Calibration multiplanaire d'une caméra : augmenter la stabilité en utilisant la sélection de modèles

Javier-Flavio Vigueras-Gomez, Marie-Odile Berger, Gilles Simon

Calibration multiplanaire d'une caméra: Augmenter la stabilité en utilisant la sélection de modèles

Multi-planar Camera Calibration: Improving Stability using Model Selection

J.F. Vigueras [1] M.O. Berger[1] G. Simon[1]

[1] LORIA/INRIA Lorraine

LORIA BP 239, 54506 Vandoeuvre-lès-Nancy, France

e-mail: {vigueras,berger,gsimon}@loria.fr

Résumé

L'objectif de la réalité augmentée est d'intégrer de façon réaliste des objets virtuels dans des séquences vidéo. Afin d'atteindre cet objectif, il est nécessaire d'avoir un bon alignement entre objets réels et objets virtuels: les objets virtuels doivent en effet être incrustés dans la scène avec une caméra virtuelle identique à la caméra effectuant les prises de vue. Pour de nombreuses applications à caractère interactif et temps réel, cette estimation des paramètres de la caméra doit de plus être effectuée de manière séquentielle. Malheureusement, les points de vue calculés sont souvent affectés de fluctuations statistiques, ce qui nuit à l'impression visuelle de la scène augmentée. Dans le cadre d'environnements de type multiplanaire, nous proposons d'utiliser des méthodes de sélection de modèles pour améliorer la précision et la stabilité des trajectoires obtenues. Des vidéos montrant l'intérêt de ce travail sont disponibles à l'URL:
http://www.loria.fr/~vigueras/orasis2003.html

Mots Clés

Calibration planaire, sélection de modèles, réalité augmentée.

Abstract

*The main objective of augmented reality is to combine virtual representations of objects and real images with an acceptable level of realism. Hence, it is necessary to reach geometric coherence between virtual and real objects, so that camera calibration becomes an important task. Commonly, it exhibits some fluctuations and lack of precision in the estimation of the parameters because of noise and approximations at the data extracted from images. Based on sequential planar calibration, we suggest some additions to improve simultaneously precision and stability: using several planes and model selection.Videos demonstrating the efficiency of the methods are available at:
http://www.loria.fr/~vigueras/orasis2003.html*

Keywords

Planar calibration, model selection, augmented reality.

1 Introduction

L'augmentation de flux vidéo avec des objets de synthèse est le but de très nombreuses applications telles que les visites virtuelles, l'aide à la maintenance, le design architectural ou les systèmes d'apprentissage [15]. Toutes ces applications nécessitent que la scène augmentée soit continuellement mise à jour en fonction des mouvements de l'utilisateur ou de la caméra dans la scène. Il est donc primordial de pouvoir calculer à chaque instant les paramètres de la caméra pour restituer les objets de synthèse avec les paramètres calculés pour avoir une composition réaliste.
Dans ce papier, nous considérons le problème du calcul du point de vue pour des applications de nature interactive et temps réel. Bien que le problème du calcul du point de vue ait reçu beaucoup attention dans la communauté vision, on est loin d'une solution précise et robuste dans le cadre séquentiel. Idéalement, un système de RA devrait fonctionner dans n'importe quel environnement, sans besoin de préparer la scène ni de restreindre les mouvements de l'utilisateur. Des systèmes temps réel ont été proposés [7] mais ils nécessitent la plupart du temps la présence de marqueurs permettant l'identification simple et rapide d'indices et donc un calcul rapide et sûr du point de vue. Ces contraintes restreignent évidemment considérablement l'effectivité de ces systèmes. Il y a donc un réel besoin concernant la recherche de méthodes fonctionnant en environnements quelconques et ne nécessitant pas de connaissances sur la scène difficiles à acquérir.

1.1 État de l'art

Les approches pour le calcul séquentiel du point de vue peuvent être divisées en deux grandes catégories: les approches à base de modèle et les approches de type move-matching. Les approches à base de modèle reposent sur l'identification dans les images d'indices dont le modèle 3D est connu. Le point de vue pour chaque image est donc directement calculé par rapport au repère global dans lequel est exprimé le modèle [7, 11, 9]. Cette capacité à traiter les images indépendamment rend cette méthode attractive pour des applications de nature séquentielle. De plus, elle évite les problèmes de dérive du calcul du point de vue au cours du temps. Cependant, seul un petit nombre de pri-

mitives est en général disponible pour effectuer le recalage. Ceci rend le calcul du point de vue assez sensible au bruit sur les indices image ainsi qu'à la disparition ou l'apparition des primitives au cours de la séquence. Enfin, il faut souligner que l'acquisition de connaissances 3D sur l'environnement est en général assez fastidieux.

A l'opposé, les méthodes de move-matching tentent de calculer le mouvement relatif entre deux images successives, par exemple en utilisant des structures planaires [10]. Si la position de la caméra pour le premier repère image est connue, la position absolue de la caméra pour une image quelconque est obtenue par composition des mouvements relatifs antérieurement calculés. Ces systèmes sont attractifs car ils ne nécessitent que peu ou pas de connaissances sur l'environnement. Cependant, ils souffrent souvent d'un problème de dérive car les erreurs s'accumulent au cours du temps.

Un bon moyen d'apprécier la qualité du point de vue calculé est de considérer l'impression visuelle de l'utilisateur sur la scène augmentée. Malgré les progrès réalisés dans le domaine du calcul du point de vue, des fluctuations statistiques sur le point de vue conduisent à des effet de sautillements ou à des effets de glissement de l'objet incrusté dans son environnement. Ce problème est particulièrement perceptible quand le mouvement de la caméra est lent, le bruit sur les primitives extraites conduisant à des fluctuations assez grandes du point de vue. Ce problème de stabilisation de la caméra a été considéré dans [4]. L' idée de Kanatani et Matsuaga était de classifier les mouvements de la caméra par un certain nombres de modèles (stationnaire, panoramique, zoom,...) de façon à fixer un certain nombre de paramètres. Ainsi, la stabilité et la précision du point de vue s'améliore a priori puisque le nombre de degrés de liberté de la fonction à optimiser est moindre.

Dans ce papier, nous reprenons cette idée et nous apportons les contributions suivantes:

(i) nous proposons une méthode pour le calcul du point de vue basée sur l'observation de plusieurs plans dans la scène. De telles structures sont très courantes en intérieur mais également dans des environnements de type urbain et le domaine d'application de cette méthode est donc assez large. (ii) poursuivant l'approche de Matsunaga et Kanatani, nous étudions les performances d'un certain nombre de critères de sélection de modèles différents de ceux envisagés dans leur approche (iii) nous proposons une amélioration de la méthode de sélection de modèle qui utilise la persistance temporelle du choix d'un modèle sur plus de deux vues, ce qui améliore la stabilité de la trajectoire calculée.

Le schéma général de notre approche du calcul du point de vue est présenté en section 1.2. Dans cette section, les relations géométriques de base utilisées dans ce papier sont également rappelées. La méthode utilisant la structure multiplanaire de la scène pour calculer le point de vue est décrite en section 2. Les tests des critères de sélection ainsi que la méthode de sélection du modèle sont décrites dans les sections3 et 4. Enfin, des résultats de scène augmentée viennent conclure cet article.

1.2 Schéma général

Cette section décrit les grandes lignes de notre méthode de recalage (Fig. 1). Nous supposons que la structure multiplanaire de la scène est décrite par un ensemble de polygones 3D $\mathcal{L}_p(1 \leq p \leq n)$. Nous supposons ici que les paramètres intrinsèques de la scène sont constants et sont déterminés préalablement. Enfin, la position de la caméra pour la première image de la séquence est déterminée, par exemple en utilisant un poster rectangulaire fixé dans la scène. Ceci suffit en effet à calculer la position de la caméra au début du processus.

Ces étapes de prétraitement accomplies, le processus de recalage décrit une boucle comportant quatre étapes:(i) les points d'intérêt sont extraits et mis en correspondance entre deux images, (ii) les homographies correspondant à chacun des plans observés sont calculés en utilisant une estimation robuste de type RANSAC (iii) le mouvement de la caméra est choisi en utilisant les critères de sélection de modèles (iv) le point de vue de la caméra est réestimé en considérant le modèle de mouvement sélectionné. Dans la suite de ce papier, chacune de ces étapes est décrite en détail.

Etape de prétraitement:

1. Fournir l'équation des plans observés et utilisés pour le recalage.
2. Calculer la matrice de projection pour la première image $\mathtt{P}^0$,

Calcul de la matrice de projection $\mathtt{P}^i$ pour $i > 0$:

1. Déterminer l'ensemble des points d'intérêt en correspondance entre les images $i-1$ et i pour chaque plan observé.
2. Calculer les homographies induites par chaque plan entre $i-1$ et i.
3. Déterminer le meilleur modèle de mouvement selon le critère de sélection qui réalise un compromis entre précision et simplicité du modèle
4. Calculer la matrice P_i à partir de P_{i-1} et des homographies en utilisant le modèle sélectionné.

FIG. 1 – *Schéma général de la méthode de recalage multiplanaire basée sur la sélection de modèle*

1.3 Relations géométriques de base

Nous supposons que la position, l'orientation et les paramètres internes de la caméra sont connus pour la première image de la séquence. Nous rappelons ici succinctement les relations de base liant les points de vues des caméras et les homographies planaires induites par les plans observés.

Etant donnés deux matrices de projection $P_1 = [I|0]$ et $P_2 = [A|a]$ et un plan défini par le vecteur v tel que $v^T X +$

$1 = 0$, alors l'homographie induite par ce plan s'exprime sous la forme [5]:

$$H = K_2(A - av^T)K_1^{-1}$$

où A et a sont les matrices de rotation et le vecteur de translation et

$$K_i = \begin{pmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

est la matrice des paramètres intrinsèques pour l'image i;

2 Calcul du point de vue par observation multiples de plans

2.1 Méthode

Etant donnés les paramètres intrinsèques et un ensemble de points en correspondance entre deux images, une méthode classique pour calculer le point de vue est de minimiser l'erreur moyenne entre les points en correspondance par rapport aux paramètres de la transformation [4]:

$$A,a = argMinJ(\mathbf{A},\mathbf{a}) = \frac{1}{N}\sum_{j=1}^{N} ||x'_j - Z(Hx_j)||^2$$

où Z est la fonction de normalisation qui transforme la troisième composant d'un vecteur 3D en 1. Nos expérimentations nous ont montré que la précision obtenue grâce à l'observation d'un seul plan n'est pas suffisante pour obtenir une bonne impression visuelle de la scène augmentée. En effet, la précision dépend des positions relatives de la caméra par rapport au plan observé et également du nombre de points en correspondance déterminés sur ce plan. De plus, les erreurs s'accumulent au fil du temps puisque la matrice de projection à l'étape i est obtenue par composition de tous les déplacements relatifs la précédant. Pour de longues séquences, l'accumulation des erreurs peut ainsi conduire le processus de calcul à diverger.

Pour cette raison, nous proposons d'utiliser plusieurs plans pour le recalage. Ceci donne en effet plus d'informations sur l'espace tridimensionnel considéré et va conduire à réduire la variabilité des paramètres estimés. Ceci permet également de prendre en compte des environnements plus larges et des mouvements plus complexes. En effet, lors de certains mouvements, le plan observé peut devenir peu visible ou conduire à des points mal détectés. L'utilisation de plusieurs plans permet d'apporter une solution à ce problème, l'un au moins des plans considérés étant généralement bien visible dans l'image.

Dans le cadre multi-planaire,la fonction de coût utilisée est alors:

$$J(\mathbf{A},\mathbf{a}) = \frac{1}{N_1 + \cdots + N_n}\sum_{k=1}^{n}\sum_{j=1}^{N_k} ||x'_{kj} - Z(H_k x_{kj})||^2$$

# Image	Mouvement
0 - 19	Rotation et Translation
20 - 29	Stationnaire
30 - 39	Translation selon l'axe des Y
40 - 49	Translation selon l'axe des X
50 - 64	Panoramique
65 - 69	Stationnaire

TAB. 1 – *Description des mouvements de la caméra au cours de la séquence.*

où n est le nombre de plans, N_k est le nombre de points en correspondance pour le plan k, H_k est l'homographie planaire induite par le plan k et Z est la fonction de normalisation qui transforme la troisième composant d'un vecteur 3D en 1.

Une méthode classique pour optimiser cette fonction non linéaire est la méthode de Newton mais elle est sensible à l'estimée initiale. C'est pourquoi nous utilisons plutôt la méthode de Levenberg-Marquardt qui est plus stable et presque aussi rapide que la méthode de Newton.

2.2 Résultats

Afin de prouver l'efficacité de notre méthode, nous avons d'abord considéré une séquence d'images synthétiques utilisant le modèle de notre mire de calibration. La scène est donc constituée de trois plans d'équations plan 1, $X - 0.577Y = 0$; plan 2, $Y = 0$;et plan 3, $Z = 0$. Cette séquence sera utilisée tout au long de cet article pour valider nos différents apports. Le mouvement de la caméra dans cette séquence est décrit dans la la table 1. De nombreux types de mouvements sont considérés: translationnels, panoramiques et stationnaires. La méthode a été testée en bruitant les points images avec différents écart-types.

Les figures 2 et 3 montrent les composantes translationnelles t_x et t_y calculées avec un bruit additif d'écart type 0.5, ainsi que les valeurs réelles de t_x et t_y. Ces graphiques montrent que l'utilisation de plusieurs plans améliore notablement la précision du point de vue calculé. Plus précisément, ces graphiques montrent l'amélioration de la précision lorsque un, deux ou trois plans sont utilisés pour le calcul du point de vue. Ces expérimentations prouvent que l'utilisation d'un seul plan conduit à des estimations assez instables du point de vue. L'utilisation de deux ou trois plans accroît considérablement la précision et la stabilité de l'estimation.

Afin de quantifier les performances de l'algorithme en terme de temps de calcul, la table 2 fournit le temps passé dans les différentes étapes de l'algorithme pour une image de la mire de calibration de taille 765×576: extraction et mise en correspondance des points d'intérêt (nous utilisons ici l'algorithme MIC [14] pour l'extraction des points), extraction des points appartenant aux structures planaires grâce a un algorithme de type RANSAC et enfin calcul du point de vue. Environ 500 points d'intérêt sont extraits dans chaque image. Parmi ceux ci, seuls une centaine appartiennent aux

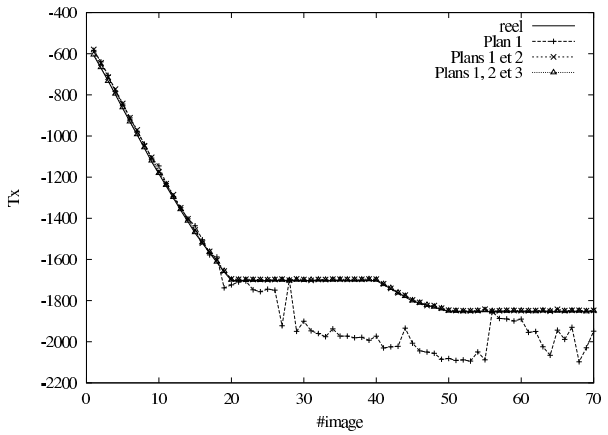


FIG. 2 – *La composante translationnelle t_x calculée au cours de la séquence en utilisant un, deux ou trois plans.*

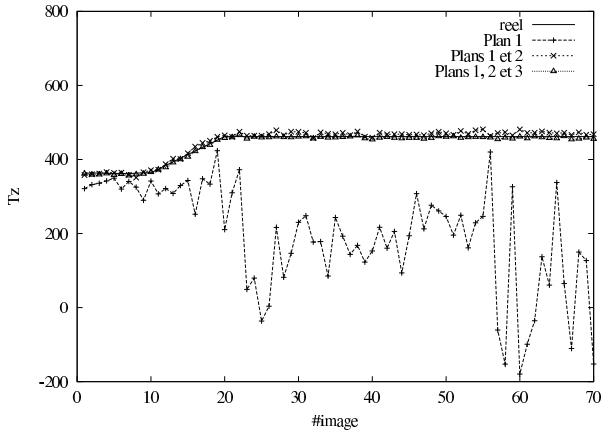


FIG. 3 – *La composante translationnelle t_y calculée au cours de la séquence en utilisant un, deux ou trois plans.*

plans considérés et entrent dans le processus de calcul du point de vue. Le temps total est donc d'environ 66 ms, ce qui conduit à un calcul du point de vue pouvant traiter environ 16 images par seconde.

MIC	30,33 ms
Mise en correspondance	18,62 ms
RANSAC	12,13 ms
Calcul du point de vue	4,8 ms
Total	65,88 ms

TAB. 2 – *Temps de calcul pour le processus complet pour un Pentium III 900 Mhz (moyenne réalisée sur 100 images).*

3 Utilisation de méthodes de stabilisation

3.1 But

Même si la précision du calcul du point de vue est améliorée en considérant plusieurs plans, on observe très souvent des fluctuations importantes sur les paramètres calculés ce qui se traduit pour l'observateur par l'impression de voir l'objet ajouté osciller ou bien glisser dans la scène. Ces phénomènes sont particulièrement visibles quand le mouvement de la caméra est lent. Par le passé, plusieurs papiers ont préconisé l'utilisation du filtre de Kalman pour prédire et stabiliser les trajectoires. Un tel filtre n'est cependant pas toujours avantageux pour la RA, en particulier parce que les modèles dynamiques considérés en général (modèle à vitesse ou accéleration constante) ne sont pas satisfaisants pour décrire les mouvements humains.

Poursuivant les travaux de Matsunaga and Kanatani[4] et [13], nous cherchons ici à utiliser les méthodes de sélection de modèles pour réduire les fluctuations sur les paramètres calculés et améliorer ainsi l'impression visuelle de la scène augmentée. L'idée sous jacente à la sélection de modèles part de la constatation suivante: un modèle d'ordre élevé approche toujours mieux un ensemble de données qu'un modèle d'ordre inférieur. Cependant, les modèles d'ordre élevé approximent en fait une partie du bruit qu'ils sont censés pourtant éliminer. Un modèle d'ordre élevé, bien que théoriquement plus précis, est donc en fait moins stable aux perturbations aléatoires des données. Une bonne méthode de sélection de modèles doit donc réaliser un compromis entre précision et stabilité. Le principe des méthodes de sélection de modèles est donc d'exiger que le modèle choisi explique bien les données et ait en même temps la structure la plus simple possible.

Dans le cas du calcul du point de vue, les homographies (une pour chaque plan observé) sont paramétrées par 9 paramètres (R,t,u_0,v_0,f). Si les mouvements ou les intrinsèques sont contraints (par exemple si on considère certains mouvements comme les panoramiques, les mouvements stationnaires ou les zooms), les homographies auront un nombre de degrés de liberté moindre et donc un nombre de paramètres à estimer plus petit. La stabilité du point de vue sera donc améliorée si le bon modèle de mouvement est utilisé. Il est donc nécessaire de déterminer le mouvement réel de la caméra en utilisant seulement les observations, c'est-à-dire dans notre cas, deux ou trois images consécutives.

3.2 Etat de l'art

De nombreux critères de sélection de modèles pour réaliser un compromis entre le résidu de l'approximation et le degré du modèle ont été proposés dans la littérature [3]. Tous sont la somme d'un terme mesurant la précision du modèle et d'un terme mesurant sa complexité. La plupart d'entre eux sont basés sur des critères statistiques ou sur des critères issus de la théorie de l'information. Parmi eux, les plus utili-

sés sont sans doute le critère géométrique d'Aikaike (AIC) ainsi que le critère de description de longueur minimale (MDL). Le critère AIC peut être vu comme une approximation d'un critère entropique (la distance de Kullbak-Leibler), alors que le critère MDL favorise le modèle dont la description en terme de bits est minimale:

$$G_{AIC} = \hat{J} + 2k\epsilon^2$$
$$G_{MDL} = \hat{J} - k\epsilon^2 log\epsilon^2$$

où k est le nombre de degrés de liberté du mouvement. Le niveau de bruit ϵ^2 est habituellement estimé à partir du résidu $\hat{J}$ (celui correspondant au modèle d'ordre le plus élevé) [4][2]). Quel que soit le critère considéré, la valeur associée à chaque modèle est calculée comme la somme du résidu et du facteur de complexité de chaque modèle $E_{critere} = \hat{J} + \epsilon^2 c(M_k)$. L'usage d'un modèle trop complexe est donc pénalisé.
Cette approche a déja été utilisée par kanatani[4] pour le problème de la calibration à partir d'un seul motif plan observé. Il y classifie les mouvements en 6 catégories, quatre d'entres elles correspondant au cas d'une focale fixe:

Mouvement	**Paramètres connus**	**Variables**
stationnaire	$\mathbf{A}_i = \mathbf{A}_{i-1}$,$\mathbf{a}_i = \mathbf{a}_{i-1}$	$-$
panoramique	$\mathbf{a}_i = \mathbf{a}_{i-1}$	$\mathbf{A}_i$
$t-$ prédit	$t_i = 2t_{i-1} - t_{i-2}$,$\mathbf{a}_i = \mathbf{a}_{i-1}$	$\mathbf{A}_i$
modèle général	$-$	$\mathbf{A}_i$,$\mathbf{a}_i$

Dans le cas du modèle prédit, la position de la caméra est extrapolée linéairement par $t_i = 2t_{i-1} - t_{i-2}$, et l'optimisation n'est effectuée que sur la rotation.
Dans l'approche de Kanatani, seuls deux critères de sélection sont étudiés, G_{AIC} and G_{MDL}, en considérant un seul motif plan pour le recalage. Cependant, il existe beaucoup d'autres critères de sélection, en particulier ceux utilisant la matrice de covariance ou la matrice d'information sur les paramètres calculés.

3.3 Notre approche de la sélection de modèles

Nous suggérons d'utiliser conjointement les méthodes de sélection de modèle et la stratégie multi-planaire pour accroître la précision et la stabilité des paramètres calculés. La sélection de modèles a été utilisée dans de nombreux domaines. Cependant, il n'existe aucun critère qui soit reconnu comme meilleur dans tous les cas, comme le montrent plusieurs papiers évaluant ces critères pour des problèmes différents: approximation par des polynômes [2], fusion de données surfaciques [3], sélection de mouvement [12], détection de primitives géométriques [6].
Pour cette raison, nous avons commencé par comparer différents critères de sélection. Nous avons plus particulièrement considéré les critères qui prennent en considération la matrice de covariance sur les paramètres estimés $(V(\theta_k))$ et la matrice d'information de Fisher $(I(\theta) = E(\frac{\partial}{\partial\theta}J(X|\theta))^t \frac{\partial}{\partial\theta}J(X|\theta)))$. En effet, des critères tels qu'AIC sont seulement des approximations asymptotiques de critères considérant la matrice de covariance ou la matrice d'information. Nous espérons donc que de tels critères amélioreront la sélection de modèles.

Critère	**Terme de complexité**
Akaike AIC [1]	$2k$
Bozdogan CAIC [2]	$k(\log n + 1)$
Bozdogan CAICF [2]	$k(\log n + 2) + \log \lvert\mathbf{I}(\theta_k)\rvert$
Schwarz BIC [8]	$2k \log n$
Kanatani gMDL [4]	$-k \log \epsilon^2$

3.4 Résultats expérimentaux

Les expériences que nous décrivons ont été conduites sur la séquence synthétique précédemment décrite avec différents niveaux de bruit et sur des séquences réelles. Par souci de simplicité dans la présentation des résultats, nous avons affecté un entier à chaque type de mouvement: 0 pour le mouvement stationnaire, 1 pour le panoramique, 2 pour le mouvement général (rotation + translation) et 3 pour le mouvement prédit.

Utilisation du modèle prédictif. Dans de nombreuses expérimentations, les variations des paramètres sont relativement petites et certaines parties peuvent donc apparaître linéaires par morceau à première vue. Le modèle prédictif ayant beaucoup moins de paramètres que le modèle général, le modèle prédictif est alors sélectionné dès que le niveau de bruit s'élève. La sélection de ce modèle trop restrictif peut alors conduire à la divergence du processus. A titre d'exemple, la figure 4 montre le comportement du processus de calcul du point de vue pour la même séquence que précédemment quand on prend en compte ou non le modèle prédictif dans la liste de modèle. Il apparaît clairement sur cette figure que la présence du modèle prédictif tend à faire diverger le processus du point de vue attendu. Ces résultats ont été obtenus en considérant le critère CAICF, les autres critères donnant des résultats similaires. Pour ces raisons, nous n'utilisons pas dans la suite de modèle prédictif. Il est préférable en effet d'avoir quelques fluctuations dans les paramètres plutôt que de faire diverger le processus.

Comparaison des critères de sélection. Afin de comparer les critères, nous avons utilisé la séquence synthétique bruitée. Pour chaque image i, nous avons appliqué l'étape de sélection de modèle. Comme cela requiert la minimisation de la fonction de coût, nous avons utilisé la véritable valeur du point de vue à l'étape $i-1$ comme initialisation afin d'éviter que les problèmes éventuels de dérive ne perturbent l'évaluation des modèles. Le véritable modèle étant connu, nous montrons dans les tables 3 et 4 le pourcentage de modèles correctement choisis pour chacun des critères évalués. Le terme + *complexe signifie qu'un modèle d'ordre plus élevé que le vrai modèle a été choisi. Alors que le terme* - complexe signifie qu'un modèle d'ordre moins élevé, donc trop restrictif, a été choisi.
Ces deux tables montrent que pour une valeur de bruit modérée, la plupart des critères se comportent bien, c'est-à-dire qu'un modèle d'ordre supérieur ou égal au véritable

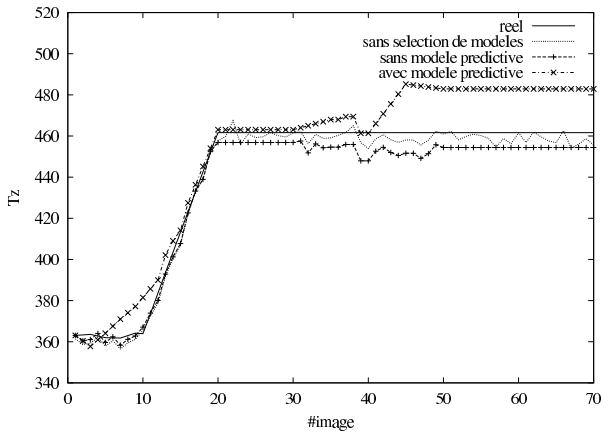


FIG. 4 – *La translation t_z calculée avec et sans le modèle prédictif.*

mouvement	critère	$\sigma = 0.3$		
		- général	**correct**	+ général
static	AIC	-	83.1%	16.9%
	CAIC	-	98.7%	1.3%
	CAICF	-	100.0%	0.0%
	BIC	-	100.0%	0.0%
	gMDL	-	77.5%	22.5%
pan	AIC	0.0%	85.3%	14.7%
	CAIC	0.0%	99.3%	0.7%
	CAICF	0.0%	98.7%	1.3%
	BIC	0.0%	100.0%	0.0%
	gMDL	0.0%	84.7%	15.3%
general	AIC	0.0%	100.0%	-
	CAIC	1.5%	98.5%	-
	CAICF	1.3%	98.7%	-
	BIC	5.4%	94.6%	-
	gMDL	0.0%	100.0%	-

TAB. 3 – *Pourcentage de bonnes sélections de modèles pour le niveau de bruit $\sigma = 0.3$.*

modèle est presque toujours choisi. On note cependant que les critères AIC et $gMDL$ ont tendance à produire des modèles trop généraux, ce qui n'est pas très intéressant dans une optique de stabilisation.

Lorsque le bruit augmente (table 4), les performances de certains critères se dégradent nettement. Cependant on peut noter que les deux critères CAIF ET CAICF se comportent le mieux: ils se comportent très bien dans le cas stationnaire, n'ont que peu tendance à sélectionner un modèle trop général dans le cas panoramique et se comportent très honorablement dans le cas général. Le critère CAICF apparaît comme supérieur au critère CAIC, puisqu'il induit moins de sélections de modèles d'ordre inférieur que CAIC. C'est la raison pour laquelle les expérimentations qui suivent seront faites en utilisant le critère CAICF. Ces résultats tendent donc à montrer que l'introduction de la matrice d'information sur les paramètres calculés améliore la sélection du modèle.

motion	criterion	$\sigma = 1.0$		
		- général	**correct**	+ général
static	AIC	-	83.7%	16.3%
	CAIC	-	100.0%	0.0%
	CAICF	-	100.0%	0.0%
	BIC	-	100.0%	0.0%
	gMDL	-	0.0%	100.0%
pan	AIC	0.0%	86.7%	13.3%
	CAIC	0.0%	100.0%	0.0%
	CAICF	0.0%	97.3%	2.7%
	BIC	0.0%	100.0%	0.0%
	gMDL	0.0%	0.0%	100.0%
general	AIC	11.5%	88.5%	-
	CAIC	24.1%	75.9%	-
	CAICF	20.3%	79.7%	-
	BIC	33.6%	66.4%	-
	gMDL	0.0%	100.0%	-

TAB. 4 – *Pourcentage de bonnes sélections de modèles pour le niveau de bruit $\sigma = 1.$.*

Les figures 5 et 6 comparent les performances des critères de sélection des modèle avec le véritable modèle de mouvement de la séquence. L'axe des x correspond au numéro des images dans la séquence et l'axe des y le modèle de mouvement choisi. On peut constater dans ces graphiques que certaines petites translations sont étiquetées par le processus en modèle panoramique. Ceci est du au fait que ces deux types de mouvement sont assez difficiles à discerner quand les mouvements sont faibles. Mais comme la complexité du modèle panoramique est moindre que celle de la translation (qui rentre dans le modèle général), le mouvement panoramique est détecté. Une solution pour remédier à ce problème serait d'introduire un modèle translationnel. Cependant, ce mouvement apparaît rarement dans les séquences tournées librement. Nous ne l'avons donc pas introduit dans la liste des modèles.

4 Amélioration de la sélection du modèle

4.1 Méthode

Cette méthode améliore très sensiblement la stabilité des points de vue et la qualité visuelle des incrustations. Cependant, il peut exister des erreurs dans l'étiquetage du mouvement, par exemple la confusion entre panoramique et translation dans le cas de petits mouvements. La répétition de choix inappropriés de mouvements pouvant conduire à la divergence du processus, nous souhaitons améliorer le processus de sélction. Pour cela, nous proposons d'utiliser la cohérence temporelle des modèles détectés et d'utiliser donc plus de deux vues pour valider la sélection du modèle. Ceci a pour effet d'augmenter la complexité du processus

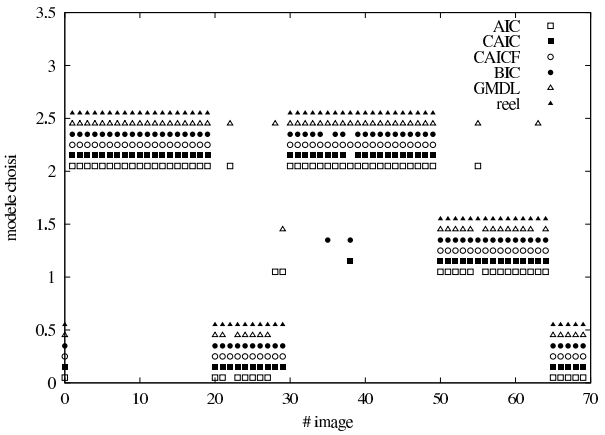


FIG. 5 – *Comparaison des critères de sélection sur une séquence synthétique bruitée* $\sigma = 0.5$.

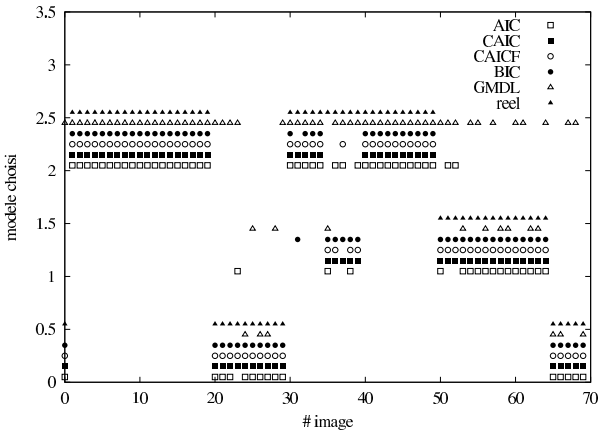


FIG. 6 – *Comparaison des critères de sélection sur une séquence synthétique bruitée* $\sigma = 1.$.

(les points en correspondance doivent être suivis dans plus de deux images) mais cela contribue à éviter qu'un modèle inapproprié soit choisi.
Dans notre approche, un modèle ne sera donc validé que si ce choix est le même pour au moins 3 images consécutives, ce qui présuppose qu'un modèle de mouvement persiste dans au moins trois images consécutives. Dans le cas où des modèles différents seraient trouvés, le modèle le plus général sera choisi, sachant qu'il est préférable de choisir un modèle plus général plutôt que préférer un modèle trop restrictif qui peut conduire à la divergence. Plus précisément, notre algorithme est le suivant:

1. Les paramètres de la caméra sont connus pour les images $i-1$et $i-2$.
2. Sélectionner le modèle de mouvement $M_{i,i-1}$ entre l'image courante i et la précédente $i-1$.
3. Sélectionner le modèle de mouvement $M_{i,i-2}$ entre l'image courante i et l'image $i-2$.
4. Le modèle sélectionné est le modèle M' le plus simple tel que l'espace des paramètres de $M_{i,i-1}$ et $M_{i,i-2}$ soient des sous espaces de M'. Si les espaces sont emboîtés, ceci signifie que M' sera le plus général des deux modèles $M_{i,i-1}$ et $M_{i,i-2}$.

Cette méthode comporte toutefois un léger inconvénient quand on passe d'un modèle complexe à un plus simple, la première image de la transition est alors toujours affectée du modèle le plus complexe. La transition au modèle le plus simple se fait donc avec un temps de retard.

4.2 Ajustement

Une fois le modèle estimé sur la base de trois images, la position de la caméra courante est recalculée de la façon suivante en tenant compte des points en correspondance (x_{i-1},x_i) et (x_{i-2},x_i) entre les trois images:

$$\begin{aligned} J(\mathbf{A},\mathbf{a}) &= \sum_{k=1}^{n}\sum_{j=1}^{N_k} ||x_{kj}^i - Z(H_k^{i,i-1}x_{kj}^{i-1})||^2 \\ &+ \sum_{k=1}^{n}\sum_{j=1}^{N_k} ||x_{kj}^i - Z(H_k^{i,i-2}x_{kj}^{i-2})||^2 \end{aligned}$$

où k désigne le plan observé et H_k^{ij} l'homographie induite par le k^{ieme} plan entre les images i et j..

4.3 Résultats

Nous présentons d'abord des résultats de sélection de modèles utilisant des triplets d'images sur la séquence synthétique additionnée d'un bruit de variance ($\sigma = 1.0$). La figure 7 montre le modèle sélectionné sur la séquence en utilisant deux ou trois images pour la sélection. La figure 8 montre la composante translationnelle T_Z calculée. Nous pouvons constater qu'entre les images 30 et 50, si nous utilisons seulement deux images pour la sélection, il y a un certain nombre de confusions entre modèle panoramique et modèle général. Ces problèmes s'atténuent visiblement quand les triplets d'images sont utilisés pour la sélection. De manière générale, la probabilité de sélectionner un mauvais modèle décroît quand les triplets sont utilisés. Enfin le diagramme montrant T_Z prouve que l'utilisation de triplets améliore la précision du point de vue calculé par rapport à l'utilisation de deux vues.

Séquence réelle: la mire. Nous montrons ici des résultats concernant une séquence réelle de 260 images montrant une mire de calibration. La figure 9 montre le résultat de la sélection de modèle en utilisant deux ou trois images. L'utilisation de triplets permet d'éliminer un certains nombre de sélections erronées. La figure montre la composante T_x calculée entre les images 130 et 170. Le point de vue calculé avec les triplets est plus précis qu'avec le processus à deux images.

Séquence de la pièce. Enfin, nous considérons une séquence de 225 images d'une pièce en modèle réduit. Cette séquence était constituée successivement d'un mouvement

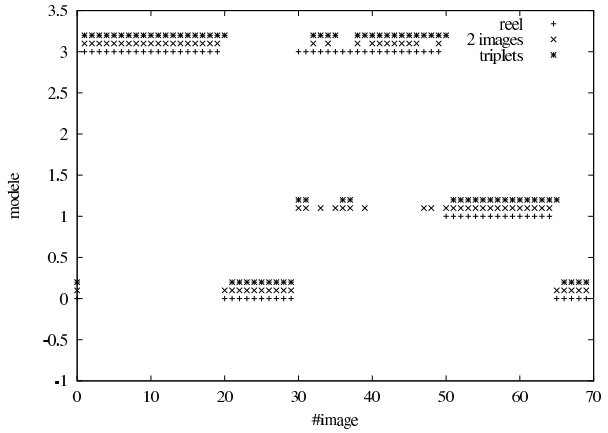


FIG. 7 – *Sélection du modéle en utilisant deux ou trois vues sur la séquence synthétique ($\sigma = 1$).*

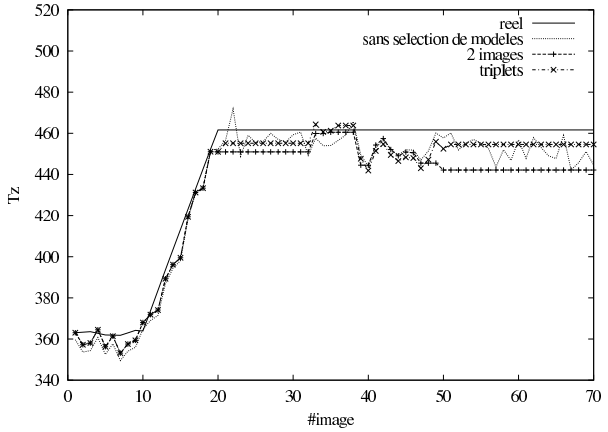


FIG. 8 – *Composante T_Z calculée en utilisant la sélection sur deux ou trois images pour la séquence synthétique.*

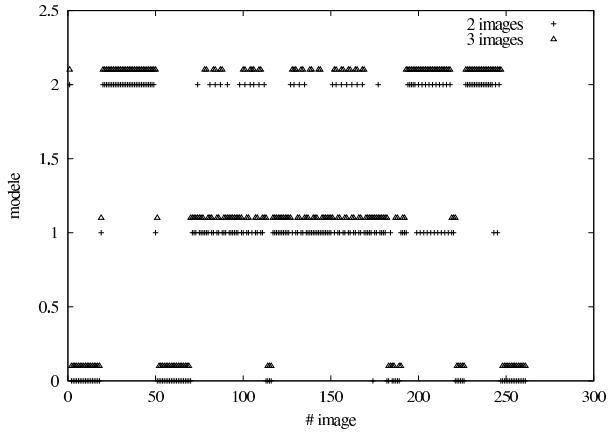


FIG. 9 – *Séquence de la mire: sélection de modèle avec deux ou trois images.*

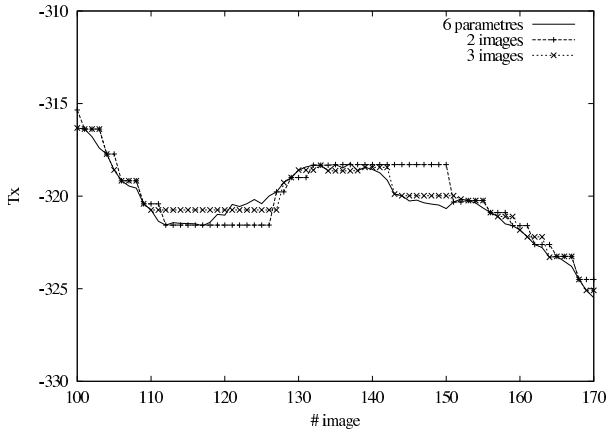


FIG. 10 – *Séquence de la mire: la composante T_X calculée.*

stationnaire, général, stationnaire, panoramique, stationnaire, panoramique et enfin stationnaire.
la figure 11 montre que l'utilisation de triplets améliore légèrement la sélection du modèle. Les images de 30-110 montrent que certains mouvement, identifies à tort comme panoramique ou stationnaire quand on utilise deux vues, sont bien classifiés en modèle général quand on utilise trois vues.
La figure 12 montre que la sélection du point de vue améliore la stabilité de la trajectoire. Ce graphique montre les points de vue calculés lorsque le modèle général est toujours utilisé et lorsque la sélection de modèles est utilisée. Ce graphique montre clairement l'effet stabilisateur de la sélection de modèle puisque la section 0-30 est bien stationnaire. Les images de la figure 13 montrent également l'impact de la sélection sur la robustesse du processus: l'image (a) montre en effet la scène augmentée au bout de 200 images lorsqu'aucune sélection de modèle n'est utilisée (c'est à dire quand le modèle général est toujours utilisé). L'image (b) montre l'image augmentée quand la sélection est utilisée. Il est bien clair sur ces deux images que la sélection apporte beaucoup de robustesse et le lecteur pourra s'en convaincre en regardant sur le site http://www.loria.fr/~vigueras/orasis2003.html les vidéos complètes. De façon générale, la sélection de modèle réduit les variations aléatoires de certains paramètres du point de vue ce qui permet de diminuer les accumulations d'erreur en composant les mouvements relatifs; ceci conduit donc au final à une évaluation plus robuste du point de vue.
Finalement, la figure 14 montre quelques exemples d'augmentation d'une scène avec un cube. La séquence complète est disponible sur notre page web; Un symbole dans le coin supérieur gauche de l'image indique le modèle sélectionné: la croix rouge désigne le modèle stationnaire, le cercle vert le panoramique et le bleu le modèle général.

5 Conclusion

Nous avons proposé dans ce papier plusieurs améliorations aux méthodes existantes de calcul séquentiel du point de

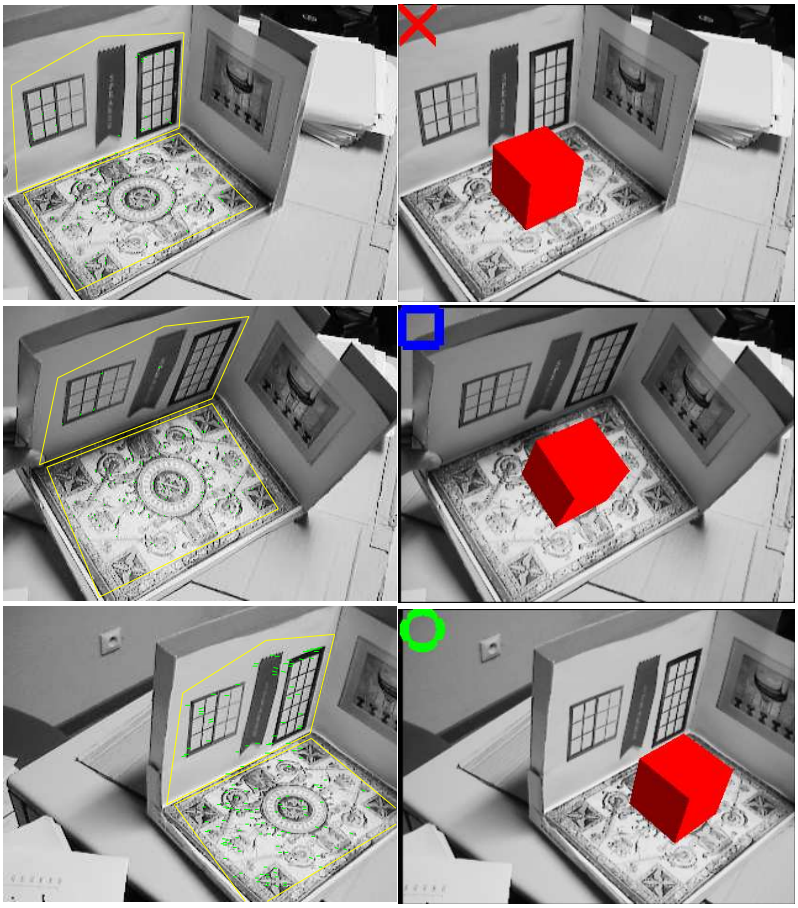


FIG. 14 – *Quelques exemples de la scène augmentée. La première colonne montre le suivi des points dans les trois plans utilisés, la second colonne montre la scène augmentée.*

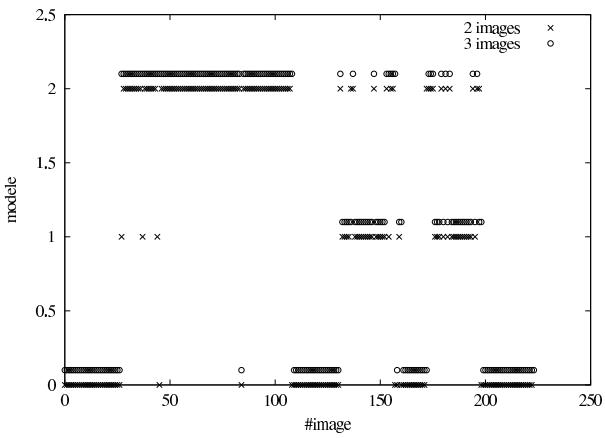


FIG. 11 – *Séquence de la pièce: sélection du modèle.*

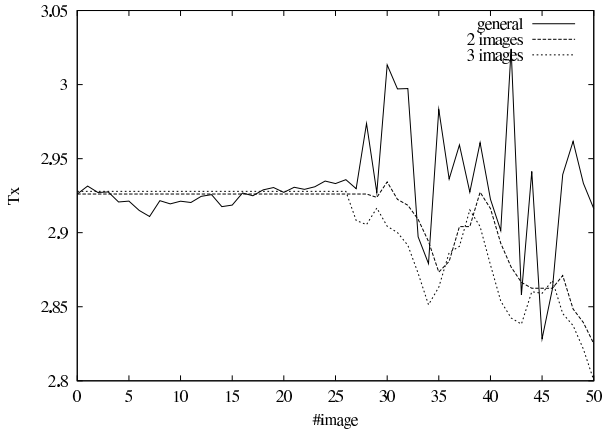


FIG. 12 – *Séquence de la pièce: la translation en x calculée avec la sémection de modèle.*

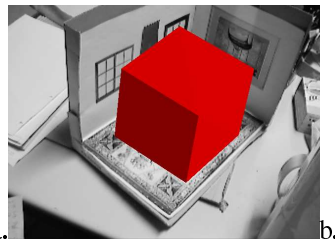
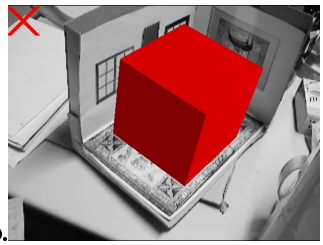
a.  b. 

FIG. 13 – *Un cube incrusté sur l'image 200: (a) sans sélection de modèle; (b): avec sélection*

vue pour des scènes multi-planaires: contexte multiplanaire pour le calcul du point de vue, test de divers critères de sélection de modèle pour améliorer la stabilité de la trajectoire,proposition d'un critère sur plus de deux images pour améliorer la sélection du modèle. Les résultats de cette étude montrent que cette méthode améliore de façon importante la précision et la stabilité de la trajectoire calculée. Le test de différents critères de sélection de modèles a mis en évidence que l'usage de critères impliquant l'information sur la covariance des paramètres calculés améliorait la précision et la robustesse des trajectoires calculées. Nous cherchons maintenant à étendre cette étude au cas de caméras à focale variable. Ce cas semble plus délicat à prendre compte que le cas à focale fixe car les différents modèles considérés ne sont pas emboîtés.

Références

[1] H. Akaike. A new look at the statistical model identification. *IEEE Trans Aut Ctrl*, 19(6):716–723, 1974.

[2] H. Bozdogan. Model Selection and Akaike's Information Criterion (AIC); The General Theory and its Analytical Extensions . *Psychometrika*, 52(3):345–370, 1987.

[3] K. Bubna et C.V. Stewart. Model selection and surface merging in reconstruction algorithms. In *Proceedings of 6th International Conference on Computer Vision, Bombay (India)*, pages 895–902, 1998.

[4] K. Kanatani C. Matsunaga. Calibration of a Moving Camera Using a Planar Pattern: Optimal Computation, Reliability Evaluation and Stabilization by Model Selection. In *Proceedings of 6th European Conference on Computer Vision, Trinity College Dublin (Ireland)*, pages 595–609, 2000.

[5] R. I. Hartley et A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049, 2000.

[6] K. Kanatani. Model Selection for Geometric Inference. In *Proceedings of 5th Asian Conference on Computer Vision, Melbourne, Australia*, pages 23–25, 2002.

[7] U. Neumann et Y. Cho. A selftracking augmented reality system. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, pages 109–115, 1996.

[8] G. Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464, 1978.

[9] G. Simon et M.-O. Berger. A Two-stage Robust Statistical Method for Temporal Registration from Features of Various Type. In *Proceedings of 6th International Conference on Computer Vision, Bombay (India)*, pages 261–266, January 1998.

[10] G. Simon et M.-O. Berger. Registration with a Zoom Lens Camera for Augmented Reality Applications. In *Proceedings of 6th European Conference on Computer Vision, Trinity College Dublin (Ireland)*, June 2000.

[11] A. State, G. Hirota, D. Chen, W. garett, et M. Livingston. Superior Augmented Reality Registration by Integrating Landmark Tracking and Magnetic Tracking. In *Computer Graphics (Proceedings Siggraph New Orleans)*, pages 429–438, 1996.

[12] P.H.S. Torr. An Assessment of Information Criteria for Motion Model Selection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, PR (USA)*, pages 47–52, June 1997.

[13] P.H.S. Torr, A.W. Fitzgibbon, et A. Zisserman. Maintaining multiple motion model hypotheses over many views to recover matching and structure. In *Proceedings of 6th International Conference on Computer Vision, Bombay (India)*, pages 485–491, 1998.

[14] M. Trajkovic et Mark Hedley. Fast corner detection. *Image and Vision Computing*, (16):75–87, 1998.

[15] J. Vallino. *Interactive Augmented Reality*. Thèse de doctorat, University of Rochester, December 1998.