Coopération multi-caméras pour la recherche puis le positionnement par rapport à un objet

Claire Dune[1,2] Eric Marchand[1] Christophe Leroux[2]

[1] INRIA, IRISA, projet Lagadic, F-35042 Rennes, France
[2] CEA List, F-92265 Fontenay Aux Roses, France
claire.dune@irisa.fr

Résumé

La plupart des systèmes de contrôle multi-caméra reposent sur l'hypothèse forte que la zone d'intérêt est vue depuis tous les capteurs à l'instant initial. Cet article propose une approche pour garantir la véracité de cette hypothèse pour un système de préhension hybride comprenant deux caméras, l'une déportée (dite "eye-to-hand"), l'autre embarquée (dite "eye-in-hand"). On suppose que l'objet à saisir est dans la champ de vision de la caméra déportée, tandis que la caméra embarquée a une pose quelconque qui ne lui permet pas nécessairement de voir l'objet. On souhaite garantir la présence de l'objet d'intérêt dans le champ de vision de la caméra déportée sans connaître de modèle de l'objet et sans utiliser de base de données. La scène considérée est complexe et aucune hypothèse n'est faite sur l'arrière plan. La méthode proposée a été testée et validée sur un système robotique multi caméra.

Mots Clef

Asservissement visuel, géométrie multi vue, théorie Bayésienne, positionnement automatique de capteur, robotique.

Abstract

A critical assumption of many multi-view control systems is the initial visibility of the regions of interest from all the views. This paper proposes an initialization step for a hybrid eye-in-hand/eye-to-hand grasping system to overcome this requierement. In this paper, the object of interest is assumed to be within the eye-to-hand field of view, whereas it may not be within the eye-in-hand one. The object model is unknown and no database is used. The object lies in a complex scene with a cluttered background. A method to automatically focus the object of interest is presented, tested and validated on a multi view robotic system.

Keywords

Visual Servoing, multiple view geometry, Bayesian theory, autonomous sensor positionning, robotics.

1 Introduction

Nos travaux s'inscrivent dans le cadre d'un projet d'aide à la saisie pour les personnes handicapées. Dans un tel contexte, il est indispensable de limiter l'action de l'utilisateur au minimum d'intervention. Cet article propose une approche nécessitant un unique "clic" dans la vue de la caméra déportée afin de sélectionner l'objet à saisir.

Le système robotique considéré est constitué d'un bras manipulateur commandé par un système de vision hybride qui comporte une caméra déportée et une caméra embarquée fixée sur l'organe effecteur du bras. La première offre une vue large de la zone de travail du bras et la seconde donne une vue détaillée d'une partie de la scène. L'objectif est d'utiliser les données issues des deux caméras pour contrôler le bras. Quelques articles [14, 8, 6, 10] traitent le problème de la coopération d'une caméra déportée et d'une caméra embarquée. L'utilisation combinée de ces deux types de caméras a déjà fait ses preuves en donnant de meilleurs résultats pour l'exécution de tâches complexes que des méthodes reposant sur un système de stéréo-vision embarqué [3, 11] ou une unique caméra [2]. En effet, les caractéristiques des deux caméras sont complémentaires. Si la caméra embarquée permet une grande précision dans l'asservissement visuel, la caméra déportée peut faciliter des tâches plus haut niveau comme l'évitement d'obstacles et de butées.

La première étape vers une saisie semi automatique d'objets pour les personnes handicapées est le positionnement des caméras afin que l'objet à saisir soit dans leurs champs de vision. À l'instant initial, l'objet à saisir est dans le champ de vision de la caméra déportée. La caméra embarquée, quant à elle, a une pose quelconque, par conséquent aucune hypothèse ne peut être faite a priori sur la présence de l'objet dans son champ de vision.

Cet article propose une méthode permettant de de centrer l'objet à saisir dans le champ de vision de la caméra embarquée sans considérer de modèle de l'objet, sans émettre d'hypothèse sur sa texture et sa forme et sans utiliser de base de données d'images. La scène peut contenir d'autres objets que l'objet à saisir et l'arrière plan est quelconque. De plus, l'objet peut être mobile. Par ailleurs, le système est entièrement calibré et l'objet est supposé appartenir à la zone de travail du bras, on connaît un point de l'objet,sélectionné par l'utilisateur à l'instant initial, dans la vue de la caméra déportée.

Si on connaît un point de l'objet dans l'image déportée, alors on connaît les coordonnées X et Y d'un point de l'objet dans le repère de la caméra déportée. Cependant, sans aucun a priori sur l'objet, on ne peut accéder directement à la coordonnée Z du point cliqué. On sait seulement que l'objet est sur une ligne passant par le centre optique de la caméra déportée et le point cliqué. Cet article propose d'utiliser un asservissement visuel reposant sur la géométrie épipolaire pour rechercher l'objet le long cette ligne de vue à l'aide de la caméra embarquée. Les commandes robotiques utilisant la géométrie épipolaire ont été étudiées par de nombreux groupes de recherche ces dernières années, tout particulièrement pour des applications de retour à la base [16, 17, 1]. Dans [16, 17] l'asservissement visuel est basé sur la mise en correspondance de la position de l'épipole désirée et de sa position courante. Dans [1] c'est la mise en correspondance des lignes epipolaires courante et désirée qui guide l'asservissement. Cependant toutes ces méthodes supposent qu'un ensemble d'éléments de la scène sont communs à la vue désirée et à la vue courante à l'instant initial pour pouvoir construire la géométrie épipolaire entre ces vues. La plupart des articles utilisant un système hybride composé d'une caméra embarquée et d'une caméra déportée [14, 8, 6, 1] font également cette hypothèse. Dans [10] une phase d'initialisation permet de centrer le champ de vision d'une caméra mobile sur un objet en mouvement détecté dans une caméra fixe.

Notre approche peut être vue comme une extension de [1] et [10] où l'asservissement visuel consiste à faire "surfer" la caméra sur le plan épipolaire tout en recherchant une zone d'intérêt. De la même manière que [10], l'asservissement ne se base pas sur la mise en correspondance de la vue courante et d'une vue désirée mais sur la mise en correspondance de lignes épipolaires. L'asservissement visuel est virtuel. De plus le système est calibré et la géométrie épipolaire n'est pas estimée mais calculée explicitement. Dans [10], la caméra mobile a deux degrés de libertés, la base de la géométrie épipolaire est petite et fixe, tandis que la caméra mobile de notre application est montée sur un bras à 6 degrés de liberté. Ainsi la base de la géométrie varie au cours du temps et est de plus grande dimension. De plus, dans [10], la commande repose sur un couplage géométrique et cinématique défini pour le système de caméras considéré alors que nous utiliserons dans cet article la commande classique par asservissement visuel telle que proposée dans [7].

Les méthodes de reconnaissance d'objet classiques permettent de retrouver l'objet dans les images acquises au cours du déplacement de la caméra mobile. L'odomètrie du système nous donne la position de la caméra mobile à chaque instant. Il est ainsi possible d'estimer la profondeur de l'objet sur le segment 3D.

La deuxième partie de cet article sera consacrée à l'asservissement visuel utilisant la géométrie épipolaire. Nous y proposons une solution complète pour garantir que l'objet est sur une ligne centrée et horizontale dans la vue mobile. La troisième partie traitera de la recherche de l'objet le long de la ligne de vue. La quatrième partie présente quelques résultats expérimentaux qui valident la méthode proposée. Enfin, la cinquième partie dresse les conclusions de ce travail et en présente les perspectives.

2 Asservissement visuel reposant sur la géométrie épipolaire

Cette partie présente la coopération de deux caméras embarquée et déportée, permettant de maintenir la ligne de vue issue du point cliqué centrée et horizontale dans la vue de la caméra embarquée. Dans un premier temps, nous présenterons le système robotique. Ensuite, nous rappellerons les principes de la géométrie épipolaire et enfin nous présenterons une commande par asservissement visuel virtuel utilisant la géométrie épipolaire.

2.1 Un système hybride : eye-in-hand/eye-to-hand

Soit $\mathscr{R}_f$, le référentiel du laboratoire. Soit c_f, une caméra déportée, et c_m, une caméra mobile embarquée sur l'organe effecteur d'un bras manipulateur à six degrés de liberté (voir Figure 1). Soit $\mathscr{R}_{c_f}$ et $\mathscr{R}_{c_m}$ les référentiels attachés respectivement à chacune de ces deux caméras. Soit $\mathscr{R}_o$ le repère lié à l'objet. Enfin, soit ${}^{c_f}\mathbf{M}_{c_m}$ la matrice de passage entre les repères des deux caméras. Notons que la matrice de passage entre le repère fixe et la caméra fixe ${}^{f}\mathbf{M}_{c_f}$ est constante (elle peut être évaluée au cours d'une procédure d'estimation de la pose classique). D'autre part, l'odomètrie du système robotique donne une estimation assez précise de la matrice de passage entre le repère fixe et le repère attaché à la caméra mobile ${}^{c_m}\mathbf{M}_f$ a chaque instant. Alors ${}^{c_m}\mathbf{M}_{c_f}$ peut être calculée par ${}^{c_m}\mathbf{M}_{c_f} = {}^{c_m}\mathbf{M}_f\,{}^{f}\mathbf{M}_{c_f}$ à chaque instant.

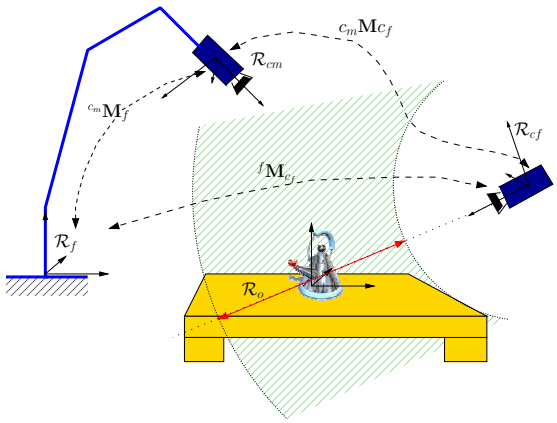

FIG. 1 – Référentiels et notations pour le système hybride eye-in-hand/eye-to-hand

2.2 Géométrie multi-vues

On souhaite que la ligne de vue issue du point cliqué soit centrée et horizontale dans le plan image de la caméra mo-

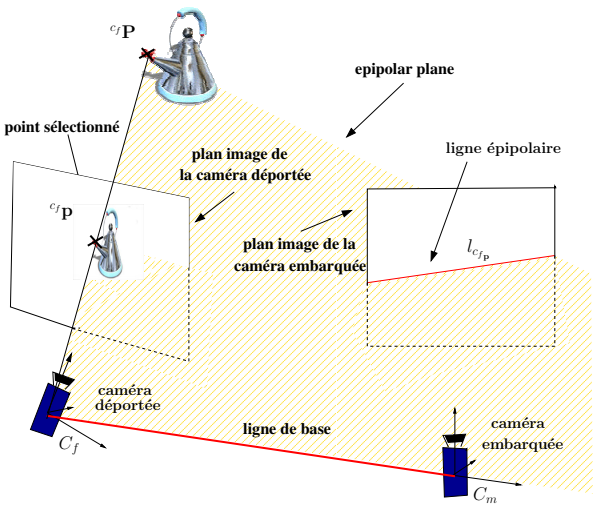

FIG. 2 – Géométrie épipolaire

bile.

Soit $\mathbf{C}_m$ et $\mathbf{C}_f$ les centres optiques respectifs de c_m et c_f. La géométrie épipolaire établit la relation entre un point 3D $^{c_f}\mathbf{P}$ et sa projection dans les deux plans images des caméras composant le système $^{c_f}\mathbf{p}$ and $^{c_m}\mathbf{p}$. La ligne $\mathbf{C}_f\mathbf{C}_m$ est dite ligne de base de la géométrie épipolaire (voir Figure 2). La contrainte épipolaire peut s'écrire de la manière suivante [9] :

$$^{c_f}\mathbf{p}^T\,{}^{c_f}\mathbf{E}_{c_m}\,{}^{c_m}\mathbf{p} = 0 \tag{1}$$

où $^{c_m}\mathbf{E}_{c_f}$ est la matrice essentielle. Elle est définie ainsi :

$$^{c_m}\mathbf{E}_{c_f} = [{}^{c_m}\mathbf{t}_{c_f}]_\times\,{}^{c_m}\mathbf{R}_{c_f} \tag{2}$$

avec $^{c_m}\mathbf{t}_{c_f}$ et $^{c_m}\mathbf{R}_{c_f}$ les matrices de translation et de rotation entre les référentiels des deux caméras et $[.]_\times$ la matrice antisymétrique ou de pré-produit vectoriel.

La matrice essentielle (2) met en relation la contrainte épipolaire et les paramètres extrinsèques du système. Connaissant $^{c_m}\mathbf{M}_{c_f}$ à chaque instant on peut calculer la matrice essentielle à chaque itération de la commande. De plus, dans (1), $^{c_f}\mathbf{E}_{c_m}\,{}^{c_m}\mathbf{p}$ est l'équation de la ligne passant par $^{c_f}\mathbf{p}$ et l'épipole $^{c_m}\mathbf{e}$ (la projection de $\mathbf{C}_f$ sur le plan image de c_m). On peut aussi définir la ligne épipolaire associée à $^{c_f}\mathbf{p}$ comme la projection sur le plan image de c_m de la ligne passant par $^{c_f}\mathbf{p}$ et $\mathbf{C}_f$, le centre optique de la caméra déportée. La matrice essentielle est alors la relation entre les points et les lignes épipolaires dont nous avons besoin pour commander les déplacements de la caméra mobile. Les deux caméras sont suffisamment éloignées l'une de l'autre pour assurer la stabilité de la géométrie épipolaire qui peut être alors utilisée de manière robuste pour réaliser une commande par asservissement visuel.

Étant donné le point $^{c_f}\mathbf{p}$ connu à chaque instant et la calibration du système, on peut déduire l'équation de la ligne épipolaire dans le plan image de la caméra embarquée en négligeant les mouvements potentiels de $^{c_f}\mathbf{p}$.

Dans la suite de cet article, la ligne épipolaire est représentée par les deux paramètres (ρ,θ). L'équation de la ligne qui sera utilisée dans la commande par asservissement visuel est alors $xcos\theta + ysin\theta = \rho$ où ρ et θ sont obtenus à partir de l'équation (1) et x et y sont les coordonnées des points du plan image appartenant à la ligne.

2.3 Asservissement visuel reposant sur la géométrie épipolaire

Rappelons tout d'abord la spécification des tâches à effectuer : la ligne épipolaire associée à $^{c_f}\mathbf{p}$ doit être amenée au milieu du plan image de la caméra mobile. Tout en maintenant la ligne épipolaire horizontale et centrée dans sa vue, la caméra mobile parcourt l'espace de travail du bras à la recherche de l'objet à saisir.

Formalisme et notation l'asservissement visuel multi tâches. L'asservissement visuel est une commande robotique reposant sur des indices visuels extraits d'images acquises par une ou plusieurs caméras. Soit $\mathbf{s}$ les information visuelles courantes et $\mathbf{s}^*$ les informations visuelles désirées. La tâche principale consiste à faire tendre le vecteur d'erreur $\mathbf{e_1}$=$\mathbf{s}-\mathbf{s}^*$ vers zéro. La matrice jacobienne image (ou matrice d'interaction) associée à une tâche relie les variations des informations visuelles considérées au torseur cinématique de la caméra dont on veut commander les déplacements. Elle est définie par [7] :

$$\dot{\mathbf{s}} = \mathbf{L}_s\mathbf{v} \tag{3}$$

Puis, si l'on considère une caméra embarquée, la loi de commande qui optimise $\mathbf{e_1}$ est [7] :

$$\mathbf{v} = -\lambda\widehat{\mathbf{L_1}}^+\mathbf{e_1} + \mathbf{P}\mathbf{z} \tag{4}$$

où $\widehat{\mathbf{L_1}}^+$ est la pseudo inverse d'une approximation ou d'un modèle de $\mathbf{L_1}$, $\mathbf{z}$ est un vecteur de contrôle arbitraire et $\mathbf{P} = \mathbf{I} - \mathbf{L_1}^+\mathbf{L_1}$ est l'opérateur de projection qui garantit que le vecteur de contrôle $\mathbf{z}$ n'affecte pas la réalisation de la tâche principale $\mathbf{e_1}$. Ainsi, si la tâche principale ne contraint pas les six degrés de liberté de la caméra mobile, il est possible d'appliquer une tâche secondaire au système par redondance.

Introduisons une deuxième tâche $\mathbf{e_2}$ et sa matrice d'interaction $\mathbf{L_2}$ dans le vecteur $\mathbf{z}$ qui devient :

$$\mathbf{z} = -\lambda\widehat{\mathbf{L_2}}^+\mathbf{e_2} \tag{5}$$

Si on inclut (5) dans (4), la loi de commande permettant de réguler les deux tâches est :

$$\mathbf{v} = -\lambda\left(\widehat{\mathbf{L_1}}^+\mathbf{e_1} + \mathbf{P}\widehat{\mathbf{L_2}}^+\mathbf{e_2}\right) \tag{6}$$

Dans le cas de notre étude, la tâche principale est la tâche de centrage de la ligne épipolaire tandis que la deuxième tâche permet de la parcourir.

Tâche principale : centrer la ligne de vue. L'information visuelle considérée est une ligne virtuelle $\mathbf{s} = (\rho, \theta)$. Sa matrice d'interaction est la suivante [7] :

$$\mathbf{L_1} = \begin{pmatrix} \lambda_\rho c\theta & \lambda_\rho s\theta & -\lambda_\rho \rho & (1+\rho^2)s\theta & -(1+\rho^2)c\theta & 0 \\ \lambda_\theta c\theta & \lambda_\theta s\theta & -\lambda_\theta \rho & -\rho c\theta & -\rho s\theta & -1 \end{pmatrix} \tag{7}$$

avec $s\theta = \sin(\theta)$ et $c\theta = \cos(\theta)$. λ_ρ et λ_θ sont obtenus par les équations suivantes :

$$\begin{cases} \lambda_\rho &= (a)\rho\cos\theta + b\rho\sin\theta + c)/d \\ \lambda_\theta &= (a\sin\theta - b\cos\theta)/d \end{cases} \tag{8}$$

où $aX + bY + cZ + d = 0$ est l'équation, exprimée dans le repère $\mathscr{R}_{c_m}$, du plan qui contient $\mathbf{C}_f$, ${}^{c_f}\mathbf{p}$ et qui est perpendiculaire au plan épipolaire.
Lorsque la tâche principale est régulée, elle garantit que le plan épipolaire coupe le plan image de la caméra mobile horizontalement en son milieu. Tout en régulant cette tâche, on peut déplacer la caméra mobile de manière à parcourir la ligne de vue pour rechercher l'emplacement de l'objet à saisir.
On peut noter que la boucle fermée permet de considérer un objet mobile. En effet, il suffit de rajouter un suivi du point cliqué ${}^{c_f}\mathbf{p}$ pour pouvoir mettre à jour à chaque instant la ligne épipolaire associée. Ceci n'aurait pas été envisageable si nous avions choisi une simple commande 3D en boucle ouverte.

Tâche secondaire : parcourir la ligne de vue. La recherche de l'objet est limitée à un segment 3D qui est l'intersection entre la ligne épipolaire et l'espace de travail du robot. Les deux extrémités de ce segment sont des points 3D virtuels ${}^{c_f}\mathbf{P}_1 = (X_1, Y_1, Z_1)$ et ${}^{c_f}\mathbf{P}_2 = (X_2, Y_2, Z_2)$, exprimés dans le repère $\mathscr{R}_{c_f}$. Ils se projettent dans le plan image de la caméra mobile en deux points $\mathbf{p_1} = (x_1, y_1)$ et $\mathbf{p_2} = (x_2, y_2)$.

La tâche principale de centrage de la ligne épipolaire associée à ${}^{c_f}\mathbf{p}$ contraint 3 degrés de liberté. Il reste donc 3 degrés de liberté disponibles pour l'application d'une tâche secondaire dédiée au parcours de la ligne de vue. En utilisant le formalisme de la redondance introduit dans la partie 2.3, le centrage successif des deux extrémités du segment ${}^{c_f}\mathbf{P}_1$ et ${}^{c_f}\mathbf{P}_2$ peut être réalisé sans perturber la régulation de la tâche principale.
Pour centrer les points ${}^{c_f}\mathbf{P}_i$, où $i = 1, 2$ dans le plan image de la caméra mobile, on construit une tâche secondaire $\mathbf{e_2}$. Les informations visuelles considérées sont la projection $\mathbf{p}_i$ du point virtuel ${}^{c_f}\mathbf{P}_i$ dans le plan image de la caméra mobile et le centre de l'image $\mathbf{p}^* = (0, 0)$. L'erreur à réguler est alors $\mathbf{e_2} = \mathbf{p}_1 - \mathbf{p}^*$. La matrice d'interaction $\mathbf{L_2}$ associée à cette tâche est la matrice d'interaction classique d'un point 2D [7].
Si on utilise une boucle ouverte, comme cela a été fait dans [1], une fois la tâche principale réalisée, la tâche secondaire consiste à faire effectuer à la caméra mobile une rotation autour de l'axe perpendiculaire au plan épipolaire. En d'autres termes pour parcourir la ligne de vue, il suffit de pivoter la caméra autour de l'axe y de son repère. Cependant, si le centrage n'est pas parfait, cette rotation perturbe petit a petit le centrage de la ligne épipolaire. L'erreur de la tâche principale augmente et on ne peut plus garantir que la ligne épipolaire est dans l'image de la caméra mobile. Le formalisme de la redondance permet de résoudre ce problème en assurant que la tâche secondaire ne perturbe pas la bonne régulation de la tâche principale.

L'asservissement visuel à plusieurs tâches reposant sur la géométrie épipolaire nous permet donc de contrôler les déplacements de la caméra mobile le long de la ligne de vue issue du point ${}^{c_f}\mathbf{p}$. À tout instant, la ligne épipolaire correspondant à ${}^{c_f}\mathbf{p}$ est horizontale et centrée dans l'image de la caméra déportée. En d'autres termes, la ligne de vue passant pas $\mathbf{C}_f$ et ${}^{c_f}\mathbf{p}$ est centrée et horizontale dans l'image. L'objet à saisir est quelque part sur cette ligne de vue. La prochaine partie est consacrée à la recherche de l'objet le long de cette ligne.

3 Localisation de l'objet sur la ligne épipolaire

L'objet à saisir est au voisinage du point ${}^{c_f}\mathbf{p}$ dans l'image de la caméra déportée. On dispose donc d'une vue de l'objet à saisir. Les techniques classiques de reconnaissance d'objet peuvent alors être mises en oeuvre pour permettre la mise en correspondance de l'apparence de l'objet dans les plans image des caméras embarquée et déportée.
Dans un premier temps, nous rappelons rapidement la méthode de reconnaissance utilisée. Puis nous expliquons comment fusionner les données issues des différentes vues de la caméra mobile à l'aide d'une chaîne Bayésienne.

3.1 Extraction et mise en correspondance des points clef

Dans le cadre de cette étude, les deux caméras ont des angles de vues différents et des positions éloignées. L'objet apparaît avec une grande différence d'échelle et d'orientation. Il est donc nécessaire d'utiliser une méthode robuste à ces variations d'apparence.
Notre choix s'est porté, dans un premier temps, sur une mise en correspondance des points clef SIFT [15]. Ils sont connus pour leur propriétés d'invariance aux changements d'échelle et aux rotations. Ils permettent une mise en correspondance robuste aux transformations affines, aux changements de point de vue, au bruit additionnel, et aux changements d'illumination. En contrepartie, le temps de calcul des SIFT est assez élevé. Il existe d'autres méthodes de mise en correspondance de point d'intérêt moins coûteuses [12, 13]. Le propos de cet article n'est pas de développer une méthode de reconnaissance d'objet mais de montrer qu'une méthode de reconnaissance classique basée sur l'apparence permet retrouver l'emplacement de l'objet sur la ligne de vue.

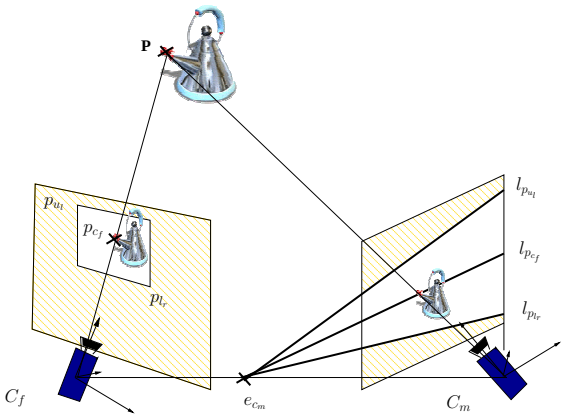

FIG. 3 – Sélection des zones d'intérêt dans les deux vues

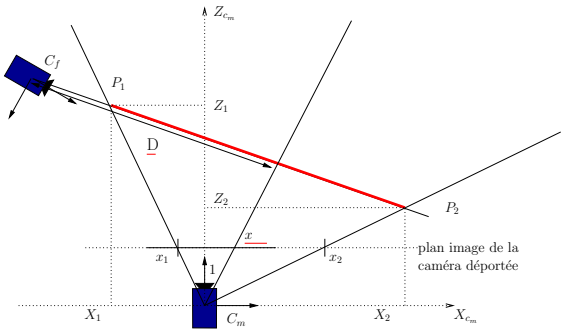

FIG. 4 – Vue du dessus : soit l'abscisse x d'un point dans la vue de la caméra embarquée, on peut alors calculer sa projection sur le segment 3D.

Pour réduire le temps de calcul, on limite l'extraction des points clef à des zones d'intérêt dans les deux images. L'objet est situé au voisinage du point ${}^{c_f}\mathbf{p}$ dans la vue de la caméra déportée. Pour délimiter la portion de l'image correspondant à la projection de l'objet à saisir, on peut s'appuyer sur l'hypothèse que l'objet se situe dans la zone de l'image qui a la plus forte densité de contours, à la manière de [2]. On peut également utiliser des méthodes de croissance de région, des contours dynamiques, des ensemble de niveaux ou encore calculer l'échelle intrinsèque de l'objet afin de segmenter l'objet au plus près. Dans la suite, la zone entourant l'objet est arbitrairement sélectionnée comme un rectangle couvrant 1/5 de l'image et centré sur l'objet. La zone considérée contient alors un motif comportant au moins une partie de l'objet et l'arrière plan à son voisinage. C'est ce motif que l'on va rechercher dans la vue de la caméra mobile pour estimer la profondeur de l'objet à saisir sur la ligne de vue issue de ${}^{c_f}\mathbf{p}$. Dans les images acquises par la caméra mobile, on limite la recherche de l'objet a la zone comprise entre les droites épipolaires extrêmes correspondant aux points du contour de la zone d'intérêt de l'image acquise par c_f(voir Figure 3).
La mise en correspondance est composée de trois étapes :

1. délimiter les zones d'intérêt dans les deux vues
2. extraire les points clef
3. sélectionner les points mis en correspondance

L'objet se situe dans la zone de plus forte concentration de points mis en correspondance. Les points clef donnent une bonne indication de la présence de l'objet et de sa profondeur sur la ligne de vue.

3.2 Un repère commun à toutes les vues : le segment 3D

La caméra mobile explore l'intersection de la ligne de vue issue de ${}^{c_f}\mathbf{p}$ et de l'espace de travail du bras, i.e. un segment 3D allant de longueur L. Elle acquiert des images tout a long de son déplacement et le processus de reconnaissance d'objet est appliqué à chacune de ces images. Ainsi, des points clef sont détecte et mis en correspondance a chaque pas de la boucle de commande, donnant une indication concernant la profondeur le l'objet. Il s'agit alors de conserver et d'utiliser ces informations. Ce paragraphe présente une méthode permettant de fusionner les données issues des différentes image afin d'estimer la profondeur $\widehat{P}$ de l'objet sur la ligne de vue.
Afin de conserver les informations au cours du temps, il est nécessaire de les représenter dans un repère commun. Nous avons choisi de projeter les points clef sur le segment 3D. Si un point clef appartient à l'objet alors il est porteur d'une information sur la position de l'objet. Les points mis en correspondance sont projetés sur le segment 3D (voir Figure 4).
Soient ${}^{c_m}\mathbf{P_1}$, (X_1,Y_1,Z_1) et ${}^{c_m}\mathbf{P_2}$, (X_2,Y_2,Z_2), les deux extrémités du segment. Dès que le centrage de la ligne épipolaire est assuré, les coordonnées du plan épipolaire dans le repère $\mathscr{R}_{c_m}$ sont (0,1,0,0). De plus, $\mathbf{C}_m$, et les deux points ${}^{c_m}\mathbf{P_1}$ et ${}^{c_m}\mathbf{P_2}$ sont sur le plan épipolaire. Alors Y_1 et Y_2, les coordonnées des points selon l'axe y du repère $\mathscr{R}_{c_m}$ sont nulles. Soit (x,y) les coordonnées d'un point clef ${}^{c_f}k$ mis en correspondance avec le motif de la caméra déportée. Soit ${}^{c_m}\mathbf{K}$, sa projection sphérique sur le segment de coordonnées $(X_k,0,Z_k)$ ((voir Figure 4).

$$\begin{aligned} X_k &= X_1 + (X_2 - X_1)D/L \\ Z_k &= Z_1 + (Z_2 - Z_1)D/L \end{aligned} \tag{9}$$

Cette équation établit la relation entre un point clef ${}^{c_f}k$ du plan image de la caméra mobile et sa profondeur D s'il provenait de l'objet à saisir. Afin de modéliser les erreurs de mesure issues des erreurs dans l'estimation des paramètres extrinsèques et intrinsèques du système, on représentera la projection du point par une densité de probabilité Gaussienne (μ,σ). La courbe Gaussienne est centrée sur D ($\mu = D$) et sa variance σ représente l'incertitude sur l'estimation de la profondeur.
Si on considère un ensemble de N points k_i, $i \in 1..N$, mis en correspondance et D_i la profondeur de leur projection sur le segment 3D, alors, on peut estimer la profondeur de l'objet dans cette vue par :

$$\widehat{P} = max\{\sum_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}(\frac{x-\delta_i}{\sigma})^2}\} \quad (10)$$

Pour chaque vue, la projection des points mis en correspondance donne un mélange de courbes Gaussiennes (voir Figure5). Ainsi chaque jeu de points mis en correspondance dans chaque image donne une estimation de la profondeur de l'objet $\widehat{P}$. Si aucun point n'est trouvé, alors, l'objet doit se trouver sur une partie du segment qui n'est pas vu depuis la position de la caméra considérée. Il reste maintenant à trouver une méthode pour utiliser les informations issues de toutes les vues afin d'affirmer ou d'infirmer les hypothèses de présence de l'objet le long de la ligne de vue.

3.3 Processus de décision Bayésien

La recherche de la profondeur d'un objet a partir d'une séquence de vue a été étudiée récemment dans [4] comme une phase d'initialisation de la profondeur des amers pour le la localisation et la cartographie simultanée. A la différence de notre approche, dans [4] l'amer est contenu dans toutes les images de la séquence. Dans cette études, il n'est pas possible de balayer tout le segment en une seul image. A chaque pas du processus de reconnaissance, certaines parties du segment ne sont pas visibles depuis la position courante. Notre méthode impose que le segment 3D soit balayé au moins une fois en entier pour estimer la profondeur de l'objet. Afin d'utiliser l'information extraite de toutes les vue, nous nous placerons dans le cadre de l' inférences Bayésienne. La formule de Bayes pour deux ensembles de variables aléatoires est :

$$p(B_i|A_j) = \frac{p(A_j|B_i)p(B_i)}{\sum p(A_j|B_i)p(B_i)} \quad (11)$$

où B_{ii} et $A_{j\,j}$ sont des ensemble d'évènements aléatoires indépendants, p est une densité de probabilité. $p(A_j|B_i)$ est la fonction de vraisemblance. Elle provient directement de la mesure. $p(B_i)$ est la probabilité a priori.$\sum p(A_j|B_i)p(B_i)$ est un facteur de normalisation. $p(B_i|A_j)$ est la densité de probabilité a posteriori.

La Figure 5 présente le processus de décision Bayésien qui permet l'estimation de la profondeur de l'objet au fur et a mesure du déplacement de la caméra embarquée. A l'instant initial, on ne dispose d'aucun a priori sur la position de l'objet sur le segment. La fonction de densité de probabilité a priori de présence de l'objet est donc uniforme sur le segment 3D et nulle en dehors. A chaque itération, on calcule une fonction de vraisemblance a partir du mélange de Gaussiennes généré par la projection des points mis en correspondance. Ces deux fonctions sont des fonctions de probabilité. On peut donc appliquer la formule de (Bayes 11) et calculer une fonction de probabilité a posteriori [5]. Cette fonction servira de probabilité a priori pour l'itération suivante du processus. Ainsi, on peut fusionner les informations issues de vues successives.

Soit D la profondeur de l'objet et p un densité de probabilité. Le théorème de Bayes donne alors à chaque itération t du processus :

$$p(D_{t+1}|k_{i_{t+1}}, D_{0..t}) = \frac{p(D_t|k_{i_t}, p(D_{0..t-1}))p(k_{i_{t+1}}|D_{t+1})}{\int_L p(D_t|k_{i_t}, p(D_{0..t-1}))p(k_{i_{t+1}}|D_{t+1})dD} \quad (12)$$

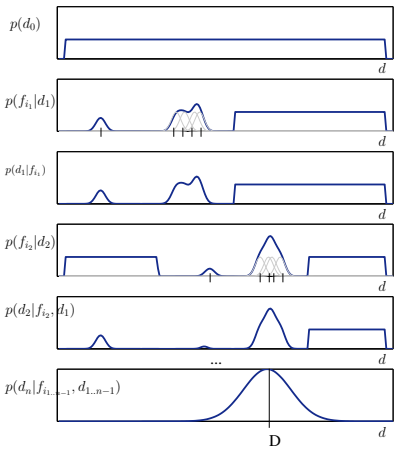

FIG. 5 – Processus Bayésien : 1)connaissance a priori à l'instant initial. La fonction de densité de probabilité (fdp) est uniforme sur le segment et nulle partout ailleurs. 2) fdp issue de la première mesure. 3) (12) permet de calculer la probabilité a posteriori, probabilité a priori pour l'itération suivante. 4) vraisemblance d'une nouvelle mesure 5) une fdp a posteriori est calculée a partir de 4) et 3) et ainsi de suite jusqu'à ce que la distribution soit Gaussienne. La profondeur de l'objet sur la ligne de vue peut être estimée en recherchant le maximum de la fdp a posteriori.

Dés que le segment a été parcouru entièrement, on peut obtenir une première estimation de la profondeur de l'objet en recherchant le maximum de la probabilité a posteriori. Pour affiner cette estimation, on peut parcourir le segment plusieurs fois en sous échantillonnant le voisinage de la profondeur estimée de l'objet a la manière des filtres a particules. Deux critères d'arrêts peuvent être utilises : on peut fixer arbitrairement un seuil sur la valeur du maximum a posteriori, ou sur une mesure de la forme de la distribution, telle que l'entropie de Shannon. Plus le nombre d'images prises le long de la ligne de vue est important, plus la précision de l'estimation est grande. Il faut trouver un compromis entre temps de calcul et précision de l'estimation.

Lorsque le critère d'arrêt est atteint, on dispose d'une estimation de la profondeur de l'objet. Une tâche de centrage de l'objet se substitue a la tâche de centrage des extrémités. L'objet appartient alors au champ de vision des deux caméras. Dans la partie suivante, nous présentons quelques résultats expérimentaux validant l'utilisation de cette méthode.

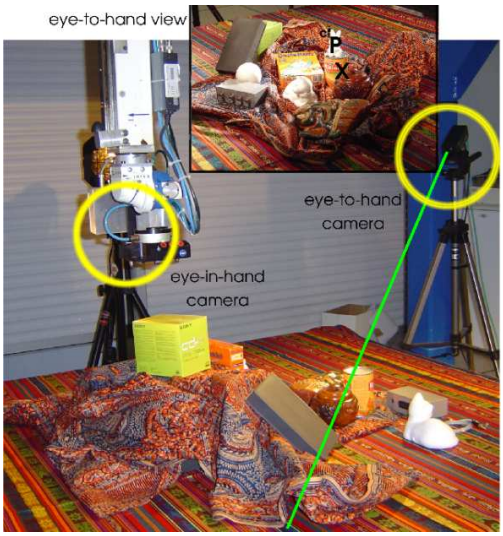

FIG. 6 – Dispositif expérimental : la scène est complexe. Les positions des deux caméras sont entourées d'un cercle jaune. En haut à droite : vue de la caméra déportée avec le point sélectionné ${}^{c_f}\mathbf{p}$. La ligne verte est la ligne de vue associée à ${}^{c_f}\mathbf{p}$

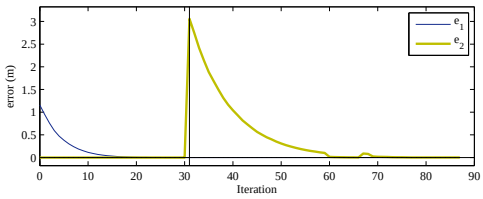

FIG. 7 – Évolution des erreurs des tâches pendant l'asservissement visuel pour un objet à saisir fixe.

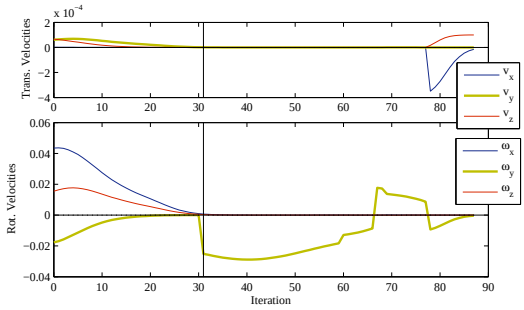

FIG. 8 – Vitesse de la caméra embarquée pendant l'asservissement visuel pour un objet à saisir fixe.

4 Résultats expérimentaux

Cette partie présente une application de la méthode de placement de la caméra embarquée à un système robotique commandé par vision hybride : caméra embarquée/caméra déportée. Le dispositif expérimental est présente Figure 1 et 6. La caméra embarquée est fixée sur l'organe effecteur d'un bras a 6 degrés de liberté. La caméra déportée est fixe et possède un objectif grand angle qui offre une large vue de la zone de travail du bras. La scène étudiée est complexe (cf. Figure 6), l'arrière plan est texture par une structure périodique. L'algorithme est déclenche par la sélection du point ${}^{c_f}\mathbf{p}$ dans l'image acquise par la caméra fixe.

Dans un premier temps la ligne épipolaire associée a ${}^{c_f}\mathbf{p}$est centrée, puis, quand l'erreur chute en deçà d'un certain seuil, la tâche secondaire est activée et la caméra parcourt le segment 3D. Des que le segment 3D entre dans le champ de vision de la caméra embarquée, le processus de reconnaissance de l'objet est lancé. La caméra embarquée parcourt le segment jusqu'à ce que la profondeur de l'objet soit estimée.

Tout d'abord nous présentons les résultats de l'asservissement visuel, puis les résultats du processus de détection de l'objet.

4.1 Un asservissement visuel basé sur la géométrie épipolaire

Les résultats de la commande sont présentés par les Figures 7, 8 and 9. La Figure 7 présente l'évolution des erreurs des deux tâches $\mathbf{e_1}$ et $\mathbf{e_2}$ au cours de l'asservissement visuel. La tâche principale de centrage de la ligne épipolaire associée a ${}^{c_f}\mathbf{p}$est activée. A l'itération 30, l'erreur de la tâche principale atteint un seuil suffisamment petit, fixé arbitrairement, et la tâche secondaire est activée. Le point a centrer est alors la projection $\mathbf{p}_1$ de $\mathbf{P}_1$, la première extrémité du segment. L'erreur de la tâche secondaire décroît alors tandis que la tâche principale est toujours régulée. A l'itération 60, l'extrémité $\mathbf{P}_1$ du segment est centrée dans le champ de vision ; le point à centrer est maintenant $\mathbf{P}_2$. L'erreur de la tâche secondaire augmente brutalement quand le point de référence change. Elle décroît ensuite exponentiellement jusqu'à ce que $\mathbf{P}_2$ soit centré dans le champ de vision de la caméra déportée. Si l'objet a été trouvé avec une confiance suffisante (seuil sur le maximum a posteriori ou sur l'entropie), le processus s'arrête. Sinon, le point de référence est a nouveau $\mathbf{P}_1$ et le segment est parcouru une nouvelle fois.

La Figure 8 donne les vitesses de l'organe effecteur du robot, i.e. de la caméra embarquée. Le lancement de la tâche secondaire aux itérations 30 et 60 entraîne, comme prévu, un mouvement de rotation pur autour de l'axe y du repère $\mathscr{R}_{c_m}$.

Une seconde expérience a été réalisée avec un objet mobile pour éprouver la robustesse de la commande proposée. Un algorithme de suivi utilisant un descripteur local d'apparence donne les coordonnées du point ${}^{c_f}\mathbf{p}$ à chaque pas de la boucle de commande. La ligne épipolaire et les points virtuels $\mathbf{P}_1$ et $\mathbf{P}_1$ sont donc recalculés à chaque itération. Ainsi les déplacements de l'objet sont pris en compte dans la boucle de commande.

A l'instant initial, lorsque la tâche principale est lancée, l'objet à saisir est immobile. L'objet est déplacé à l'itération 90 alors que la tâche principale n'est pas encore régulée. Il redevient immobile à l'itération 190 et se déplace à

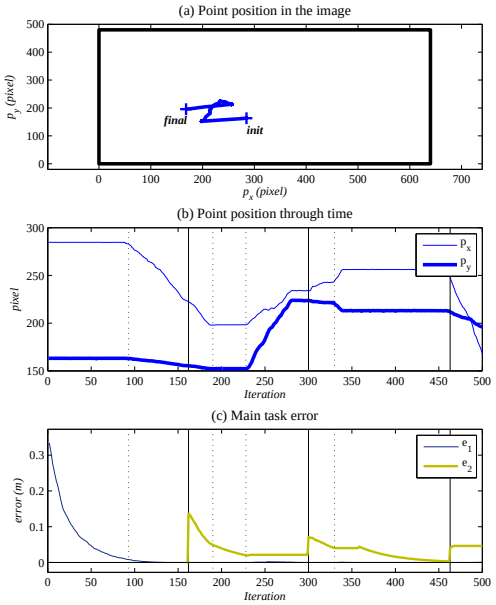

FIG. 9 – Asservissent visuel pour un objet mobile : le graphe du haut illustre le mouvement de ${}^{cf}\mathbf{p}$ dans le plan image de la caméra déportée. Le graphe du milieu présent l'évolution des coordonnées de ${}^{cf}\mathbf{p}$ dans le temps. Le graphe du bas illustre l'évolution des erreurs associées aux tâches

nouveau de l'itération 230 à l'itération 330. La tâche secondaire est déclenchée à l'itération 160 lorsque la tâche principale a convergé.
La tâche principale est très peu perturbée par le mouvement de l'objet. Comme le montre la Figure 9 la décroissance de l'erreur de la tâche principale est exponentielle jusqu'à l'itération 160 et est correctement régulée par la suite. La tâche secondaire est un peu plus perturbée par le déplacement. L'erreur décroît mais n'est pas immédiatement régulée à zéro. On observe alors un phénomène de traînage des itérations 200 à 330 (voir Figure 9). Dès que l'objet s'immobilise, l'erreur est à nouveau ramenée à zéro.

4.2 Estimation de la profondeur

Dés que le segment entre dans le champ de vision de la caméra déportée, l'algorithme de mise en correspondance est lancé et la recherche de l'objet commence. Les points clef SIFT sont extraits de l'image de la caméra déportée et mis en correspondance avec ceux extraits du motif de l'image de la caméra déportée. La fonction de vraisemblance de la profondeur de l'objet sur le segment est calculée à partir de la projection des points clef sur le segment 3D. La probabilité a posteriori est obtenu par (12). Le processus est itéré jusqu'à ce que le maximum a posteriori atteigne un seuil fixée arbitrairement à 0.5. La Figure 10 illustre l'évolution de la fonction de densité de probabilité a posteriori de la profondeur de l'objet au cours du temps. Un

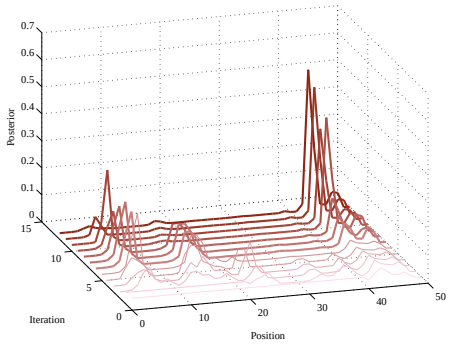

FIG. 10 – Estimation de la profondeur par un processus Bayésien de décision : évolution de la probabilité a posteriori au cours du temps. Ce graphe permet d'estimer la profondeur de l'objet sur la ligne de vue. Le segment 3D mesure $1,5m$ de long et a été échantillonnée en 50 parties. Après quelques itérations, un maximum apparaît autour de la position 43 qui correspond a une profondeur de $1.37m$.

maximum global apparaît rapidement et l'objet est trouvé en moins de 10 itérations du processus de reconnaissance (le segment est alors parcouru 2 fois). L'objet à saisir est ramené au centre de la vue de la caméra embarquée. Sa profondeur estimée est $1,37m$.

5 Conclusion

Cet article propose une phase d'initialisation pour un système de commande coopératif caméra déportée, caméra embarquée. Elle est la première étape vers un système semi autonome d'aide à la saisie d'objet sans a priori.

La caméra mobile est déplacée afin que l'objet à saisir soit centré dans son champ de vision. La commande est un asservissement visuel reposant sur la géométrie épipolaire du système. Ainsi, l'algorithme est robuste a de petits mouvements de l'objet. La profondeur de l'objet sur la ligne de vue issue du point ${}^{cf}\mathbf{p}$ est estimée en utilisant une mise en correspondance classique de points entre le motif sélectionné dans l'image de la caméra déportée et l'image de la caméra embarquée. Un processus Bayésien permet de fusionner les données obtenues sur l'ensemble des images acquises par la caméra déportée au fur et à mesure de son déplacement.

Cette méthode sera testée sur le système robotique d'aide à la saisie pour les personnes handicapées développé par le CEA dans le cadre des projets européen ITEA-ANSO et national AVISO.

Références

[1] R. Basri, E. Rivlin, and I. Shishoni. Visual homing : Surfing on the epipoles. *IJCV*, 33(2) :117–137, fevrier 1999.

[2] M. Becker et al. GripSee : A gesture-controlled robot for object perception and manipulation. *Autonomous Robots*, 6(2) :203–221, 1999.

[3] R. Cipolla and N. Hollinghurst. Visually guided grasping in unstructured environments. *Robotics and Autonomous Systems*, 19 :337–346, 1997.

[4] A.J. Davison, W. Mayol, and D. Murray. Real-time localisation and mapping with wearable active vision. In *ACM/IEEE ISMAR'03*, pages 18–27, Tokyo, octobre 2003.

[5] R. O. Duda, P.E. Hart, and Stork D. G. *Pattern Classification, second edition*. Wiley Interscience Publication, 2001.

[6] M. Elena, M Critiano, F. Damiano, and M. Bonfe. Variable structure pid controler for cooperative eye-in-hand/eye-to-hand visual servoing. In *IEEE. Int. Conf. on Control Applications, ICCA'03*, pages 989–994, Istambul, Turkey, 2003.

[7] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE Trans. on Robotics and Automation*, 8(3) :313–326, June 1992.

[8] G. Flandin, F. Chaumette, and E. Marchand. Eye-in-hand / eye-to-hand cooperation for visual servoing. In *IEEE ICRA'00*, pages 2741–2746, San Francisco, avril 2000.

[9] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge Univ. Press, 2001.

[10] R. Horaud, Knossow D., and M. Michaelis. Camera cooperation for achieving visual attention. *Machine Vision and Applications*, 16(6) :1–12, 2006.

[11] R. Horaud, F. Dornaika, and B. Espiau. Visually guided object grasping. *IEEE Trans. on Robotics and Automation*, 14(4) :525–532, aout 1998.

[12] F. Jurie and C. Schmid. Scale-invariant shape features for recognition of object categories. In *IEEE CVPR'04*, Washington, juillet 2004.

[13] V. Lepetit, P. Lagger, and P. Fua. Randomized trees for real-time keypoint recognition. In *Int. IEEE CVPR'05*, San Diego, juin 2005.

[14] V. Lippiello, B. Siciliano, and L. Villani. Eye-in-hand/eye-to-hand multi-camera visual servoing. In *IEEE CDC'05*, pages 5354 – 5359, Seville, Spain, decembre 2005.

[15] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2) :91–110, 2004.

[16] J. Piazzi, D. Prattichizzo, and N. J. Cowan. Auto epipolar visual servoing. In *IEEE IROS'04*, volume 1, pages 363–368, Sendai, octobre 2004.

[17] P. Rives. Visual servoing based on epipolar geometry. In *IEEE IROS'00*, pages 602–607, Takamatsu, novembre 2000.