

# Une loi de commande robuste aux mesures aberrantes en asservissement visuel

Andrew Comport, Muriel Pressigout, Éric Marchand, François Chaumette  
IRISA - INRIA Rennes  
Campus de Beaulieu, 35042 Rennes Cedex, France  
E-Mail: Prenom.Nom@irisa.fr

## Résumé

Les techniques actuelles d'asservissement visuel ne prennent pas en compte d'éventuelles mesures aberrantes, laissant souvent cet aspect au seul traitement d'images. Nous présentons dans cet article une loi de commande qui permet simultanément de réaliser la tâche d'asservissement visuel et de prendre en compte des données erronées. Cette méthode permet de considérer une grande variété d'erreur : bruit dans l'extraction des indices visuels, faibles erreurs de suivi, et importantes erreurs de mise en correspondance entre primitives courantes et désirées. La méthode retenue repose sur l'utilisation des techniques d'estimation statistique robuste et, en particulier, des M-estimateurs. Ces M-estimateurs sont introduits directement dans la loi de commande. Une étude sur la robustesse et la stabilité locale de cette nouvelle loi de commande est proposée. Des résultats expérimentaux sont finalement présentés et démontrent l'efficacité de la loi de commande pour des tâches de positionnement en présence d'erreurs plus ou moins importantes dans les données image.

## Mots Clef

Asservissement visuel, robuste, M-estimateurs.

## Abstract

*A fundamental step towards broadening the use of real world image-based visual servoing is to deal with the important issues of reliability and robustness. In order to address this issue a closed loop control law is presented that simultaneously accomplishes a visual servoing task and is robust to a general class of external errors. This generality allows concurrent consideration of a wide range of errors including: noise from image feature extraction, small scale errors in the tracking and even large scale errors in the matching between current and desired features. This is achieved with the application of widely accepted statistical techniques of robust M-estimation. The M-estimator is integrated by an iteratively re-weighted method. The Median Absolute Deviation is used as an estimate of the standard deviation of the inlier data and is compared with other methods. This combination is advantageous because of its high efficiency, high breakdown point and desirable influence functions. The robustness and sta-*

*bility of the control law is shown to be dependent on a subsequent measure of position uncertainty. Furthermore the convergence criteria of the control law are investigated. Experimental results are presented which demonstrate visual servoing tasks which resist severe outlier contamination.*

## Keywords

Visual servoing, robust, M-estimators.

## 1 Introduction

L'asservissement visuel [3, 10] consiste à contrôler les mouvements d'un système robotique à partir d'un ensemble d'informations visuelles extraites des images acquises par une caméra vidéo montée sur (ou observant) l'effecteur d'un robot. Une tâche classique d'asservissement visuel est, par exemple, de déplacer la caméra de façon à ce que l'image perçue corresponde à une image désirée. Ceci est réalisé en minimisant l'erreur entre deux vecteurs décrivant l'état de primitives visuelles calculés à partir des deux images. Bien que très efficace, il apparaît cependant que la précision de positionnement est très sensible aux erreurs inhérentes au processus d'extraction des données. L'efficacité de l'asservissement visuel dépend en effet de la précision de l'appariement entre la position courante et désirée de l'information visuelle. Si cette correspondance est entachée d'erreur, la tâche de positionnement sera au mieux imprécise, au pire impossible.

Traditionnellement, la robustesse d'une loi de commande est définie par : “*stability results which remain true in the presence of modeling errors or certain classes of disturbance*” [18]. Deux solutions peuvent donc être exhibées pour assurer la robustesse de la loi de commande : la première est de créer un modèle le plus précis possible du système considéré (perturbations potentielles comprises) et la seconde est de traiter (limiter) au mieux les perturbations. Dans le premier cas, il est raisonnable de penser qu'une modélisation et une estimation correcte des paramètres intrinsèques du système permettent d'améliorer les résultats. En asservissement visuel, ce type d'approche a conduit à modéliser la caméra par un modèle de projection perspective, à disposer d'une formulation analytique de la matrice d'interaction [3], à estimer en ligne l'informa-

tion de profondeur présente dans cette matrice [16, 6, 21], etc... D'autres sources d'erreurs proviennent du bruit dans l'extraction des indices visuels, ou d'erreurs de suivi voire d'importantes erreurs de mise en correspondance entre primitives courantes et désirées. La prise en compte de ces erreurs se fait le plus souvent en aval de la loi de commande, c'est-à-dire au niveau de l'extraction des indices visuels (voir Figure 1a) : amélioration de la qualité des algorithmes de suivis [22] ou sélection de primitives particulières [17], fusion d'informations redondantes (par des approches de vote ou de consensus [11]).

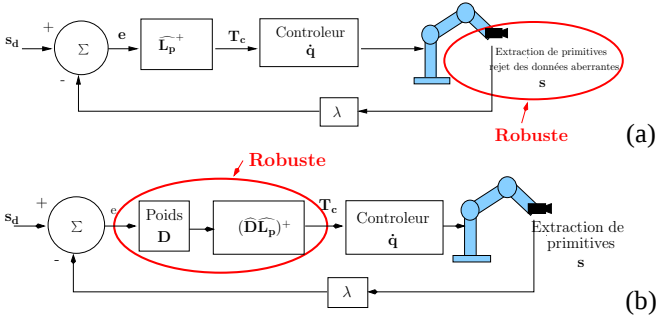


FIG. 1 – (a) Asservissement visuel robuste “classique” : le rejet des données aberrantes se fait dans l’extraction des données, (b) Nouvelle loi de commande : le rejet des données aberrantes se fait dans la loi de commande (voir Section 2.2 pour plus de détails)

Les solutions mentionnées dans le paragraphe précédent sont des solutions partielles pouvant prendre en compte un certain type d’erreur bien défini. Une séquence d’images acquises à la cadence vidéo est cependant une source quasi infinie d’erreurs qu’il est impossible de caractériser et de traiter de manière exhaustive. Ceci inclut les problèmes dus au mouvement plus ou moins rapide des objets, aux occultations éventuellement multiples, aux changements d’illumination, etc. Il semble évident qu’établir un catalogue analytique de toutes les sources de perturbation possibles et de proposer une solution pour traiter chacune d’entre elles est un travail complexe voire impossible à réaliser. Dans cet article, nous considérerons donc le problème de l’asservissement visuel robuste en introduisant directement dans la loi de commande des estimateurs robustes permettant de quantifier la confiance dans chacune des informations visuelles et, si nécessaire, de les rejeter (voir Figure 1b). L’incertitude sur chaque primitive est donc modélisée statistiquement, ce qui permet de prendre en compte tout type de perturbations dans l’extraction des données.

Dans la littérature portant sur les statistiques ou la vision par ordinateur, différentes approches ont été proposées pour considérer la détection et le traitement des sources de perturbation. L’approche la plus classiquement utilisée pour le rejet de données aberrantes est l’algorithme RANSAC [5]. Cet algorithme consiste à estimer les paramètres recherchés avec le minimum de mesures nécessaires puis à vérifier si d’autres mesures confirment cette première es-

timisation. Si un consensus est obtenu, l’estimation est retenue. D’autres méthodes reposent sur des approches statistiques. C’est le cas des moindres carrés médian (LMedS) ou des M-estimateurs [9]. Ces approches cherchent à estimer en ligne et de façon robuste l’écart type des “bonnes” mesures ou des mesures non aberrantes (à la différence de Ransac qui considère cet écart type comme une constante connue). Le lecteur pourra se référer à [20] pour une analyse des différentes techniques d’estimation robuste appliquées à la vision par ordinateur. Les approches de type Ransac ou LMedS se prêtant mal à une intégration dans une loi de commande d’asservissement visuel, nous proposons dans cet article une méthode reposant sur l’utilisation des M-estimateurs. Les M-estimateurs peuvent être considérés comme une formulation plus générale d’un estimateur au maximum de vraisemblance [9]. Ils sont plus généraux car ils permettent l’utilisation de différentes fonctions de minimisation qui ne correspondent pas nécessairement à une distribution normale des données. Un grand nombre de fonctions ont été proposées dans la littérature qui permettent de considérer comme peu vraisemblables des mesures incertaines et, dans certains cas, de les rejeter complètement. La loi de commande intégrant les M-estimateurs se formule de manière similaire à un algorithme d’estimation reposant sur les moindres carrés pondérés itérés (*Iteratively Re-weighted Least Square*). Afin d’améliorer la précision de détection des données aberrantes, la valeur de la variance du bruit de mesure (pour les données non aberrantes) sera estimée au cours de la minimisation. Pour cette estimation, puisque les données peuvent contenir des “outliers”, nous utilisons la valeur médiane de la déviation absolue (Median Absolute deviation ou MAD) qui représente un estimateur robuste de l’écart type du bruit de mesure.

Précisons que l’asservissement visuel d’un robot manipulateur est un problème très proche du problème de calcul de pose ou de localisation par des méthodes non-linéaires [12][19][1][4]. On peut en effet voir le calcul de pose comme un problème d’asservissement visuel virtuel. On considère l’image de l’objet comme le motif désiré à atteindre, le motif courant étant obtenu par la reprojection du modèle 3D de l’objet sur le plan image d’une caméra virtuelle [15]. Ces méthodes de calcul de pose non-linéaires sont très efficaces et permettent d’intégrer de façon naturelle des M-estimateurs [1][4].

Après avoir exposé les principes généraux de notre méthode, nous décrivons la nouvelle loi de commande en détail dans la section 2.2 où une matrice de pondération, reflétant la confiance que l’on peut affecter à chacune des données extraites des images, est introduite. La Section 2.3 détaille la manière de calculer les poids en utilisant des M-estimateurs robustes. Des résultats expérimentaux portant sur diverses tâches de positionnement sont finalement présentés dans la Section 3.

## 2 Asservissement visuel robuste

### 2.1 Principes et motivations

Nous considérons la tâche qui consiste à déplacer une caméra pour observer un objet à une position donnée dans l'image. Ceci est accompli en minimisant l'erreur  $\Delta$  entre un état désiré des primitives dans l'image ( $s^*$ ) et leur état courant ( $s$ ).

$$\Delta = s(r) - s^*. \quad (1)$$

où  $r$  représente la pose entre la caméra et l'objet d'intérêt. Si le vecteur des primitives visuelles est bien choisi, une seule position finale de la caméra permet d'accomplir cette minimisation.

L'hypothèse que  $s$  et  $s^*$  soient calculés (à partir des images) avec une précision suffisante est une supposition importante. En asservissement visuel, la loi de commande qui réalise la minimisation de  $\Delta$  est traitée habituellement par une approche aux moindres carrés [3, 10]. Cependant, s'il y a des données aberrantes, une estimation robuste est nécessaire. Comme évoqué en introduction, notre approche repose sur l'utilisation des M-estimateurs. La fonction à minimiser est modifiée afin de réduire la sensibilité aux données aberrantes. L'erreur à minimiser est alors donnée par :

$$\Delta_{\mathcal{R}} = \rho(s(r) - s^*), \quad (2)$$

où  $\rho(u)$  est une fonction robuste [9].

Afin de combiner une minimisation robuste avec un asservissement visuel, une modification de la loi de commande est nécessaire pour que l'algorithme puisse rejeter les données aberrantes. La méthode des Moindres Carrés Pondérés Itérés est une méthode bien connue pour l'utilisation du M-Estimateur dans un problème d'estimation. Elle convertit le problème d'estimation robuste en un problème équivalent à celui des moindres carrés pondérés. De façon similaire, nous introduisons dans la loi de commande une matrice de pondération, où les poids reflètent la confiance dans chaque primitive. La nouvelle loi de commande ainsi construite est présentée dans la section suivante. Le calcul des poids sera présenté dans la section 2.3.

### 2.2 Loi de commande robuste

En asservissement visuel classique, une fonction de tâche  $e$  permettant de minimiser l'erreur  $\|s - s^*\|$  est définie par la relation:

$$e = C(s(r) - s^*), \quad (3)$$

où la matrice  $C$  est appelée matrice de combinaison. Elle permet de prendre en considération plus de primitives visuelles que le nombre de degrés de liberté contrôlés.

Nous proposons une nouvelle fonction de tâche  $e$  qui assure une minimisation robuste de  $\Delta$  définie par l'équation (2). Elle est définie par la relation :

$$e = CD(s(r) - s^*), \quad (4)$$

où

- comme dans le cas classique la matrice  $C$  est une matrice de combinaison de taille  $m \times k$  et de rang  $m$  où  $k$  est le nombre de primitives visuelles et où  $m$  est le nombre de degrés de liberté contrôlés du robot (6 pour atteindre une unique position désirée),
- $D$  est une matrice diagonale définie par

$$D = \begin{pmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_k \end{pmatrix}$$

Le calcul du poids  $w_i$  associé à chaque information visuelle sera décrit dans la section 2.3.

En supposant  $C$  et  $D$  constantes, la dérivation de l'équation (4) conduit à :

$$\dot{e} = \frac{\partial e}{\partial s} \frac{\partial s}{\partial r} \frac{dr}{dt} = CDL_s v, \quad (5)$$

où  $L_s$  s'appelle la matrice d'interaction associée à  $s$  [3, 10]. Elle lie le mouvement des primitives à la vitesse  $v$  de la caméra :

$$\dot{s} = L_s v.$$

Si nous spécifions une décroissance exponentielle de l'erreur  $e$ , nous avons :

$$\dot{e} = -\lambda e, \quad (6)$$

où  $\lambda$  est un coefficient proportionnel qui fixe la vitesse de décroissance. Nous pouvons maintenant dériver la loi de commande. En effet, la combinaison de (6) et (5) nous donne :

$$CDL_s v = -\lambda e. \quad (7)$$

On en déduit une loi de commande idéale :

$$v = -\lambda(CDL_s)^{-1}e. \quad (8)$$

Dans la pratique, un modèle ou une approximation  $\widehat{L}_s$  de  $L_s$  et un modèle  $\widehat{D}$  de  $D$  sont utilisés, et nous obtenons :

$$v = -\lambda(C\widehat{D}\widehat{L}_s)^{-1}e. \quad (9)$$

Dans tous les cas considérés ici, la dimension  $k$  du vecteur des primitives visuelles  $s$  est supérieure à  $m$  (les primitives visuelles choisies sont redondantes). Puisque la matrice de combinaison doit être de dimension  $m \times k$  et de rang  $m$ , le choix le plus simple est de définir  $C$  comme la pseudo inverse  $(\widehat{D}\widehat{L}_s)^+$  de  $\widehat{D}\widehat{L}_s$ . Ceci implique  $C\widehat{D}\widehat{L}_s = (\widehat{D}\widehat{L}_s)^+\widehat{D}\widehat{L}_s = I_m$  et finalement:

$$v = -\lambda e = -\lambda(\widehat{D}\widehat{L}_s)^+ D(s(r) - s^*). \quad (10)$$

La convergence et la stabilité sont des questions importantes lorsque l'on applique une telle loi de commande. En utilisant (9) dans (10). Un comportement approché du système en boucle fermée est obtenu :

$$\dot{e} = -\lambda(CDL_s)e. \quad (11)$$

Sous l'hypothèse d'un modèle exact, la condition de positivité

$$(\mathbf{CDL}_s) > 0 \quad (12)$$

serait suffisante pour assurer la décroissance de  $\|\mathbf{e}\|$  et la stabilité asymptotique globale du système.

En posant  $\mathbf{C} = (\widehat{\mathbf{DL}}_s)^+$  la condition (12) peut être écrite sous la forme :

$$(\widehat{\mathbf{DL}}_s)^+ \mathbf{DL}_s > 0 \quad (13)$$

Cependant, pour obtenir (5), nous avons supposé que  $\mathbf{C}$  et  $\mathbf{D}$  étaient constantes, ce qui n'est pas le cas. Comme pour toute technique actuelle d'asservissement visuel 2D, la stabilité globale ne peut donc pas être démontrée. Il est possible de considérer la stabilité locale pour deux choix du modèle  $\widehat{\mathbf{DL}}_s$  [2] :

- $[\widehat{\mathbf{DL}}_s]^+ = [\mathbf{D}(\mathbf{s})\mathbf{L}_s(\mathbf{s}, \hat{\mathbf{r}})]^+$  : les poids et la matrice d'interaction sont calculés à chaque itération avec une estimation (si disponible) de la valeur actuelle de la pose  $\hat{\mathbf{r}}$  et des primitives visuelles  $\mathbf{s}$ . Le comportement du système est alors aussi proche que possible du comportement attendu ( $\dot{\mathbf{e}} = -\lambda\mathbf{e}$ ). Nous pourrions penser que la stabilité globale est démontrée avec  $(\mathbf{DL}_s)^+ \mathbf{DL}_s = \mathbf{I}_6 > 0$ , quelle que soit la valeur de  $\mathbf{s}$ . Cependant, dans ce cas, la matrice  $\mathbf{C}$  n'est pas constante et l'équation (5) devrait tenir compte de la variation de  $\mathbf{C}$ . Ceci mène à des calculs inextricables, et, seule la stabilité locale peut être obtenue.
- $[\widehat{\mathbf{DL}}_s]^+ = [\mathbf{L}_s(\mathbf{s}^*, \mathbf{r}^*)]^+$  : la matrice d'interaction est calculée seulement une fois avec les valeurs désirées de la pose  $\mathbf{r}^*$  et des primitives visuelles  $\mathbf{s}^*$ . Ce choix est le plus classique en asservissement visuel. La matrice de poids est, elle, fixée à l'identité :  $\widehat{\mathbf{D}} = \mathbf{I}_k$ . Ce choix est intéressant parce que  $(\widehat{\mathbf{DL}}_s)^+$  est calculée seulement une fois ce qui réduit le temps de calcul. La loi de commande est, elle, simplifiée :

$$\mathbf{v} = -\lambda \widehat{\mathbf{L}}_s^+ \mathbf{D}(\mathbf{s} - \mathbf{s}^*) \quad (14)$$

et on obtient finalement une condition suffisante de convergence plus simple :

$$\mathbf{L}_s(\mathbf{s}^*, \mathbf{r}^*)^+ \mathbf{DL}_s > 0. \quad (15)$$

Même si la matrice  $[\widehat{\mathbf{DL}}_s]^+ = [\mathbf{L}_s(\mathbf{s}^*, \mathbf{r}^*)]^+$  est constante, l'évolution des poids pendant la réalisation de la loi de commande est prise en compte au travers du calcul de  $\mathbf{e}$  par  $\mathbf{D}(\mathbf{s} - \mathbf{s}^*)$  (cf (14)). Par ailleurs, la condition de positivité (13) est satisfaite seulement dans un voisinage de la position désirée, ce qui démontre la stabilité locale du système.

Précisons qu'il est bien évidemment nécessaire de s'assurer qu'un nombre suffisant d'informations visuelles ne sera pas rejeté par les estimateurs robustes afin que  $\mathbf{DL}_s$  soit toujours de rang plein (6 pour contrôler les 6 degrés de liberté du robot). La pseudo-inverse  $(\mathbf{DL}_s)^+$  étant calculée via une décomposition SVD, on a facilement accès au rang de la matrice  $\mathbf{DL}_s$  ce qui permet de vérifier que le rang est

toujours supérieur à 6. Finalement, comme notre approche repose sur la redondance d'informations il est malheureusement impossible de la considérer dans le cadre de loi de commande 2D 1/2 [13] ou 3D puisque seulement 6 informations visuelles sont utilisées dans ce type d'approche.

Dans les résultats présentés par la suite, nous avons retenu la loi de commande avec une matrice d'interaction constante (équation (14)). Les expériences ont montré que, bien que la convergence ne puisse être théoriquement démontrée pour de grands déplacements à réaliser, celle-ci est en général obtenue quand le déplacement de la caméra a une erreur d'orientation inférieure à  $30^\circ$  sur chaque axe.

### 2.3 Calcul du degré de confiance

Les poids  $w_i$ , éléments de la matrice  $\mathbf{D}$ , reflètent la confiance en chaque primitive et sont souvent définis par [9] :

$$w_i = \frac{\psi(\delta_i/\sigma)}{\delta_i/\sigma} \quad (16)$$

où  $\psi(\delta_i/\sigma) = \frac{\partial}{\partial \mathbf{r}} \rho(\delta_i/\sigma)$ ,  $\psi$  est la fonction d'influence et  $\delta_i$  est le résidu normal donné par  $\delta_i = \Delta_i - \text{Med}(\Delta)$  ( $\text{Med}(\Delta)$  correspond à la valeur médiane des résidus). Le paramètre  $\sigma$  représente la valeur de l'écart type du bruit sur les "bonnes" mesures.

**Fonctions d'influence.** Parmi les diverses fonctions d'influence qui existent dans la littérature, nous avons retenu les fonctions de Huber et Tukey. La fonction d'influence de Huber diminue de façon asymptotique l'influence des données aberrantes. Cette fonction est donnée par :

$$\psi(u) = \begin{cases} u, & \text{si } |u| \leq a \\ a \frac{u}{|u|}, & \text{si } |u| > a \end{cases} \quad (17)$$

où le facteur de proportionnalité  $a$  est fixé à  $a = 1.2107$  ce qui représente 95% d'efficacité dans le cas d'un bruit Gaussien [8].

La fonction d'influence de Tukey rejette complètement les données aberrantes et leur donne un poids nul [9] (voir Figure 2). Il est en effet souhaitable que les données aberrantes n'aient **aucun** effet sur le mouvement de la caméra. Des erreurs dans l'image, même petites, peuvent en effet entraîner un écart très important dans la précision du positionnement final. Cette fonction d'influence est donnée par :

$$\psi(u) = \begin{cases} u(C^2 - u^2)^2, & \text{si } |u| \leq C \\ 0, & \text{sinon} \end{cases} \quad (18)$$

où le facteur de proportionnalité pour la fonction de Tukey est  $C = 4,6851$ . Ce facteur représente une efficacité de 95% dans le cas du bruit Gaussien.

Il existe d'autres fonctions d'influence robustes (Cauchy, Geman, McLure, ...) La figure 2 montre les fonctions de poids  $w(u)$  de ces différents M-estimateurs.

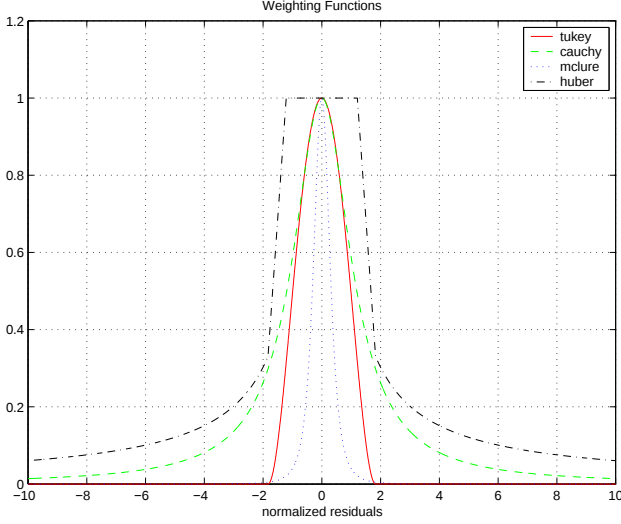


FIG. 2 – Fonction de poids  $w(u)$  pour différents M-estimateurs et pour  $\sigma = 0.5$

**Estimation de l'écart-type du bruit.** Afin d'obtenir une fonction robuste et d'améliorer la précision de détection des données aberrantes, il est nécessaire de définir un degré de confiance des mesures ou valeurs du bruit. Ceci revient à déterminer l'écart type du bruit sur les "bonnes" mesures  $\sigma$  (pour les données non aberrantes). Cette valeur  $\sigma$  (souvent appelée facteur d'échelle) peut varier énormément au cours du processus de minimisation.  $\sigma$  est souvent traitée comme une variable d'ajustement qui est choisie manuellement en fonction d'une application particulière. Dans notre cas, afin d'améliorer la précision de détection des données aberrantes, la valeur de  $\sigma$  est estimée parallèlement à la minimisation de l'erreur en utilisant une statistique robuste. Pour cette estimation, puisque les données peuvent contenir des "outliers", nous utilisons la valeur médiane de la déviation absolue (Median Absolute Deviation ou MAD) qui représente un estimateur robuste de l'écart type du bruit de mesure. Il est donné par :

$$\hat{\sigma} = \frac{1}{\Phi^{-1}(0.75)} \text{Med}_i(|\delta_i - \text{Med}_j(\delta_j)|). \quad (19)$$

où  $\frac{1}{\Phi^{-1}(0.75)} = 1,48$  permet d'avoir un biais nul sur l'estimation de  $\sigma$  lorsque la distribution des résidus est Gaussienne.  $\Phi(\cdot)$  est la fonction de distribution cumulée dans le cas Gaussien.

Malheureusement, une preuve de la convergence ne peut être obtenue que si l'on calcule la valeur de  $\sigma$  une seule fois. Ceci est dû à la propriété non asymptotique de la médiane [8]. Par contre, les expérimentations ont montré que le calcul de  $\sigma$  à chaque itération donne de meilleurs résultats.

### 3 Résultats expérimentaux

La validation de la loi de commande a été réalisée sur un banc d'essai expérimental composé d'une caméra CCD montée sur l'effecteur d'un robot AFMA à six degrés de liberté. La mise en œuvre a été faite en utilisant la plate-forme d'asservissement visuel ViSP développée à l'IRISA [14]. Nous avons considéré ici uniquement des tâches de positionnement contraignant les 6 ddl. D'une position initiale, le robot doit atteindre une position désirée exprimée sous la forme d'un motif à atteindre dans l'image.



FIG. 3 – Robot 6 axes de l'IRISA

**Primitives visuelles et calcul des poids.** Dans ces expériences, les primitives retenues sont les coordonnées de points extraits des images. Si  $n$  points sont pris en compte dans la loi de commande,  $s$  est un vecteur défini par  $s = (x_1, y_1, x_2, y_2, \dots, x_n, y_n)$  où  $(x_i, y_i)$  sont les coordonnées du  $i^{\text{ème}}$  point. La matrice d'interaction  $L_s$  est une matrice  $2n \times 6$  donnée par  $L_s = (L_{s1}, \dots, L_{sn})^T$  où :

$$L_{si} = \begin{pmatrix} -\frac{1}{z_i} & 0 & \frac{x_i}{z_i} & x_i y_i & -(1 + x_i^2) & y_i \\ 0 & -\frac{1}{z_i} & \frac{y_i}{z_i} & 1 + y_i^2 & -x_i y_i & -x_i \end{pmatrix}$$

Les pondérations sont calculées en utilisant l'équation (16) où

$$\forall k = 1 \dots n, \begin{cases} \Delta_{2k} = x_k - x_k^* \\ \Delta_{2k+1} = y_k - y_k^* \end{cases}$$

Les poids  $w_{2k}$  et  $w_{2k+1}$  reflètent la confiance que l'on peut avoir dans le même point  $k$ . Il est donc possible de définir les éléments de la matrice de pondération  $D$  par :

$$D_{2k,2k} = D_{2k+1,2k+1} = \min(w_{2k}, w_{2k+1}).$$

**Expériences avec une mire.** Dans ces premières expérimentations, nous considérons des tâches de positionnement reposant sur l'utilisation d'une mire très simple. Cette mire est composée de 12 points blancs sur fond noir. Considérer un tel objet permet de valider l'efficacité de la loi de commande. En effet, grâce à sa simplicité, si aucun bruit n'est introduit artificiellement dans la mise en correspondance ou le suivi, la tâche de positionnement se passera parfaitement (dans les limites de l'asservissement visuel).

classique [2]) et une expérience de “référence” est donc disponible (voir Figure 4).

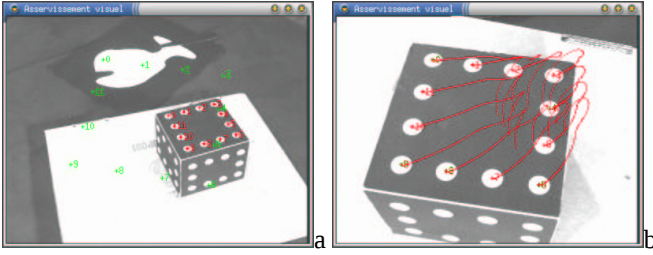


FIG. 4 – Asservissement visuel classique : images initiale et finale (les trajectoires rouges correspondent aux trajectoires des points dans les images et les croix vertes à la position désirée de ces points dans l’image.)

La première expérience considère des erreurs importantes dans l’appariement entre les points courants et désirés. Deux points (ceux numérotés 0 et 2 sur les figures 4 et 5) ont été volontairement inversés au moment de l’initialisation. Avec une loi de commande classique, l’asservissement visuel échoue (voir Figure 5). Les trajectoires, tant des points dans l’image que du manipulateur dans l’espace cartésien ne sont pas du tout satisfaisantes : on assiste en fait à des oscillations de la caméra et après 2500 itérations, il n’y a toujours pas eu de convergence vers une position finale (même fausse).

Une loi de commande robuste a ensuite été utilisée. Nous avons testé les estimateurs de Cauchy, Huber et Tukey. Cauchy et Huber détectent correctement la présence des données aberrantes mais comme l’illustre la courbe de la Figure 2, la valeur  $w_i$  des poids n’atteint jamais zéro. Toutes les mesures participent donc toujours (même faiblement) au calcul de la vitesse de la caméra, ce qui implique finalement une erreur non négligeable dans la précision de repositionnement (voir Figure 6a et Table 1, lignes 5 et 6). On a cependant, contrairement au cas précédent, une convergence de la caméra vers une position d’équilibre et, comme le montre la figure 5, l’erreur finale dans l’image est très faible. Nous avons ensuite considéré l’estimateur de Tukey avec ou sans estimation en ligne de l’écart type du bruit de mesure (MAD). Dans les deux cas, les deux points sont détectés comme étant des points aberrants, et plus vite que dans les deux cas précédents (en particulier quand le MAD  $\sigma$  est estimé). On peut surtout noter que les poids  $w_i$  deviennent nuls ce qui implique un repositionnement bien meilleur. On notera que l’erreur est même plus faible que dans le cas de l’asservissement visuel classique (où aucune erreur n’avait été introduite). Un troisième point (le point 4 sur la Figure 5) est en effet lui aussi considéré comme étant un “outlier”, son extraction est, semble-il, considérée comme entachée de bruit par les M-estimateurs. Ce résultat montre l’efficacité de la nouvelle loi de commande et souligne l’intérêt de l’estimateur de Tukey par rapport aux autres estimateurs.

La figure 7 montre les trajectoires du centre optique de

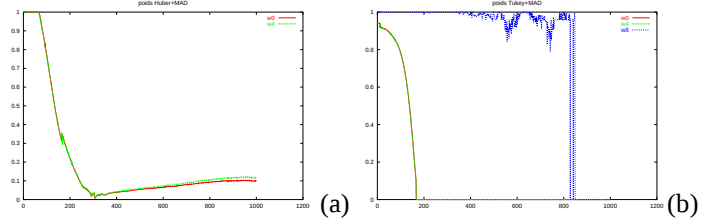


FIG. 6 – Évolution des poids des points considérés comme aberrants (Point 0, 2, et 5 : notons que seuls les points 0 et 2 ont été inversés)

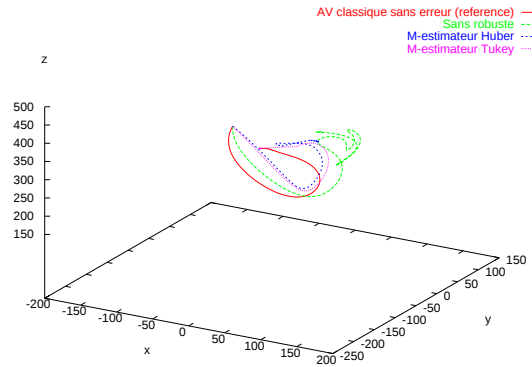


FIG. 7 – Trajectoire des caméras pour des erreurs importantes de mise en correspondance

la caméra pour ces différentes expériences : trajectoire de “référence” correspondant à un asservissement visuel classique sans “erreur” (en rouge). Après l’introduction de l’erreur de mise en correspondance, la trajectoire verte correspond à l’asservissement visuel classique (cette trajectoire est très bruitée, instable et ne converge pas vers une position d’équilibre). La trajectoire bleue correspond à l’introduction d’un estimateur de Huber dans la loi de commande (les résultats sont similaires avec l’estimateur de Cauchy), et s’il y a convergence de la caméra vers une position d’équilibre, ce n’est malheureusement pas la position désirée. Finalement la trajectoire rose correspond à l’introduction de l’estimateur de Tukey. Cette trajectoire n’est évidemment pas identique à celle de l’asservissement visuel classique (rouge) puisque les points aberrants ne sont pas détectés dès la première itération. Par contre la position finale est à quelques millimètres près la même. Le tableau 1 synthétise les erreurs de repositionnement pour ces différentes expériences.

Une seconde expérience montre l’influence de petites erreurs (qui peuvent simuler des erreurs de “dérive” dans un processus de suivi). Quatre points sont partiellement occultés sur les images acquises pendant l’asservissement ce qui modifie la position de leur centre de gravité (voir Figure 8) alors que l’image désirée (Figure 4a) a été acquise sans occultation. Même si l’erreur est très faible (de l’ordre de 3 à

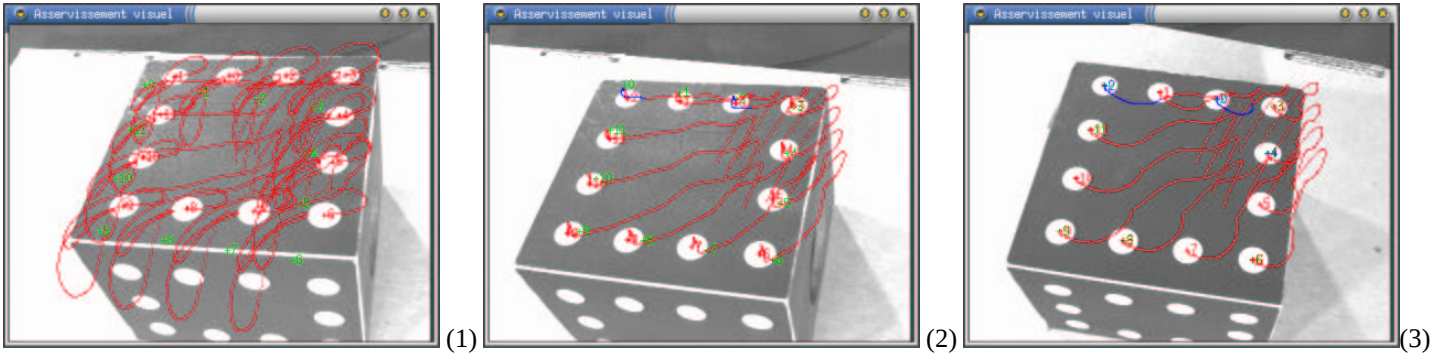


FIG. 5 – Importantes erreurs de mise en correspondance sur deux points (les points 0 et 2 ont été volontairement inversés). Images finales acquises par la caméra : (a) sans méthode robuste (non convergence), (b) avec la fonction de Huber, (c) avec la fonction de Tukey

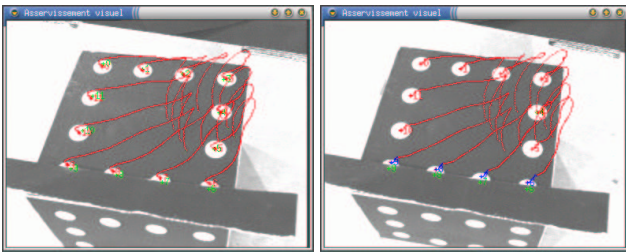


FIG. 8 – Petites erreurs dans l'extraction des centres de gravité de quatre points (images acquises depuis la position finale atteinte par la caméra (sans robuste et avec un estimateur de Tukey))

4 pixels pour chacun de ces points) l'erreur finale de positionnement avec une loi de commande classique est importante (près de 3 cm en translation et 4 degrés en rotation, voir Tableau 1). Nous avons considéré uniquement l'estimateur de Tukey. Si la variance du bruit de mesure (MAD) n'est pas recalculée à chaque itération, alors, un résultat similaire à celui de l'asservissement classique est obtenu puisque les points aberrants sont presque impossibles à détecter lors de la première itération (les erreurs sont trop petites). Par contre en recalculant  $\hat{\sigma}$ , les "outliers" sont correctement détectés après quelques itérations (l'erreur  $\delta$  diminuant, les erreurs deviennent statistiquement significatives) et le positionnement final est correct (voir Figure 9).

**Suivi de points par corrélation.** L'utilisation de mires avec des points permet l'introduction aisée de bruit et de perturbations "artificielles" dans le suivi ou dans la mise en correspondance et permet, par voie de conséquence, la validation de la loi de commande proposée. En pratique le suivi de telle mire est très fiable et il est peu probable que des erreurs aussi importantes se produisent. C'est pourquoi nous considérons maintenant une tâche de positionnement reposant sur un algorithme de traitement d'image (suivi par corrélation ou SSD) moins fiable mais pouvant être utilisé sur des images quelconques plus complexes.

La position à atteindre est encore exprimée sous la forme

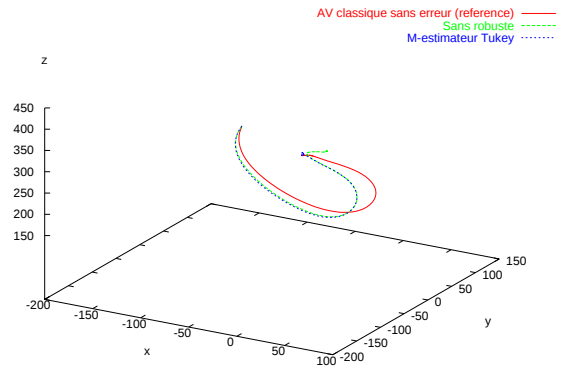


FIG. 9 – Trajectoire des caméras pour de petites erreurs dans l'extraction des centres de gravité de certains points

d'une image de référence. Des points d'intérêts sont extraits de cette image (en utilisant un détecteur de Harris [7]) et mis en correspondance avec un ensemble de points (extraits de façon similaire) dans l'image acquise depuis la position initiale. La mise en correspondance est réalisée en utilisant le logiciel Image-Matching [23]. Le suivi de ces points pendant l'asservissement visuel se fait en utilisant un algorithme classique de suivi par corrélation. Afin de "compliquer" l'expérience, la caméra placée habituellement sur l'effecteur de notre manipulateur (et utilisée pour les expériences précédentes) a été remplacée par une simple *web cam* à 30 euros qui fournit une image de piètre qualité. Les Figures 10a et 10b montrent l'image initiale acquise par la caméra et l'image désirée.

Dans ces conditions, le suivi n'est pas fiable (certains points "dérivent" de plusieurs pixels et d'autres sont perdus) ce qui met en échec une loi de commande classique d'asservissement visuel. Même si les points définitivement perdus (ce qui est facilement détectable) sont retirés du processus de minimisation, la caméra ne peut atteindre la position désirée. La Figure 10c montre l'image de différence

| Expérience                       | $T_x$  | $T_y$  | $T_z$ | $R_x$ | $R_y$ | $R_z$  | $\ \mathbf{T}\ $ |
|----------------------------------|--------|--------|-------|-------|-------|--------|------------------|
| 1 position initiale              | -155.2 | 124.4  | 218.2 | 30.56 | 46.84 | -0.98  |                  |
| 2 position désirée               | 61.1   | -185.2 | 441.5 | 0.0   | 29.45 | 0      |                  |
| 3 AV classique sans erreur       | -0.9   | -0.3   | 0.6   | -0.29 | -0.04 | 0.16   | 1.12             |
| 4 Grandes erreurs sans robuste   | 47.9   | 69.1   | 19.3  | 6.23  | 10.54 | -13.73 | 86.26 (E)        |
| 5 Grandes erreurs Cauchy         | 63.7   | 31.1   | 9.7   | 4.1   | 5.22  | -14.03 | 71.54            |
| 6 Grandes erreurs Huber          | 56.1   | 45.2   | 10.4  | 3.42  | 8.25  | -12.3  | 72.79            |
| 7 Grandes erreurs Tukey sans MAD | -0.2   | -0.3   | 0.6   | -0.27 | 0     | 0.01   | 0.7              |
| 8 Grandes erreurs Tukey avec MAD | -0.3   | 0.2    | 0.5   | -0.27 | 0.02  | 0.02   | 0.6              |
| 9 Petites erreurs sans robuste   | 18     | 17.6   | 7.4   | 0.9   | 3.72  | -3.97  | 26.23            |
| 10 Petites erreurs Tukey MAD     | -1.2   | 0.7    | 0.4   | -0.29 | -0.05 | 0.25   | 1.44             |

TAB. 1 – Synthèse des résultats de positionnement (erreur en mm et degré pour les différentes expériences décrites)

entre l'image finale désirée (Figure 10b) et l'image acquise depuis la position "finale" de la caméra (Figure 10d). On peut aussi voir que la caméra est très loin de sa position désirée. Une loi de commande robuste a ensuite été utilisée en considérant deux M-estimateurs différents : Huber et Tukey. Dans les deux cas, nous avons estimé en ligne le niveau de bruit (MAD) à chaque itération. Les résultats sont similaires (il est cependant difficile de les comparer précisément, l'algorithme de suivi de points étant sensible aux conditions expérimentales (éclairage, etc.) et donc non déterministe). Il reste que l'estimateur de Tukey est de manière générale préférable puisqu'il permet de rejeter totalement les points aberrants ( $w_i = 0$ ) ce qui n'est pas le cas des autres M-estimateurs. Dans les deux cas, la position désirée est atteinte avec une précision satisfaisante (moins de 3cm d'erreur en translation) malgré des conditions expérimentales très dégradées. Les différences entre les images finale et désirée sont très faibles (voir Figures 10e, 10g) ce qui implique un bon positionnement.

Précisons que sur les Figures 10d, 10f et 10h les croix rouges représentent la position initiale des points suivis et les croix bleues leur position désirée. Les croix vertes représentent leur position finale. Les trajectoires des points apparaissent en rouge (40 points sont initialement suivis).

## 4 Conclusion

Dans cet article, nous avons proposé une nouvelle loi de commande pour l'asservissement visuel qui permet le rejet de données visuelles aberrantes. Classiquement, ce problème de robustesse aux erreurs de suivi ou de mise en correspondance est traité au niveau de l'extraction des informations visuelles. Notre approche se base, elle, au niveau de la loi de commande (et ne remet pas en cause, par ailleurs, un travail toujours possible sur la robustesse de l'extraction des informations). Pour parvenir à cet objectif, nous avons introduit des M-estimateurs dans la loi de commande permettant de minimiser l'erreur dans l'image. Celle-ci se comporte alors de façon similaire à une minimisation par moindres-carrés pondérés itérés. Nous avons analysé la stabilité locale de notre loi de commande et démontré son efficacité pour des tâches de positionnement en

considérant d'une part des images simples puis des images réelles avec un traitement d'images plus complexe. Dans les deux cas, une amélioration nette de la précision de repositionnement a pu être observée.

Dans cet article, nous avons uniquement considéré des points comme primitives visuelles. Des primitives plus complexes peuvent aussi être utilisées afin de rendre la loi de commande plus robuste. En asservissement visuel virtuel et pour une application de réalité augmentée [4], nous avons utilisé avec succès cette loi de commande en considérant comme primitives visuelles des distances de points à droite ou de points à portions d'ellipse. Ce type d'information pourrait aussi être considéré dans une expérimentation d'asservissement visuel classique.

## Références

- [1] G.A. Borges, M.-J. Aldon, T. Gil. An optimal pose estimator for map-based mobile robot dynamic localization: Experimental comparison with the EKF. In *IEEE Int. Conf. on Robotics and Automation, ICRA'01*, pages 1585–1590, Seoul, Corée du sud, mai 2001.
- [2] F. Chaumette. Potential problems of stability and convergence in image-based and position-based visual servoing. In D. Kriegman, G. Hager, A.S. Morse (eds), *The Confluence of Vision and Control*, pages 66–78. LNCIS Series, No 237, Springer-Verlag, 1998.
- [3] F. Chaumette. Asservissement visuel. In W. Khalil (ed), *La commande des robots manipulateurs*, Traité IC2, chapitre 3, pages 105–150. Hermès, 2002.
- [4] A. Comport, E. Marchand, F. Chaumette. A real-time tracker for markerless augmented reality. In *ACM/IEEE Int. Symp. on Mixed and Augmented Reality, ISMAR'03*, pages 36–45, Tokyo, Japon, Octobre 2003.
- [5] N. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography. *Communication of the ACM*, 24(6):381–395, Juin 1981.

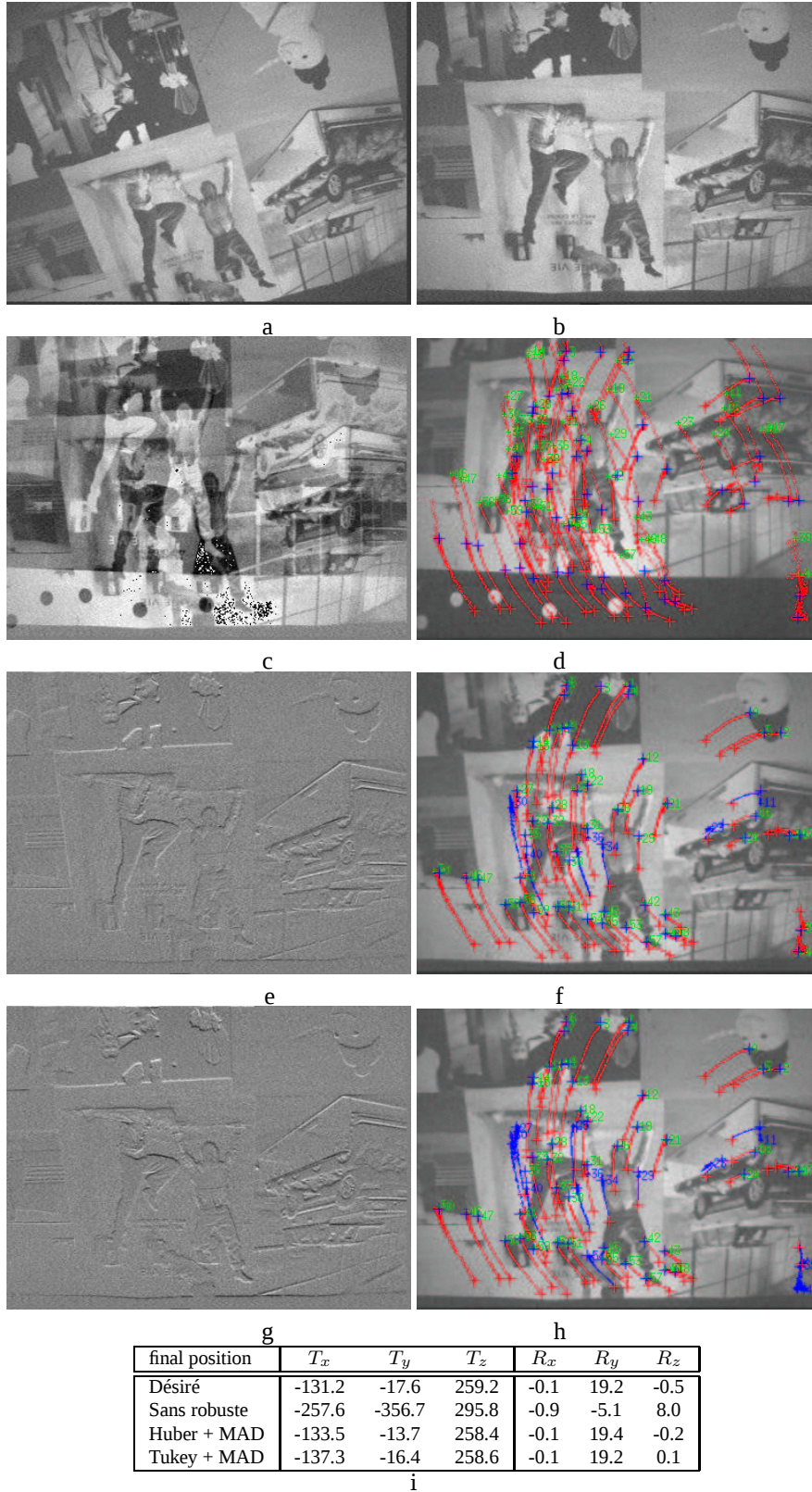


FIG. 10 – Tâche de positionnement avec extraction et suivi de points d'intérêt par corrélation: (a) image initiale, (b) image désirée, (c-d) résultats avec une loi de commande classique (image de différence et trajectoire des points sur l'image finale), (e-f) Asservissement visuel robuste avec l'estimateur de Huber, (g-h) Asservissement visuel robuste avec l'estimateur de Tukey, (i) erreur de repositionnement (mm et dg)

- [6] E. Grosso, G. Metta, A. Oddera, G. Sandini. Robust visual servoing in 3-d reaching tasks. *IEEE Trans. on Robotics and Automation*, 12:732–742, Octobre 1996.
- [7] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Conference*, pages 189–192, 1988.
- [8] P.-W. Holland and R.-E. Welsch. Robust regression using iteratively reweighted least-squares. *Comm. Statist. Theory Methods*, A6:813–827, 1977.
- [9] P.-J. Huber. *Robust Statistics*. Wiler, New York, 1981.
- [10] S. Hutchinson, G. Hager, P. Corke. A tutorial on visual servo control. *IEEE Trans. on Robotics and Automation*, 12(5):651–670, Octobre 1996.
- [11] D. Kragic and H. Christensen. Cue integration for visual servoing. *IEEE Trans. on Robotics and Automation*, 17(1):19–26, Février 2001.
- [12] D.G. Lowe. Robust model-based motion tracking through the integration of search and estimation. *Int. Journal of Computer Vision*, 8(2):113–122, 1992.
- [13] E. Malis, F. Chaumette, S. Boudet. 2 1/2 D visual servoing. *IEEE Trans. on Robotics and Automation*, 15(2):238–250, April 1999.
- [14] E. Marchand. Visp: A software environment for eye-in-hand visual servoing. In *IEEE Int. Conf. on Robotics and Automation, ICRA'99*, volume 4, pages 3224–3229, Detroit, Michigan, Mai 1999.
- [15] E. Marchand and F. Chaumette. Virtual visual servoing: a framework for real-time augmented reality. In *EUROGRAPHICS'02 Conf. Proc.*, volume 21(3) of *Computer Graphics Forum*, pages 289–298, Saarbrück, Allemagne, Septembre 2002.
- [16] S.-J. Maybank and O. Faugeras. A theory of self calibration of a moving camera. *Int. Journal of Computer Vision, IJCV*, 8(1):123–152, 1992.
- [17] N. P. Papanikolopoulos and P. K. Khosla. Selection of features and evaluation of visual measurements for 3-d robotic visual tracking. *Int. Symp. on Intelligent Control.*, pages 320–325, Août 1993.
- [18] C. Samson, M. Le Borgne, B. Espiau. *Robot Control: the Task Function Approach*. Clarendon Press, Oxford, Royaume-Uni, 1991.
- [19] G. Simon and M.-O. Berger. A two-stage robust statistical method for temporal registration from features of various type. In *IEEE Int. Conf. on Computer Vision, ICCV'98*, pages 261–266, Bombay, Inde, Janvier 1998.
- [20] C.-V. Stewart. Robust parameter estimation in computer vision. *SIAM Review*, 41(3):513–537, Septembre 1999.
- [21] C.-J. Taylor, J.-P. Ostrowski, S.-H. Jung. Robust visual servoing based on relative orientation. *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'99*, pages 574–580, 1999.
- [22] T. Tommasini, A. Fusiello, E. Trucco, V. Roberto. Making good features track better. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 178–183, Santa Barbara, USA, Juin 1998.
- [23] Z. Zhang, R. Deriche, O. Faugeras, Q.-T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78:87–119, Octobre 1995.