}$$

$$(26)$$

L'équation (26) peut être réécrite :

$$(27) \quad \mathbf{V} \times \mathbf{u} - \dot{\mathbf{X}}_0 \times \mathbf{u} + (\boldsymbol{\Omega} \times \mathbf{X}_0) \times \mathbf{u} = 0$$

la dérivation de l'équation (20) donne :

$$(28) \quad \dot{\mathbf{h}} = \dot{\mathbf{X}}_0 \times \mathbf{u} + \mathbf{X}_0 \times \dot{\mathbf{u}}$$

qui, une fois substituée dans l'équation (27) et en utilisant l'équation (25), amène à :

$$(29) \quad \dot{\mathbf{h}} = \mathbf{V} \times \mathbf{u} + \mathbf{X}_0 \times (\boldsymbol{\Omega} \times \mathbf{u}) + (\boldsymbol{\Omega} \times \mathbf{X}_0) \times \mathbf{u}$$

soit, enfin :

$$(30) \quad \dot{\mathbf{h}} = \mathbf{V} \times \mathbf{u} - \mathbf{h} \times \boldsymbol{\Omega}$$

Cette équation nous permet d'extraire deux équations indépendantes en utilisant la contrainte fournie par l'équation (22). Nous avons donc établi quatre équations différentielles indépendantes (25) et (30) décrivant complètement la variation des paramètres représentant la droite \mathcal{D} due au mouvement relatif de R_c par rapport à R_s .

Les calculs précédents ont permis d'établir un modèle cinématique de primitive 3D de type « segment ». Il est nécessaire de relier les paramètres de ce modèle à des mesures dans la séquence d'images

La projection de \mathcal{D} dans le repère image est une droite D d'équation obtenue en utilisant (4) et (5) et en éliminant Z dans deux équations fournies par (19). Lorsque la projection D de \mathcal{D} , ne passe pas par l'origine ($x=0; y=0$), c'est-à-dire lorsque $h_3 \neq 0$, son équation est de la forme générale :

$$(31) \quad px + 1 = qy$$

où p et q sont donnés par :

$$(32) \quad \begin{cases} ph_2 + qh_1 = 0 \\ \beta q - \alpha p = \gamma \end{cases}$$

$$(33)$$

Dans le cas contraire, l'équation de D est :

$$(34) \quad p'x = q'y$$

avec :

$$(35) \quad \begin{cases} p' = \beta \\ q' = \alpha \\ h_3 = 0 \end{cases}$$

$$(36)$$

$$(37)$$

Nous allons démontrer maintenant que si nous pouvons estimer (p, q, \dot{p}, \dot{q}), alors il est possible de calcu-

ler les paramètres (\mathbf{u}, \mathbf{h}) qui caractérisent \mathcal{D} ; nous traiterons dans la suite uniquement le cas général où h_3 est différent de zéro.

Pour faciliter l'écriture, posons :

$$(38) \quad \mathbf{s} = (-p \quad q \quad -1)^T$$

ce qui permet d'écrire l'équation (33) sous la forme :

$$(39) \quad \mathbf{s}^T \mathbf{u} = 0$$

En différentiant l'équation (39) et en utilisant l'équation (25), on obtient :

$$(40) \quad \mathbf{s}^T (\boldsymbol{\Omega} \times \mathbf{u}) + \dot{\mathbf{s}}^T \mathbf{u} = 0$$

qui peut être écrite (en appelant ω la matrice antisymétrique associée au produit vectoriel \times) :

$$(41) \quad \begin{pmatrix} \mathbf{s}^T \boldsymbol{\omega} - \dot{\mathbf{s}}^T \\ \mathbf{s}^T \end{pmatrix} \mathbf{u} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ -p & q & -1 \end{pmatrix} \mathbf{u} = 0$$

avec

$$(42) \quad \begin{cases} a_{11} = -q\Omega_z - \Omega_y + \dot{p} \\ a_{12} = -p\Omega_z + \Omega_x - \dot{q} \\ a_{13} = p\Omega_y + q\Omega_x \end{cases}$$

$$(43)$$

$$(44)$$

En posant

$$(45) \quad \mathbf{A} = qa_{11} + pa_{12}$$

et

$$(46) \quad \mathbf{F} = qa_{13} - p$$

l'équation (41) est alors équivalente à

$$(47) \quad \begin{pmatrix} \mathbf{A} & 0 & \mathbf{F} \\ -p & q & -1 \end{pmatrix} \mathbf{u} = 0$$

$$(48)$$

qui donne

$$(49) \quad \alpha = \frac{\mathbf{F}}{\mathbf{A}} \gamma$$

et

$$(50) \quad \beta = \frac{p\mathbf{F} + \mathbf{A}}{q\mathbf{A}} \gamma$$

Finalement, l'expression des deux premières composantes de \mathbf{u} s'écrit :

$$(51) \quad \alpha = \frac{q\mathbf{A}\mathbf{F}}{\sqrt{q^2\mathbf{F}^2(\mathbf{A}^2 + \mathbf{F}^2) + \mathbf{A}^2(p\mathbf{F} + \mathbf{A})^2}}$$

et

$$(52) \quad \beta = \frac{p\mathbf{F} + \mathbf{A}}{q\mathbf{F}} \alpha$$

Nous allons appliquer une démarche identique pour exprimer les composantes de \mathbf{h} en fonction de (p, q, \dot{p}, \dot{q}).

Posons

$$(53) \quad \tau = (q \ p \ 0)^T$$

alors, l'équation (32) devient :

$$(54) \quad \tau^T \mathbf{h} = 0$$

En différenciant l'équation (54) et en utilisant l'équation (30), on obtient :

$$(55) \quad (\dot{\tau}^T - \tau^T \omega) \mathbf{h} + \tau^T \mathbf{V} \times \mathbf{u} = 0$$

En utilisant la propriété

$$(56) \quad \tau^T \mathbf{V} \times \mathbf{u} = \gamma \mathbf{s}^T \mathbf{V}$$

alors, les équations (22), (54) et (55) forment le système suivant :

$$(57) \quad \{b_{ij}\} \mathbf{h} = \begin{pmatrix} \tau^T \omega - \dot{\tau}^T \\ \tau^T \\ u^T \end{pmatrix} \mathbf{h} = \begin{pmatrix} \gamma \mathbf{s}^T \mathbf{V} \\ 0 \\ 0 \end{pmatrix}$$

avec $\mathbf{s}^T \mathbf{V} \neq 0$ pour une solution non nulle [condition analogue à celle donnée par l'équation (17)].

En remarquant que

$$(58) \quad b_{11} = a_{12} - \Omega_x$$

$$(59) \quad b_{12} = -a_{11} - \Omega_y$$

$$(60) \quad b_{13} = -q \Omega_y + p \Omega_x$$

et en utilisant l'équation (45), le système s'exprimera :

$$(61) \quad \begin{pmatrix} 0 & A - b_{13} & -qb_{13} \\ q & p & 0 \\ \alpha & \beta & \gamma \end{pmatrix} \mathbf{h} = \begin{pmatrix} -q \gamma \mathbf{s}^T \mathbf{V} \\ 0 \\ 0 \end{pmatrix}$$

dont la solution est

$$(62) \quad h_2 = \frac{q \alpha}{F} \mathbf{s}^T \mathbf{V}$$

avec h_1 obtenu par l'équation (54) et h_2 par l'équation (22).

Remarque

Dans le cas particulier d'une translation pure du capteur, le système complet définissant (\mathbf{u}, \mathbf{h}) sera :

$$(63) \quad \left\{ \begin{array}{l} \frac{\alpha}{\beta} = \frac{\dot{q}}{p} \\ \alpha = \frac{p - q(\dot{p}/\dot{q})}{\sqrt{1 + (1 + (p/q)^2)(p - q(\dot{p}/\dot{q}))^2}} \\ h_2 = \frac{q}{q} \alpha (-pV_x + qV_y - V_z) \\ h_1 = -\frac{p}{q} h_2 \end{array} \right.$$

3. Extraction des mesures dans la séquence d'images

Nous venons d'établir les équations reliant le mouvement dans le plan image aux paramètres caractérisant des primitives 3D de type « points » et « segments », cette section est consacrée à l'extraction de ce mouvement dans la séquence d'images. Dans le cas de primitives « points », nous utiliserons une approche de type *Optic Flow* et dans le cas des segments, une approche de type *approximation polygonale*.

3.1. CAS DES PRIMITIVES « POINTS » : Estimation du mouvement dans l'image

Le champ des vitesses apparentes dans l'image (*Optic Flow*) est défini comme le champ de vecteurs mettant en correspondance les pixels entre deux images prises à des instants différents et après déplacement du capteur et/ou des objets dans la scène. Dans la pratique, en imagerie numérique, on assimilera la mesure du champ de vitesse à une mesure de déplacement.

Une hypothèse couramment utilisée consiste à faire l'hypothèse que la variation de l'intensité dans la scène est uniquement due au mouvement de la caméra, ce qui permet d'écrire l'équation bien connue liant le gradient spatial et les composantes de l'Optic Flow dans l'image :

$$(67) \quad \frac{dI}{dt} = 0$$

$$(68) \quad \langle \nabla I, \mathbf{v} \rangle + \frac{\partial I}{\partial t} = 0$$

ou $\langle . \rangle$ représente le produit scalaire.

Cette équation montre que, localement, seule la composante v^\perp orthogonale au contour dans l'image est mesurable. Cependant, si nous sommes capables d'estimer à la fois v^\perp et la direction du gradient spatial, alors nous pourrions déterminer (λ_1, λ_2) dans l'équation (16) et le calcul de la carte des profondeurs deviendra possible.

Parmi les nombreuses méthodes citées en référence, permettant d'extraire la composante orthogonale de l'optic flow, l'algorithme décrit dans [8] est un des plus efficaces. A partir d'une modélisation spatio-temporelle d'un élément de contour en mouvement, une technique de maximum de vraisemblance discret permet de déterminer un ensemble de masques qui constitue un EOF (Estimateur d'Optic Flow). Pour chaque pixel appartenant à un contour, cet opérateur fournit l'estimation de (fig. 2) :

l'orientation spatiale du contour, $\tilde{\theta}$;

la composante orthogonale au contour de l'optic flow, \tilde{v}^\perp , avec la valeur de vraisemblance associée.

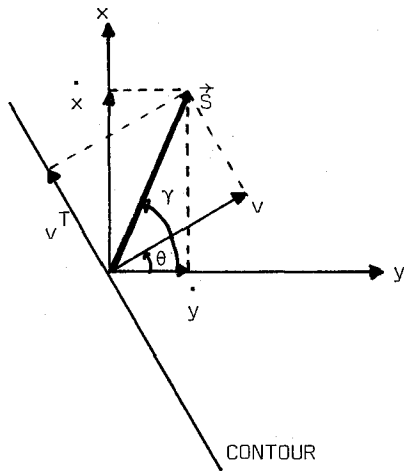
L'équation (16) devient alors :

$$(69) \quad \tilde{v}^\perp = \tilde{\lambda}_1 \dot{y} + \tilde{\lambda}_2 \dot{x}$$

avec

$$(70) \quad \tilde{\lambda}_1 = \cos \tilde{\theta}$$

$$(71) \quad \tilde{\lambda}_2 = \sin \tilde{\theta}$$



Nous utiliserons cet opérateur dans le paragraphe suivant.

3.2. CAS DES PRIMITIVES « SEGMENTS » : Estimation des segments dans l'image

L'estimation de segments dans l'image peut être décomposée en deux phases: la détection des contours spatiaux et leur approximation polygonale.

Concernant l'extraction de contours, de nombreuses méthodes peuvent être utilisées comme les techniques de gradient ou de laplacien [17, 18], ou de filtrage récursif [19, 20]; dans de nombreux cas, des étapes d'amincissement et de chaînage seront nécessaires pour obtenir des contours de longueur maximale.

La deuxième étape consiste à approximer le contour par des segments de droites. Le principe de base consiste à chercher la meilleure approximation d'un nuage de points par un segment de droite ce qui est équivalent à minimiser :

$$(72) \quad \hat{d}^2 = \min_{P, q} \left\{ d^2 = \sum_{i=1}^n d_i^2 \right\}$$

avec n égal au nombre de pixels et d_i la « distance » du pixel i à la droite cherchée, soit dans le cas d'une distance euclidienne :

$$(73) \quad d_i^2 = \frac{(px_i - qy_i + 1)^2}{p^2 + q^2}$$

La précision de l'approximation est alors donnée par la valeur du résidu.

Cette approche peut être directement étendue à l'approximation d'une courbe en utilisant soit un algorithme itératif le long de la chaîne de contours, soit une méthode globale permettant de diviser récursivement la chaîne de contour à partir de ses extrémités (« divide and conquer » [21]). Dans le cas d'une approche itérative, il est possible d'utiliser d'autres critères de distances comme, par exemple, le maximum de courbure.

4. Estimation récursive le long de la séquence

Du fait des perturbations, des erreurs de modélisation, des bruits de mesure (discrétisation, quantification), les primitives 3D reconstruites à partir de seulement un couple d'images sont souvent très bruitées. Une idée intéressante consiste à utiliser la cohérence temporelle pour affiner l'estimée au long de la séquence. La formulation du problème en termes de filtrage récursif apparaît alors naturelle.

4.1. CAS DES PRIMITIVES « POINTS »

4.1.1. Cas général

Le vecteur d'état est constitué par les coordonnées 3D des points dans le repère R_s et l'équation d'état (9) peut être réécrite sous la forme :

$$(74) \quad \dot{\mathbf{X}} = \mathbf{V}(t) + \omega(t) \mathbf{X}$$

où $\omega(t)$ est la matrice antisymétrique associée au produit vectoriel.

Les équations de mesures proviennent des équations (4), (5) et (16) et prennent la forme :

$$(75) \quad \begin{cases} \tilde{x} = \frac{X}{Z} + e_x(t) \\ \tilde{y} = \frac{Y}{Z} + e_y(t) \end{cases}$$

$$(77) \quad \tilde{v}^\perp = v^\perp + \varepsilon_v(t) = \lambda_1 (u_T + v_T) + \lambda_2 (u_r + v_r) + \varepsilon_v(t)$$

avec

$$(79) \quad u_r = \frac{1}{Z^2} (YZ \Omega_z + XY \Omega_x - (X^2 + Z^2) \Omega_y)$$

$$(80) \quad v_T = \frac{1}{Z} \left(-V_Y + \frac{Y}{Z} V_Z \right)$$

$$(81) \quad v_r = \frac{1}{Z^2} (-XZ \Omega_z - XY \Omega_y + (Y^2 + Z^2) \Omega_x)$$

Plusieurs problèmes apparaissent à ce niveau concernant :

– *les perturbations* : Le bruit d'état est dû à l'erreur de mesure sur le mouvement. Bien qu'il soit de la forme suivante :

$$(82) \quad \begin{cases} \tilde{\mathbf{V}}(t) = \mathbf{V}(t) + \varepsilon_v(t) \\ \tilde{\mathbf{\Omega}}(t) = \mathbf{\Omega}(t) + \varepsilon_\omega(t) \end{cases}$$

cela entraîne un bruit multiplicatif au niveau de l'équation (74).

De plus, les bruits sur les équations de mesure apparaissent également de façon non linéaire: $\tilde{\lambda}_1$ et $\tilde{\lambda}_2$ dépendent de l'estimation de l'orientation du contour fournie par l'EOF qui est bruitée.

– *la discrétisation* : Le passage du modèle continu au modèle discret utilisé en pratique pose des problèmes au niveau des rotations : choix de la représentation des orientations, transformation des équations (79) et (81) en termes de vecteurs de rotations discrets.

— les non-linéarités : Une autre difficulté importante vient du fait que toutes les équations de mesures sont non linéaires vis-à-vis des variables d'états :

$$(84) \quad z = f(\mathbf{X}, \mathbf{V}, \boldsymbol{\Omega}) + e(t).$$

La technique largement utilisée pour contourner cette difficulté consiste à linéariser à chaque instant les équations autour de la valeur $\mathbf{X}^*(t)$ la plus proche possible de la valeur réelle, soit :

$$(85) \quad \tilde{z} = f(\mathbf{X}^*(t), \mathbf{V}, \boldsymbol{\Omega}) + \frac{\partial f}{\partial \mathbf{X}}(\mathbf{X}^*(t), \mathbf{V}, \boldsymbol{\Omega})(\mathbf{X} - \mathbf{X}^*(t)) + \dots$$

Dans le cas discret, l'algorithme le plus classique est connue sous le nom de *filtrage de Kalman étendu*, où $\mathbf{X}^*(t)$ est la *prédiction* de $\mathbf{X}(t)$ au temps $t = t_k$. Alors, les méthodes classiques de filtrage peuvent être employées sur l'équation (84). Cependant, une telle approche peut entraîner des problèmes au niveau de la convergence.

Pour toutes ces raisons, les algorithmes d'estimation récursive seront généralement approximatés. Dans le paragraphe qui suit, nous étudierons le cas plus simple d'un capteur se déplaçant uniquement en translation.

4.1.2. Mouvement de translation

Le passage en discret est direct dans ce cas et nous pouvons écrire, pour chaque instant k et pour chaque point \mathbf{X} , l'équation d'état :

$$(86) \quad \mathbf{X}(k+1) = \mathbf{X}(k) + \mathbf{V}(k) + \varepsilon_v(k+1)$$

où nous supposons que ε_v est un bruit blanc centré de variance $q_v \mathbf{I}$.

En addition aux équations (75) et (76), l'équation principale de mesure (69) est fournie par l'EOF et il est impossible d'éviter les problèmes dus aux non-linéarités et aux bruits. En vue de rendre possible les calculs, plusieurs approximations sont requises.

L'équation (69) dans le cas d'un mouvement de translation, s'écrit :

$$(87) \quad \tilde{v}^\pm = \frac{1}{Z} \left(-(\tilde{\lambda}_1 \tilde{V}_x + \tilde{\lambda}_2 \tilde{V}_y) + \left(\tilde{\lambda}_1 \frac{X}{Z} + \tilde{\lambda}_2 \frac{Y}{Z} \right) \tilde{V}_z \right) + e_z$$

où le « tilde » représente le caractère bruité de la variable. Une première approximation consiste à considérer une absence de bruit sur les équations (75) et (76). Cela permet de remplacer X/Z et Y/Z dans l'équation (87) par les valeurs « exactes » x et y , et donc de réduire la dimension du problème à l'estimation unique de Z . La seconde approximation est destinée à linéariser l'équation de mesure, nous proposons deux façons de réaliser cela :

— Soit linéariser l'équation (87) autour d'un point donné Z^* ; on obtient alors :

$$(88) \quad \begin{cases} \tilde{v}^\pm \simeq \frac{a}{Z^{*2}}(2Z^* - Z) + e_v \\ a = -(\tilde{\lambda}_1 \tilde{V}_x + \tilde{\lambda}_2 \tilde{V}_y) + (\tilde{\lambda}_1 x + \tilde{\lambda}_2 y) \tilde{V}_z. \end{cases}$$

— Soit appliquer directement l'équation de mesures

$$(89) \quad \tilde{Z} = \frac{a}{\tilde{v}^\pm} Z + e_z.$$

Dans les deux cas e_v et e_z sont des termes globaux de bruit prenant en compte les contributions dues à la linéarisation et aux approximations. Il est nécessaire de considérer que leur covariance n'est pas constante (elle dépend même de l'état), en pratique, de la borner. Dans la suite, nous traiterons le cas de l'approximation (89) et les équations seront alors :

$$(90) \quad \begin{cases} Z(k+1) = Z(k) + V_z(k) + \varepsilon_z(k+1) \\ \tilde{Z}(k+1) = Z(k+1) + e_z(k+1) \end{cases}$$

$$(91) \quad \begin{cases} Z(k+1) = Z(k) + V_z(k) + \varepsilon_z(k+1) \\ \tilde{Z}(k+1) = Z(k+1) + e_z(k+1) \end{cases}$$

où \tilde{Z} est calculé à chaque instant en utilisant l'équation (89), et en supposant que : ε_z est centré, indépendant de variance q et $e_z(k)$ est centré, indépendant de variance $r(k)$ où k est le numéro de l'image dans la séquence.

Les équations de filtrage associées sont les équations bien connues du filtre de Kalman, dans lesquelles on peut introduire, si nécessaire, un facteur d'oubli exponentiel. La forme générale de l'équation d'estimation pour Z est alors :

$$(92) \quad \hat{Z}(k+1) = \hat{Z}(k) + V_z(k) + K(k+1)(\tilde{Z}(k+1) - V_z(k) - \hat{Z}(k))$$

où $K(k+1)$ est le gain de Kalman.

Remarque

L'approche présentée ci-dessus concerne uniquement un même point suivi au long de la séquence, nous montrerons par la suite comment l'étendre à un ensemble de points.

4.2. CAS DES PRIMITIVES « SEGMENTS »

Nous réduirons l'étude au cas des translations. Pour un segment de droite, l'état du système est fourni par les versions discrètes des équations (25) et (30) avec ($\Omega=0$). Cela nous procure quatre équations dont deux sont linéaires et deux non linéaires avec le vecteur d'état suivant :

$$(93) \quad \mathbf{R}(k) = (\alpha(k) \beta(k) h_1(k) h_2(k))^T$$

Les variables de mesures à chaque instant sont $p(k)$ et $q(k)$. Les équations de mesures (32) et (33) sont non linéaires et il est nécessaire de les linéariser avant d'appliquer le filtre récursif. Cependant $p(k)$ et $q(k)$ sont évalués grâce à une estimation *spatiale*. Si le segment est assez long et si la dispersion des points le constituant n'est pas très importante, l'erreur résultante sur le segment sera plus faible que dans le cas de l'estimation des points. Aussi, l'utilisation directe des équations (63) et (64) en tant qu'équations de mesure est possible. L'algorithme pourra alors être de la forme :

(i) A partir des mesures $\tilde{\alpha}(k)$ et de $\tilde{\beta}(k)$ fournies par les équations (63) et (64), mettre à jour l'état avec un

filtre linéaire portant sur l'équation :

$$(94) \quad \begin{pmatrix} \alpha(k+1) \\ \beta(k+1) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha(k) \\ \beta(k) \end{pmatrix} + \text{bruits.}$$

(ii) Partant des estimées courantes $\hat{\alpha}(k)$ et $\hat{\beta}(k)$, calculer $\hat{\gamma}(k)$.

(iii) Mettre à jour $(\hat{h}_1(k), \hat{h}_2(k))$ en utilisant les mesures $(\tilde{h}_1(k), \tilde{h}_2(k))$ et les équations d'état approximées :

$$(95) \quad \begin{pmatrix} h_1(k+1) \\ h_2(k+1) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} h_1(k) \\ h_2(k) \end{pmatrix} + \hat{\gamma}(k) \begin{pmatrix} V_Y(k) \\ -V_X(k) \end{pmatrix} + V_Z(k) \begin{pmatrix} -\hat{\beta}(k) \\ \hat{\alpha}(k) \end{pmatrix}.$$

5. Élaboration de la commande en boucle fermée

Une constatation évidente est que la qualité de l'estimation récursive au long de la séquence d'images est très sensible aux mouvements successifs du capteur (amplitude et direction). De façon plus précise, la nature de ce mouvement affecte les erreurs de mesure dans l'image : par exemple, le mouvement du capteur qui produira le glissement d'un segment de l'image sur lui-même n'induirait aucune information sur la structure du segment 3 D correspondant dans la scène.

Dans le cas des robots mobiles ou des manipulateurs porteurs de caméra, il est possible de contrôler le mouvement de la caméra (bien que la mesure de ce mouvement soit bruitée). Nous allons utiliser cette possibilité pour réaliser un schéma de commande en boucle fermée dans lequel nous allons superposer au mouvement « nominal » requis par la tâche, un mouvement destiné à accroître la qualité des estimations. Cette idée est, en un certain sens, proche du concept d'*extra-signal* utilisé dans certaines méthodes d'identification. Nous allons décrire cette approche dans le cas des « points » et des « segments ».

5. 1. CAS DES POINTS

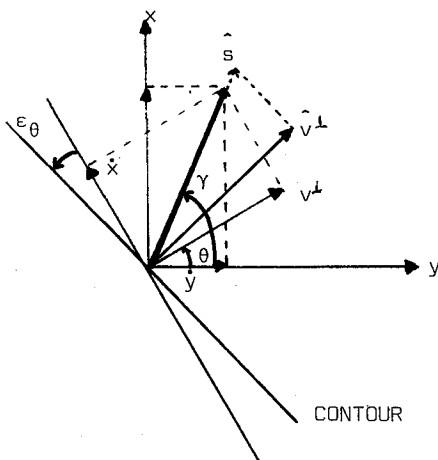


Fig. 3. — Erreurs de mesures introduites par l'EOF.

La figure 3 représente la modélisation des erreurs de mesures introduites par l'EOF, soit :

$$(96) \quad \tilde{\theta} = \theta + \varepsilon_\theta$$

$$(97) \quad \tilde{v}^\perp = \frac{v^\perp}{\cos \varepsilon_\theta} + \varepsilon_v$$

où ε_θ correspond au bruit de discrétisation sur l'orientation du contour et ε_v le bruit de quantification sur $\|\tilde{v}^\perp\|$.

Dans la pratique, ε_θ est uniquement fonction du jeu de masques utilisés dans l'EOF, il peut être borné et choisi petit :

$$(98) \quad \cos \varepsilon_\theta \simeq 1$$

$$(99) \quad \sin \varepsilon_\theta \simeq \varepsilon_\theta$$

ce qui a pour conséquence que :

$$(100) \quad \tilde{\lambda}_1 = \lambda_1 - \lambda_2 \varepsilon_\theta$$

$$(101) \quad \tilde{\lambda}_2 = \lambda_2 + \lambda_1 \varepsilon_\theta$$

Modélisons les autres erreurs de mesures (mouvement et coordonnées pixel) sous la forme additive :

$$(102) \quad \tilde{V} = V + \varepsilon_v$$

$$(103) \quad \tilde{x} = x + e_x$$

$$(104) \quad \tilde{y} = y + e_y$$

alors, l'approximation (89) de l'erreur relative sur \tilde{Z} prend la forme :

$$(105) \quad \frac{\tilde{Z} - Z}{Z} \simeq \varepsilon_\theta \frac{v^T}{v^\perp} - \frac{\varepsilon_v}{v^\perp} + \frac{h'}{Z v^\perp} + \text{termes du 2-ième ordre} \dots$$

avec

$$(106) \quad h' = -(\varepsilon_x \lambda_2 + \varepsilon_y \lambda_1) + \varepsilon_z (\lambda_1 y + \lambda_2 x) + V_z (\lambda_1 \varepsilon_x + \lambda_2 \varepsilon_y).$$

Cette erreur sera réduite si :

— ε_θ est petit : Ce bruit est lié uniquement à l'EOF, il ne dépend pas du mouvement et peut être réduit en diminuant le pas de quantification des masques utilisés dans l'EOF.

— h' est petit : Cela nécessite de minimiser ε_v , donc d'avoir la meilleure mesure possible du mouvement. Du point de vue de la commande en boucle fermée, il faut également que V_z soit le plus petit possible.

— $1/v^\perp$ et v^T/v^\perp sont petits : Cette condition traduit que localement, le vecteur d'optic flow doit être orthogonal au contour avec une amplitude maximale.

En tenant compte de ces contraintes et en gardant à l'esprit le fait que nous devons rester le plus proche possible de la trajectoire nominale $V^N(k)$ imposée par la tâche, nous allons poser le problème en termes de minimisation.

L'optic flow étant estimé entre l'image $(k-1)$ et l'image (k) , le but est de trouver le mouvement « optimal » du capteur entraînant la meilleure mesure d'optic flow entre (k) et $(k+1)$. Nous supposons le suivi des points au long de la séquence résolu et que

chacun d'eux correspond à un même point 3D dans la scène :

$$(107) \quad \begin{cases} \mathbf{p}_i(k-1) = (x_i(k-1) y_i(k-1))^T \\ i = 1 \dots N(k-1) \end{cases}$$

A chaque point $p_i(k-1)$ nous associons :

- une position estimée $\tilde{p}_i(k)$ dans l'image k ;
- la mesure fournie par l'EOF au moyen des équations (96) et (97) $\{\tilde{v}_i^+(k-1); \tilde{\theta}_i(k-1)\}$;
- un déplacement local désiré de $p_i(k)$ entre l'image k et l'image $k+1$ appelé par la suite $v_a^i(k)$;
- un paramètre de pondération $\sigma^i(k)$.

Posons :

$$(108) \quad \mathbf{E}(k) = ((v_a(k) - \tilde{v}^1(k)) \times \dots \times (v_a^N(k) - \tilde{v}^N(k)))^T$$

où $\tilde{v}^i(k)$ est la prédiction du déplacement local de $\tilde{p}^i(k)$ de l'image k à l'image $k+1$ [calculée à partir de $\tilde{x}_i(k)$, $\tilde{y}_i(k)$ et de $\tilde{Z}_i(k)$ l'estimation de $Z_i(k)$ fournie par le filtre récursif décrit précédemment], qui dépend explicitement de $\mathbf{V}(k)$, le mouvement du capteur entre k et $k+1$ qui doit être optimisé.

Notons

$$(109) \quad \mathbf{W}(k) = \begin{pmatrix} \sigma_{(k)}^1 & & 0 \\ & \dots & \\ 0 & & \sigma_{(k)}^N \end{pmatrix},$$

la fonction de coût à minimiser sera finalement :

$$(110) \quad J(k) = \lambda(k) \mathbf{E}^T(k) \mathbf{W}(k) \mathbf{E}(k) + (1 - \lambda(k)) \|\mathbf{V}(k) - \mathbf{V}^N(k)\|^2 + \xi \mathbf{V}_Z^2$$

dont la solution sera donnée par :

$$(111) \quad \frac{\partial J(k)}{\partial \mathbf{V}(k)} (\mathbf{V}^*(k)) = 0.$$

Cette méthode sera appliquée à un problème d'évitement d'obstacles dans le paragraphe suivant.

5.2. CAS DES SEGMENTS

L'estimation des paramètres de type « segments » présente quelques différences par rapport la méthode précédente : en fait la première étape (correspondant à l'EOF) est le calcul des paramètres (\hat{p}, \hat{q}) , qui est obtenu par approximation polygonale.

Cette étape est moins locale que l'EOF, de ce fait on peut espérer que les estimations obtenues seront moins bruitées que dans le cas de points. C'est pourquoi le raffinement de l'estimée au long de la séquence peut amener à de bons résultats sans nécessiter de s'écarter de la trajectoire nominale. Toutefois, si cela s'avère nécessaire, la même démarche que pour les points pourra être mise en œuvre.

L'analyse du « bruit » sur les équations de mesures (63) et (64) ne sera pas détaillée ici. Elle montre que la meilleure qualité d'estimation, en termes de S/B, est obtenue lorsque le mouvement du capteur est tel que, entre l'image k et l'image $k+1$, $p(k+1) - p(k)$ est grand et $q(k+1) - q(k)$ est petit. En appelant les

valeurs désirées de ces variations $\Delta^* p(k+1)$, $\Delta^* q(k+1)$, nous pouvons construire la fonction de coût de la même façon que précédemment :

$$(112) \quad J'(k) = \lambda(k) \sum_{j=1}^{m(k)} \{ \sigma^j(k) (\tilde{p}_j(k+1) - \tilde{p}_j(k) - \Delta^* p(k+1))^2 + \tau^j(k) (\tilde{q}_j(k+1) - \tilde{q}_j(k) - \Delta^* q(k+1))^2 \} + (1 - \lambda(k)) \|\mathbf{V}(k) - \mathbf{V}^N(k)\|^2$$

pour chaque ligne L_j , $j = 1 \dots m(k)$.

En utilisant les équations (32), (33), (94) et (95), et les résultats du filtre récursif, on obtient une expression approximée de $J'(k)$ qui dépend explicitement de $\mathbf{V}(k)$. Cependant, du fait des non-linéarités, une méthode numérique est nécessaire pour calculer $\mathbf{V}^*(k)$ minimisant $J'(k)$.

6. Application en robotique

Dans cette dernière partie, nous allons présenter deux applications dans le domaine de la robotique utilisant l'approche décrite précédemment. Les résultats ont été obtenus sur un outil de simulation développé à l'IRISA et qui permet de simuler les images perçues par une caméra mobile se déplaçant dans un univers 3D au moyen de techniques de synthèse d'image [22].

Nous allons tout d'abord décrire une simple application d'évitement d'obstacles, puis nous superposerons un mouvement destiné à renforcer l'estimation suivant l'approche décrite dans la section IV. Ces deux expérimentations utilisent des primitives 3D de type « points », l'estimation des primitives de type « segments » n'étant pas encore implantée.

6.1. COMMANDE SIMPLE EN BOUCLE FERMÉE

Dans ce premier exemple, l'estimation récursive n'est pas réalisée et les coordonnées 3D calculées à partir des équations (87), (4) et (5) sont très bruitées. Considérons la tâche locale d'évitement d'obstacles suivante : partant d'une position initiale dans le repère R_c , nous désirons atteindre un but X^g en évitant les obstacles. La trajectoire nominale est donnée par :

$$(113) \quad \mathbf{V}^M = \eta \mathbf{X}_c^g$$

où l'indice c représente l'expression des coordonnées du but dans le repère de la caméra. η peut avoir la forme suivante :

$$(114) \quad \eta = \mathbf{V}_0 \|\mathbf{X}_c^g\|^{-1}.$$

La tâche d'évitement d'obstacles est réalisée en ajoutant à \mathbf{V}^M un mouvement local \mathbf{V}_i calculé en associant aux obstacles $\{0\}$ un champ de potentiel de type newtonien, tel que :

$$(115) \quad \begin{cases} \forall \mathbf{P}_i (X^i, Y^i, Z^i) \in \{0\} \\ \mathbf{V}_i^t = -w_i \frac{\mathbf{O}_c \mathbf{P}_i}{\|\mathbf{O}_c \mathbf{P}_i\|^3} \end{cases}$$

exprimé dans R_c , avec un coefficient de pondération w_i .

Naturellement V_i est calculé à partir de la valeur estimée de \tilde{Z}_i , ce qui finalement donne à chaque instant :

$$(116) \quad V_i(k) = \frac{1}{N(k)} \times \sum_{i=1}^{N(k)} \frac{\omega_i}{\tilde{Z}_i^2(k) (\tilde{x}_i(k) + \tilde{y}_i(k) + 1)^{3/2}} x^P.$$

Le mouvement résultant est alors :

$$(117) \quad V^N(k) = V^M(k) + V_i(k).$$

La figure 4 [16] montre un exemple de trajectoire générées par cet algorithme.

Le principal intérêt de ce simple exemple est de montrer que, dans une approche en *boucle fermée*, le processus de lissage peut être situé en dehors du processus d'estimation et de mesure, quand la tâche le permet; ici l'étape de « moyennage » des mesures locales est intégrée dans le calcul du déplacement global sans qu'il y ait filtrage explicite de ces mesures.

6. 2. ESTIMATION RÉCURSIVE AVEC COMMANDE OPTIMALE

Dans cette seconde expérimentation, l'approche décrite dans l'article a été totalement implantée. La scène est du même type que précédemment et le mouvement nominal est défini par la même tâche : atteindre le but en évitant localement les obstacles. $V^N(k)$ est alors défini par l'équation (117), avec $\tilde{Z}_i(k)$ fourni par le filtre récursif. La résolution complète de l'équation (110) est donnée en Annexe B. Les différents paramètres ou variables qui apparaissent dans l'algorithme sont les suivants :

a. Le mouvement désiré $v_d^i(k)$

Il est défini par (cf. section 4)

$$(118) \quad v_d^{iT} = 0$$

$$(119) \quad v_d^{i\perp} = v_d^i$$

v_d^i est la valeur maximale autorisée par l'EOF.

Nous pouvons écrire alors :

$$(120) \quad v_d^i(k) = \begin{pmatrix} \tilde{\beta}_i(k) & \tilde{\alpha}_i(k) \\ \tilde{\alpha}_i(k) & -\tilde{\beta}_i(k) \end{pmatrix} \times \begin{pmatrix} v_d^i \\ 0 \end{pmatrix}$$

et

$$v_d^i(k) = \begin{pmatrix} \tilde{\beta}_i(k) \\ \tilde{\alpha}_i(k) \end{pmatrix} v_d^i.$$

b. Le mouvement prédit $\tilde{v}^i(k)$

Il est donné par :

$$(121) \quad \tilde{v}^i(k) = \frac{1}{\tilde{Z}_i(k)} \begin{pmatrix} -V_x(k) + \tilde{x}_i(k) V_z(k) \\ -V_y(k) + \tilde{y}_i(k) V_z(k) \end{pmatrix}$$

c. Les coordonnées pixel estimées dans l'image k

$$(122) \quad \begin{pmatrix} \tilde{x}_i(k) \\ \tilde{y}_i(k) \end{pmatrix} = \begin{pmatrix} x_i(k-1) \\ y_i(k-1) \end{pmatrix} + \tilde{v}_i(k-1)$$

$\tilde{v}_i(k-1)$ est calculé à partir de l'EOF, et défini par sa pente dans le repère R_c [angle $\tilde{\delta}_i(k-1)$] et sa norme :

$$(123) \quad \text{tg } \tilde{\delta}_i(k-1) = \frac{-V_y^*(k-1) + y_i(k-1) V_z^*(k-1)}{-V_x^*(k-1) + x_i(k-1) V_z^*(k-1)}$$

$$(124) \quad \|\tilde{v}_i(k-1)\| = \frac{\tilde{v}_i^+(k-1)}{\cos(\tilde{\delta}_i(k-1) - \tilde{\theta}_i(k-1))}$$

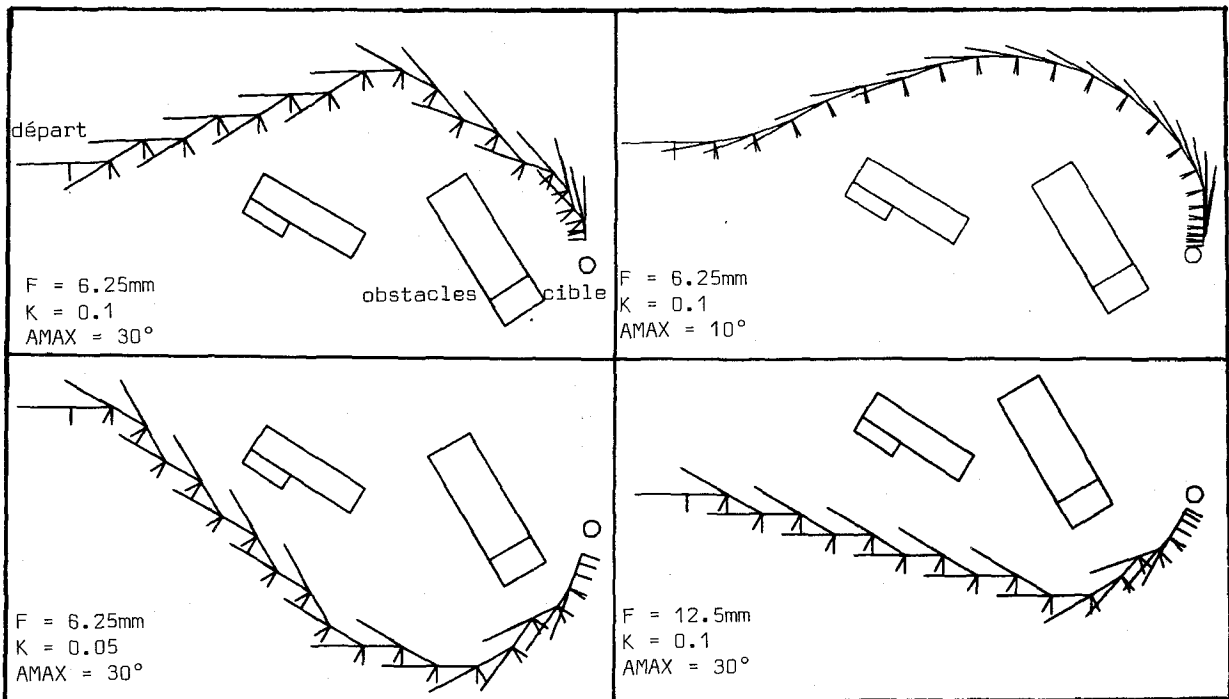


Fig. 4. Exemples de trajectoires engendrées par l'algorithme d'évitement d'obstacles.

d. Le paramètre de pondération locale $\sigma^i(k)$

Ces paramètres sont un point clé de la méthode. L'idée de base provient du fait que dans une image la direction des contours est aléatoire et, donc, les mouvements locaux désirés également (orthogonaux aux contours). Il en résultera que le mouvement global « optimal » découlant de la minimisation du critère sera oscillant ou n'aura aucun sens. C'est pourquoi, nous associons à chaque point, un paramètre $\sigma^i(k)$ mesurant la qualité de l'estimation courante. Ainsi, à chaque instant, le mouvement superposé favorisera les points les plus mal estimés. Nous avons testé deux types de fonctions :

$$(125) \quad \sigma'_i(k) = q_i(k)$$

et

$$(126) \quad \sigma'_i(k) = \frac{q_i(k)}{\hat{Z}_i(k)}$$

avec

$$(127) \quad \sigma_i(k) = \frac{\sigma'_i(k)}{\sum_{i=1}^{N(k)} \sigma'_i(k)}$$

$q_i(k)$ est la variance de l'erreur d'estimation associée au filtre pour chaque point et à chaque instant. L'introduction de la profondeur estimée dans l'équation (126) a deux avantages :

– Du point de vue de la tâche, les points les plus proches sont les plus importants : leur poids est

augmenté dans le calcul du mouvement optimal.

– Obtenir la valeur désirée v_i^d pour des points éloignés suppose des mouvements de la caméra de grande amplitude ce qui perturbe beaucoup la trajectoire nominale et rend plus difficile l'estimation des points les plus proches.

e. Le paramètre de pondération globale $\lambda(k)$

Le but de ce paramètre est d'éviter des perturbations inutiles de la tâche nominale lorsque les points sont bien estimés. Nous définissons un index global de qualité :

$$(128) \quad I(k) = \frac{1}{N(k)} \sum_{i=1}^{N(k)} q_i(k)$$

et nous choisissons, par exemple :

$$(129) \quad \lambda(k) = 1 - e^{-\theta I(k)}$$

Le dernier point important dans la mise en œuvre de cet algorithme concerne la mise en correspondance locale des points : il est nécessaire que les mesures données par l'EOF pour un point i tout au long de la séquence correspondent au même point 3D dans la scène. Deux problèmes doivent être résolus :

– *mise en correspondance* : Une fenêtre 5×5 est centrée sur la valeur prédite $\hat{x}_i(k+1)$, $\hat{y}_i(k+1)$, et une fonction de distance entre ce point et les autres points appartenant au contour contenu dans la fenêtre est calculée. Le plus proche voisin est alors mis en correspondance avec le point prédit.

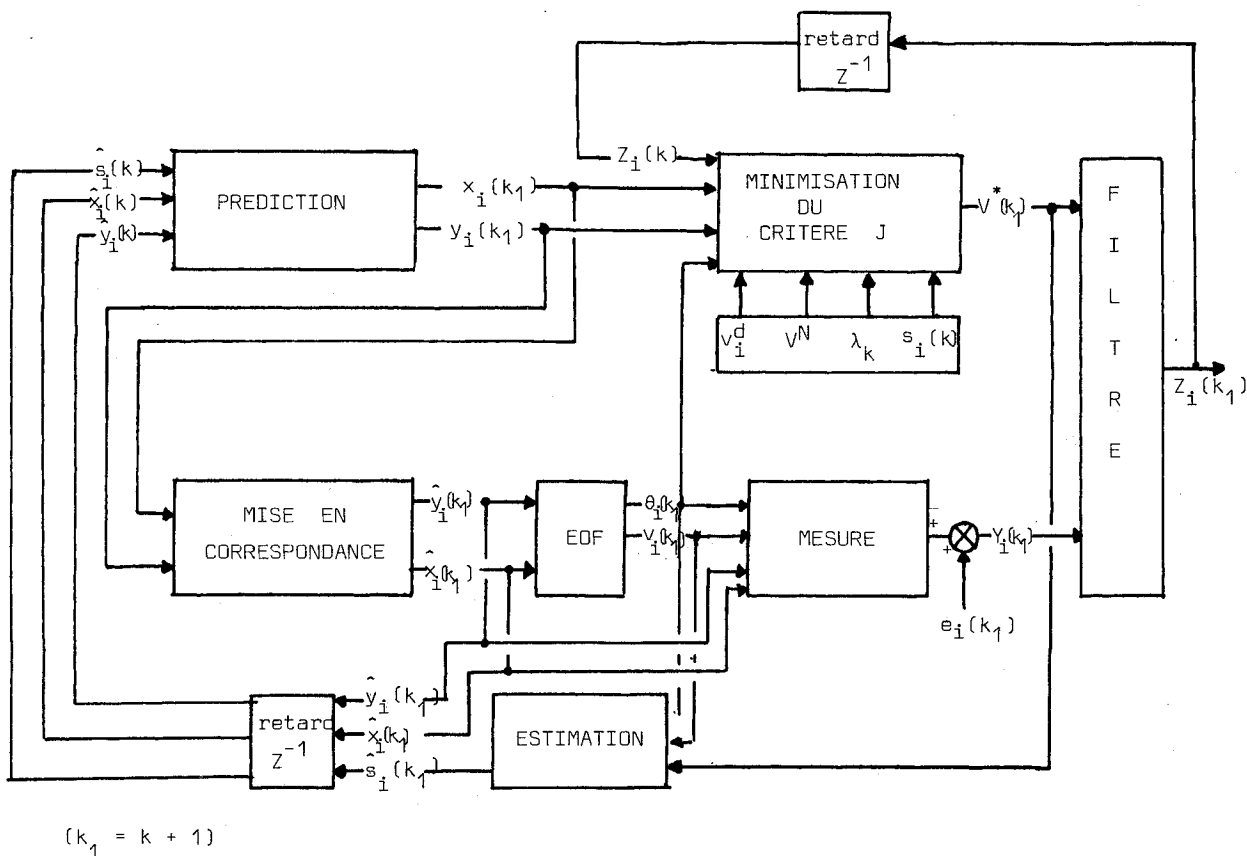


Fig. 5. — Schéma fonctionnel de l'algorithme.

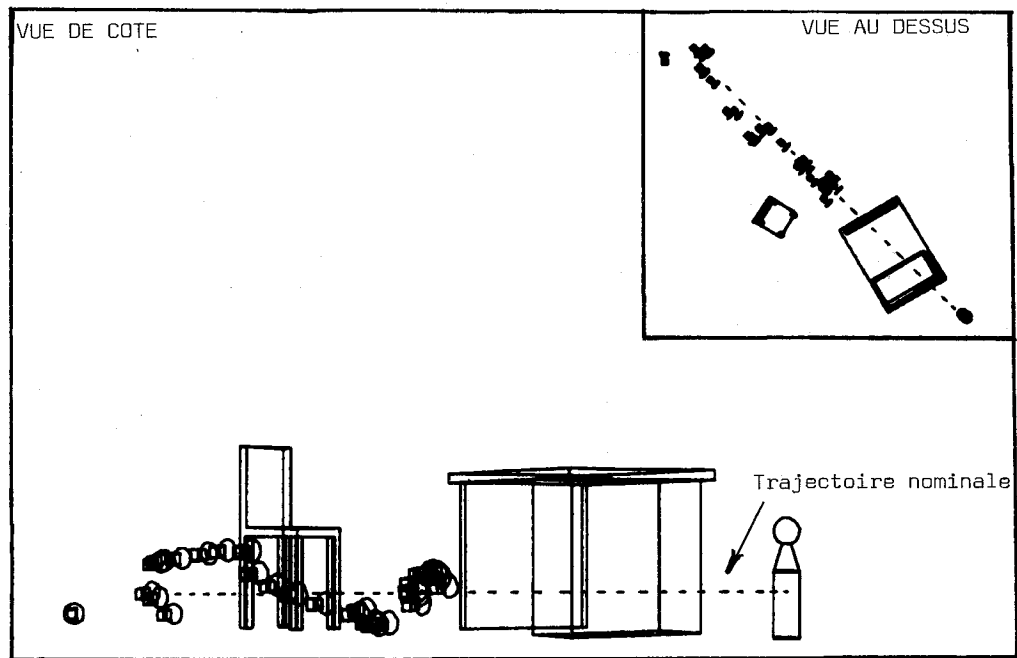


Fig. 6. — Exemple de trajectoire avec estimation récursive.

— *gestion des entrées/sorties*: Les points entrants dans l'image en cours de séquence sont initialisés avec la profondeur moyenne dans la scène et une grande variance leur est associée. Les points quittant l'image sont détectés lorsque leur prédiction sort du domaine de l'image.

La figure 5 résume l'ensemble de l'algorithme et la figure 6 montre la trajectoire obtenue dans un cas similaire à celui de la première expérimentation.

7. Conclusion

Nous avons présenté dans cet article, un schéma d'estimation récursive de primitives 3 D utilisant des informations dynamiques extraites d'une séquence d'images. Une méthode originale de contrôle en boucle fermée a été implantée dans le cas de l'estimation de la carte des profondeurs avec une caméra se déplaçant en translation. Ces méthodes ont été validées en simulation.

Cependant, de nombreuses améliorations restent à faire dans le domaine de l'estimation récursive de primitives 3 D au moyen du mouvement.

En premier, concernant les algorithmes décrits, des efforts doivent être fait pour simplifier les modèles et décroître les coûts calcul. En ce qui concerne le filtrage, les approximations faites sur les modèles et les perturbations suggèrent l'emploi de techniques plus robustes que le filtrage de Kalman, tel le gradient stochastique. De plus, la méthodologie demande à être validée sur images réelles.

Concernant les primitives de type « segments », le schéma d'estimation proposé doit être vérifié et le passage à des primitives de type « surfaces » reste à faire. Finalement, il faut également remarquer que l'extension de l'approche au cas des rotations n'est

pas directe et nécessite des précautions notamment au niveau du choix des représentations.

En conclusion, nous pressentons que des tâches complexes en univers inconnu peuvent être réalisées par un robot avec l'approche proposée qui permet d'acquérir *en ligne*, durant le déplacement et avec une caméra unique, les informations nécessaires à la bonne exécution de la tâche. Ce moyen pourrait être un bon complément ou même une alternative aux méthodes classiques de stéréovision ou de vision 3 D active.

Manuscrit reçu le 19 décembre 1986.

BIBLIOGRAPHIE

- [1] J. K. AGGARWAL, L. S. DAVIS et W. N. MARTIN, Correspondence processes in dynamic scene analysis, *Proc. of IEEE*, 69, n° 5, 1981.
- [2] M. BERTHOD et P. LONG, Stéréoscopie par mise en correspondance de graphes, *Colloque CESTA/GRETSI*, Biarritz, France, May 1984.
- [3] Y. YAKIMOVSKY et R. CUNNINGHAM, A system for extracting Three-dimensional Measurements from a stereo pair of T.V. Cameras, *Computer Graphics and Image Processing*, 7, 1978, p. 195-210.
- [4] N. AYACHE et B. FAVERJON, Fast stereo matching of edge segments, *IEEE Conf. on CVPR*, San Francisco, June 1985.
- [5] R. F. BUXTON et H. BUXTON, Monocular depth perception from optical flow by space time signal processing, *Proc. R. Soc.*, London, B218, 1983.
- [6] E. C. HILDRETH, The measurement of visual motion, *ACM distinguished dissertation*, MIT Press, 1983.
- [7] E. C. HILDRETH, Computations underlying the measurement of visual motion, *Artificial Intelligence*, n° 23, 1984.

- [8] P. BOUTHEMY, Estimation of edge motion based on local modelling, *SPIE Conference on Robot Vision*, Cannes, France, Dec. 1985.
- [9] H. H. NAGEL, Recent advances in image sequence analysis, *Colloque CESTA/GRETSI*, Biarritz, France, May 1984.
- [10] A. BRUSS et B. K. P. HORN, Passive navigation, *CVG & IP*, n° 21, 1983.
- [11] R. Y. TSAI et T. S. HUANG, Uniqueness and Estimation of Three-dimensional Motion parameters of rigid objects with curved surfaces, *IEEE Trans. on PAMI*, PAMI-6, n° 1, janv. 1984.
- [12] J. WEBB et J. K. AGGARWAL, Structure from motion of rigid and jointed objects, *Artificial Intelligence*, 19, 1982, p. 107-130.
- [13] H. P. MORAVEC, Robot visual obstacle avoidance, *IJCAI*, 1981.
- [14] J. H. RIEGER, Information in optical flow induced by curved path of observation, *Journal of Opt. Soc. Am.*, 73, n° 3, March 1983.
- [15] A. C. SANDERSON et L. E. WEISS, Adaptive visual servo-control of robots, *Robot vision*, A. Pugh, 1982.
- [16] P. RIVES et L. MARCE, Use of moving vision sensors in robotics. Application to an obstacle avoidance task, *ICAR'85*, Tokyo, Sept. 1985.
- [17] R. NEVATIA et R. BABU, Line feature extraction and description, *CGIP* 13, 1980, p. 257-269.
- [18] W. E. L. GRIMSON et E. C. HILDRETH, Comments on "Digital step edges from zero crossings of second directional derivatives", *IEEE Trans. PAMI*, PAMI 7, Jan. 1985, p. 121-129.
- [19] R. DERICHE, Optimal edge detection using recursive filtering. Submitted to *International Journal on Computer Vision*.
- [20] J. F. CANNY, Finding edges and lines in images, *MIT Artificial Intelligence laboratory*, TR-720, 1983.
- [21] D. H. BALLARD, Strip trees: a hierarchical representation for curves, *Com. of ACM*, 24, n° 5, May 1981, p. 310-321.
- [22] P. RIVES et G. HEGRON, Design of a simulation tool for robots using moving vision sensors, *Proc. of NATO Int. Adv. Res. workshop on languages for sensor-based control in robotics*, Castelvechio Pascoli, Italy, Sept. 1986.

Annexe A

Principales notations

k : index temporel dans la séquence, i : index spatial dans l'image.

PARAMÈTRES 3 D

- R_s , repère scène.
- R_c , repère image.
- $X=(XYZ)^T$, coordonnées 3 D dans R_c .
- $\tilde{Z}(k)$, estimation de Z à l'instant k .
- $q(k)$, variance de l'erreur d'estimation sur $\tilde{Z}(k)$.

- $\tilde{Z}(k)$, mesure de Z à l'instant k .
- e_z , bruit sur \tilde{Z} après linéarisation.
- $V=(V_x V_y V_z)^T$, vitesse en translation de R_c .
- $\Omega=(\Omega_x \Omega_y \Omega_z)^T$, vitesse en rotation de R_c .
- $\varepsilon_v=(\varepsilon_x \varepsilon_y \varepsilon_z)^T$, bruit sur V .
- q_v , variance de ε_v .
- ε_Ω , bruit sur Ω .
- $\tilde{V}=V+\varepsilon_v$.
- $\tilde{\Omega}=\Omega+\varepsilon_\Omega$.
- V^N , mouvement nominal.
- V^M , mouvement d'attraction de la cible.
- V^e , mouvement de répulsion des obstacles.
- V^* , mouvement optimal calculé.
- X_0 , coordonnées 3 D d'un point appartenant à \mathcal{D} .
- $u=(\alpha\beta\gamma)^T$, vecteur directeur de \mathcal{D} .
- $h=(h_1 h_2 h_3)^T=X_0 X u$.

PARAMÈTRES 2 D

- $p=h_1/h_2$, $q=(\alpha P-\beta)/\gamma$, paramètres du segment projeté dans l'image.
- $x^P=(xy 1)$, coordonnées de la projection p de P dans l'image.
- e_x, e_y , bruits sur x, y .
- $v=(\dot{x} \dot{y})^T=(v^\perp v^T)^T$, mouvement de p .
- v^\perp , composante orthogonale au contour de v .
- v^T , composante tangentielle de v .
- u_T, v_T , vitesse en translation de p .
- u_r, v_r , vitesse en rotation de p .
- θ , orientation du contour spatial.
- v^* , mesure fournie par l'EDF.
- $\tilde{v}^\perp=v^\perp+\varepsilon_v$, mesure de v^\perp fournie par l'EDF.
- $\tilde{\theta}=\theta+\varepsilon_\theta$, mesure de θ fournie par l'EDF.
- ε_θ , bruit de quantification sur θ .
- ε_v , bruit de quantification sur v .
- $\lambda_i=\text{Cos } \tilde{\theta}$.
- $\tilde{\lambda}_2=\text{Sin } \tilde{\theta}$.
- z , équation générale de mesure.
- \tilde{z} , équation générale de mesure linéarisée.
- e_p , bruit sur \tilde{v}^\perp après linéarisation.
- $(\tilde{x}_i(k), \tilde{y}_i(k))$, position estimée dans l'image k du point i .
- $\tilde{v}_i(k-1)$, estimation du vecteur déplacement entre $k-1$ et k .
- $\tilde{\delta}_i(k-1)$, pente de $\tilde{v}_i(k-1)$.
- $\tilde{v}_i(k)$, prédiction du mouvement de $\tilde{p}_i(k)$ dans l'image $(k+1)$.
- $v_d^i(k)$, mouvement désiré dans l'image.
- $\sigma_i(k)$, paramètre de pondération locale.
- $\lambda(k)$, paramètre de pondération globale.
- ξ , paramètre de pondération sur V_z .

Annexe B

Mouvement optimal

Il est donné, à chaque instant k (k est omis dans les équations par souci de simplicité) par :

$$\left. \begin{aligned} V_X^* &= \frac{1}{v} (-\lambda B + \lambda C V_Z^* + (1-\lambda) V_X^N) \\ V_Y^* &= \frac{1}{v} (-\lambda A + \lambda D V_Z^* + (1-\lambda) V_Y^N) \\ V_Z^* &= \frac{\lambda (-E + (1/v) [\lambda (BC + AD) - (1-\lambda) (V_X^N + D V_Y^N)]}{\lambda (1-F + (\lambda/2) (C^2 + D^2) - \xi - 1)} \\ &\quad - \frac{(1-\lambda) V_Z^N}{\lambda (1-F + (\lambda/v) (C^2 + D^2) - \xi - 1)} \end{aligned} \right\}$$

$$A = \sum_{i=1}^{N(k)} \frac{\sigma_i \tilde{\lambda}_1^i}{\tilde{Z}_i}; \quad B = \sum \frac{\sigma_i \tilde{\lambda}_2^i}{\tilde{Z}_i} v_i^d$$

$$C = \sum \sigma_i \frac{\tilde{x}_i}{\tilde{Z}_i^2}; \quad D = \sum \sigma_i \frac{\tilde{y}_i}{\tilde{Z}_i^2}$$

$$E = \sum \sigma_i \left(\frac{\tilde{\lambda}_2^i \tilde{x}_i + \tilde{\lambda}_1^i \tilde{y}_i}{\tilde{Z}_i} \right) v_i^d$$

$$F = \sum \sigma_i \left(\frac{\tilde{x}_i^2 + \tilde{y}_i^2}{\tilde{Z}_i^2} \right)$$

$$\mu = \sum \sigma_i \frac{1}{\tilde{Z}_i^2}; \quad v = \lambda \mu + 1 - \lambda.$$