Un panorama des techniques de suivi visuel temps réel pour la réalité augmentée

Eric Marchand * — **Muriel Pressigout**[**]

* *INRIA Rennes-Bretagne Atlantique, IRISA, F-35000 Rennes, France*
marchand@irisa.fr

** *INSA Rennes, IETR, UMR 6164, F-35000, Rennes*
muriel.pressigout@insa-rennes.fr

RÉSUMÉ. *La réalité augmentée désigne les différentes méthodes qui permettent d'incruster de façon réaliste des objets virtuels dans une séquence d'images. Pour assurer cette fusion il convient de pouvoir déterminer la position de la caméra par rapport à l'environnement ce qui revient à aligner un repère lié à la scène avec le repère de la caméra. Cet article se propose de fournir un rapide aperçu des techniques de suivi visuel spatio-temporelle qui sont utilisées pour la réalité augmentée.*

ABSTRACT. *Augmented Reality has now progressed to the point where real-time applications are being considered and needed. At the same time it is important that synthetic elements are rendered and aligned in the scene in an accurate and visually acceptable way. In order to address these issues, a real-time, robust and efficient tracking algorithm have to be considered. The tracking of objects in the scene amounts to calculating the location (or pose) between the camera and the scene.*

MOTS-CLÉS : *Réalité augmentée, suivi 2D, suivi 3D, vision par ordinateur*

KEYWORDS: *Augmented reality, 2D tracking, 3D tracking, computer vision*

RSTI - TSI – 28/2009. Réalités virtuelle et augmentée, pages 921 à 943

1. Introduction

Le concept de réalité augmentée vise à accroître notre perception du monde réel, en y ajoutant des éléments fictifs, non visibles *a priori*. La réalité augmentée désigne donc les différentes méthodes qui permettent d'incruster de façon réaliste des objets virtuels dans une séquence d'images. Ses applications sont multiples et touchent de plus en plus de domaines (voir figure 1) : jeux vidéo et « edutainment », cinéma et télévision (postproduction, studios virtuels, retransmissions sportives...), industrie (conception, design, maintenance, assemblage, pilotage, robotique et télérobotique, implantation, étude d'impact...), médical, etc.

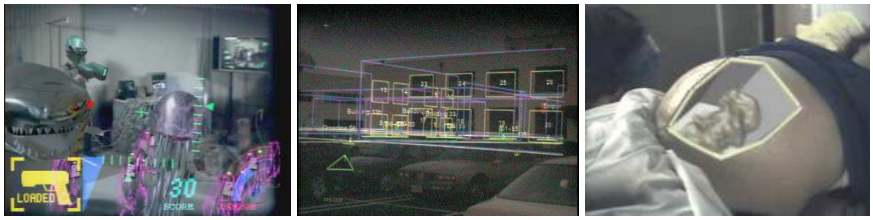
Agrémenter d'objets fictifs une séquence vidéo issue d'un plan fixe ne pose guère de problèmes. Dès lors que la caméra est en mouvement, les difficultés techniques et méthodologiques deviennent significatives. Les applications visées demandant souvent énormément de réalisme, il est indispensable que l'ajout d'objets dans une scène ne perturbe pas la cohérence du contenu. Le fait de déplacer la caméra implique cependant un mouvement de la scène dans l'image. Pour assurer la cohérence entre les deux flux réels et virtuels, un lien rigide doit être maintenu entre les deux mondes. Afin de donner l'illusion que ces objets fictifs appartiennent au même environnement, il est nécessaire de bien les placer, bien les orienter et de respecter des facteurs d'échelle. Bien placer les objets virtuels par rapport aux objets de la scène nécessite donc de connaître la position de la caméra par rapport à la scène réelle.

Le problème de la localisation de la caméra est donc crucial et peut être résolu par diverses approches. On peut utiliser un système de capteurs, comme des capteurs magnétiques qui mesurent la distorsion du champ magnétique pour calculer leur position, des capteurs optiques, des encodeurs sur les moteurs du pied des caméras ou encore, évidemment, le flux vidéo. Cependant, il s'agit dans cet article de se limiter à une approche image, ce qui ramène le problème de réalité augmentée à un problème de vision par ordinateur. Dans certains contextes applicatifs comme le cinéma, l'ensemble de la séquence vidéo est disponible avant le traitement. Dans cette optique de postproduction, des traitements lourds en termes de temps de calcul sont envisageables. Des techniques permettant à la fois la reconstruction 3D d'un certain nombre de points de la scène et la localisation 3D de la caméra sont mises en œuvre par des techniques d'autocalibration ou d'ajustement de faisceaux (*bundle adjustment*). Des logiciels commerciaux reposant sur ce principe sont d'ores et déjà disponibles (on peut citer Boujou de la société 2d3 – issu de l'université d'Oxford – et MatchMover de la société Realviz – issu du projet Robotvis de l'Inria Sophia-Antipolis –). Ces méthodes sont cependant très dépendantes de la qualité de la mise en correspondance des primitives 2D (bruit d'extraction, distribution spatiale, nombre d'erreurs d'appariement...) et l'utilisateur est parfois mis à contribution.

Dans le cadre d'applications interactives (audiovisuel dans les « conditions du direct », industrie, jeux vidéo interactifs, médical, militaire) le recours à des techniques d'autocalibration n'est plus possible. Des techniques permettant la localisation de la caméra uniquement à partir de l'image courante (et éventuellement des précédentes)

Application dans le domaine de l'audiovisuel : effets spéciaux pour le cinéma (Terminator 3) ou la publicité (Renault/Realviz), effets spéciaux interactifs pour la télévision (France 2, Total-immersion), aide à l'interprétation pour le sport, ici pour le football (TF1, Symah-vision)

Industrie : étude de conformité (Siemens) industrie automobile, analyse des déformations (Arvika), aide à l'assemblage (Arvika),



Divers : jeux vidéo (Aquagauntlet, MR-lab), applications militaires (US Marines Corps), applications médicales (UNC Chapel Hill)

Figure 1. *Exemples d'application de la réalité augmentée*

sont nécessaires (Azuma, 1997; Azuma *et al.*, 2001; Navab, 2003). Si un modèle de la scène (ou d'une partie de celle-ci) est disponible, le calcul de point de vue est évidemment une solution idéale à ce problème. Dans le cas où la structure 3D de la scène n'est que partiellement connue d'autres approches reposant, par exemple, sur le calcul du déplacement de la caméra sont envisageables (Simon *et al.*, 2002a). Les avantages de ces approches interactives sont multiples :

– elles permettent une intégration réel-virtuel en temps réel (c'est-à-dire, à la cadence vidéo) car les calculs sous-jacents sont relativement peu coûteux ;

– il n'est pas non plus nécessaire de faire un étalonnage « lourd » du système comme c'est le cas si on utilise d'autres types de capteurs (par exemple, des capteurs magnétiques, centrale inertielle ou autre), ni de disposer *a priori* de la séquence complète ;

– elles peuvent fonctionner sur des plates-formes PC standard ce qui implique un coût relativement faible.

Cependant contrairement aux techniques utilisées pour la postproduction, et si un grand nombre de prototypes ont été réalisés, et à l'exception notable de ARToolKit[1], il existe, à notre connaissance, peu de systèmes « sur étagère » distribués commercialement (on notera cependant l'existence récente des sociétés ARmusement ou Total-Immersion qui commencent à diffuser des bibliothèques logicielles pour la réalité augmentée).

Pour aboutir à un suivi et une localisation robuste et stable, la plupart des systèmes ont recourt à l'utilisation de marqueurs. Cette technique implique cependant de modifier l'environnement ce qui n'est pas toujours possible à réaliser. La suppression de la contrainte imposée par les marqueurs est un passage nécessaire pour une industrialisation effective de la réalité augmentée interactive. Cependant, si un nombre important d'algorithmes de suivi en milieu naturel existent dans la littérature ceux-ci ne sont pas toujours, pour des raisons diverses (fiabilité, temps de calcul, type de résultats...), compatibles avec une application de réalité augmentée. Dans cet article nous nous focalisons donc sur les techniques de calcul de points de vue compatibles avec ce type d'applications en considérant d'une part les approches reposant sur l'utilisation d'un modèle 3D (calcul de pose) et les approches ne nécessitant qu'un modèle simplifié (calcul du déplacement de la caméra). Dans tous les cas les méthodes présentées dans cet article permettent une implémentation temps réel (c'est-à-dire généralement à la cadence vidéo) et permettent donc le développement d'applications interactives.

Cet article ne prétend pas être une synthèse complète des différentes approches présentées dans la littérature[2] et la typologie retenue est sans doute critiquable mais il illustre les grandes classes de méthodes qui sont couramment retenues.

1. ARToolKit a été initialement développé par Hirokazu Kato de l'université d'Osaka, et est actuellement supporté par le HIT Lab. (*Human Interface Technology Laboratory*) de l'université de Washington, et par le HIT Lab NZ de l'université de Canterbury en Nouvelle Zélande.
2. Pour les lecteurs souhaitant avoir plus de détails techniques, une monographie de synthèse beaucoup plus complète a récemment été publiée (Lepetit *et al.*, 2005).

2. Calcul de points de vue

Dans un premier temps, nous considérons qu'un modèle 3D de la scène est disponible. L'idée sera donc de calculer la position et l'orientation relatives de la scène et de la caméra à partir de la projection de ce modèle sur le plan image. L'utilisation d'un modèle 3D est une contrainte forte qui permet d'obtenir une très bonne précision et de traiter des occultations temporaires de l'objet dans une des vues. Ce type d'approche reposant sur des modèles 3D se prête très bien à une intégration dans des expérimentations de réalité augmentée car ils fournissent directement la position relative entre la scène et la caméra. Nous verrons cependant dans la section 4 que des alternatives au calcul « direct » de points de vue existent.

2.1. *Principe du calcul de pose mono-image avec un modèle complet de la scène*

Cette section présente la problématique du calcul de point de vue sur un exemple simple. Afin d'illustrer le problème, nous avons considéré que le modèle de la scène est uniquement composé de points. Le problème du calcul de pose se généralise cependant à tout type de primitives (points, droites, contours, etc.).

D'un point de vue formel, nous considérons une caméra ayant un modèle de projection perspective. Un point ${}^{o}\mathbf{P} = ({}^{o}X, {}^{o}Y, {}^{o}Z, 1)^{\top}$ dont les coordonnées sont exprimées dans le repère de la scène $\mathcal{R}_o$ se projette donc sur le plan image en un $\mathbf{p} = (x, y, 1)^{\top}$ avec :

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \propto \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\mathbf{K}} \underbrace{\begin{bmatrix} {}^{c}\mathbf{R}_o & {}^{c}\mathbf{t}_o \\ 0 & 1 \end{bmatrix}}_{{}^{c}\mathbf{M}_o} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

où ${}^{c}\mathbf{t}_o$ et ${}^{c}\mathbf{R}_o$ sont le vecteur de translation et la matrice de rotation définissant la position de la caméra par rapport à la scène (${}^{c}\mathbf{R}_o$ étant une matrice de rotation elle doit évidemment respecter des propriétés d'orthogonalité) et $\mathbf{K}$ est la matrice de projection perspective[3].

L'objectif du calcul de pose, dans le cas de primitive de type point, est donc d'estimer la pose relative ${}^{c}\mathbf{M}_o$ connaissant le modèle 3D, ${}^{o}\mathbf{P}$, et les mesures dans l'image

3. Nous considérons que toutes les mesures extraites des images sont exprimées en mètre ce qui implique que les paramètres intrinsèques de la caméra (point principal, focale, taille des pixels...) sont connus *a priori*. Le processus d'étalonnage, relativement similaire à celui du calcul de pose, n'est pas décrit dans cet article. Une bonne description de ce processus est donnée dans (Lavest *et al.*, 2003).

$\mathbf{p}$. C'est un problème intrinsèquement non linéaire qui revient à minimiser l'erreur de recalage entre les mesures image $\mathbf{p}_i$ et la projection du modèle $\mathbf{K}\,{}^c\mathbf{M}_o\,{}^{\mathbf{o}}\mathbf{P}_i$:

$$\widehat{{}^c\mathbf{M}_o} = \underset{{}^c\mathbf{M}_o}{\operatorname{argmin}}\ \Delta \ \text{ avec } \ \Delta = \sum_{i=1}^{N} \left(\mathbf{p}_i - \mathbf{K}\,{}^c\mathbf{M}_o\,{}^{\mathbf{o}}\mathbf{P}_i\right)^2 \qquad [1]$$

où N est le nombre de points considéré dans le modèle.

Nous avons pris pour exemple le cas simple du calcul de pose à partir de points. C'est en effet un cas qui a largement été considéré dans la littérature (Brown, 1971; Fischler *et al.*, 1981; Horaud *et al.*, 1989; Haralick *et al.*, 1989; Dementhon *et al.*, 1995; Lu *et al.*, 2000). Notons cependant que ces techniques sont généralisables au cas des segments (Dhome *et al.*, 1989; Lowe, 1987), des coniques (Safaee-Rad *et al.*, 1992), ou des objets cylindriques (Dhome *et al.*, 1990). Peu d'approches considèrent l'utilisation conjointe de plusieurs types de primitives (voir cependant (Phong *et al.*, 1995) pour l'utilisation de points et de droites ou (Marchand *et al.*, 2002)).

Concernant les objets complexes (voir section 4.1) l'approche la plus classique est de considérer des distances aux contours (Lowe, 1991; Drummond *et al.*, 2002; Martin *et al.*, 2002; Simon *et al.*, 1998; Comport *et al.*, 2006). (Drummond *et al.*, 2002; Comport *et al.*, 2006) minimisent la distance entre les points de contour extraits de l'image courante et le projeté du modèle. Le problème d'*aperture* est ainsi mieux pris en compte que par une distance point-point. Récemment, (Vacchetti *et al.*, 2004a) puis (Wuest *et al.*, 2005) ont proposé une approche de calcul de pose basée sur un suivi des contours gardant plusieurs hypothèses de sites de contours afin de mieux gérer les erreurs d'appariements dues à la texture environnante.

S'il est nécessaire de connaître la pose ${}^c\mathbf{M}_o$ pour pouvoir « augmenter » la séquence, nous verrons dans la section 4.2.1 que, sous certaines hypothèses concernant la géométrie de la scène, celle-ci peut être calculée en intégrant le déplacement 3D de la caméra au cours du temps.

2.2. *Principes de résolution*

Le problème de calcul de pose ou de localisation 3D est très ancien en vision par ordinateur (citons les travaux historiques sur le P3P (Fischler *et al.*, 1981) – *Perspective from 3 points* –) ou en photogrammétrie (Brown, 1971). La résolution du problème de recalage sous-jacent (défini par l'équation [1]) a suscité et continue de susciter de nombreuses études.

2.2.1. *Techniques analytiques ou linéaires*

Dans le cas où un faible nombre de primitives est disponible, il existe des solutions purement analytiques à ce problème consistant à résoudre directement le système d'équations non linéaires issu de l'équation [1] (Fischler *et al.*, 1981; Horaud *et al.*, 1989; Dhome *et al.*, 1989). Par nature, ces problèmes sont non linéaires par

rapport aux paramètres de pose mais il existe des solutions reposant sur la résolution de systèmes linéaires aux moindres carrés (par exemple, (Faugeras *et al.*, 1987; Liu *et al.*, 1990)) pour estimer la pose. Dans ce cas, l'efficacité de ces approches repose principalement sur la représentation choisie pour la matrice de rotation (matrice 3×3 (Ganapathy, 1984; Faugeras *et al.*, 1987), angles d'Euler (Liu *et al.*, 1990)) et les contraintes retenues pour assurer l'orthonormalité de cette matrice. Elles sont cependant extrêmement sensibles aux bruits de mesure.

2.2.2. *Techniques d'optimisation itératives ou récursives*

Le problème étant non linéaire l'utilisation de techniques d'optimisation de type Newton doit être considérée pour obtenir des résultat optimaux ((Lowe, 1987; Lowe, 1991; Drummond *et al.*, 2002; Martin *et al.*, 2002; Comport *et al.*, 2006)). Elles consistent à minimiser l'erreur entre les observations dans l'image et la projection du modèle de l'objet pour une pose donnée. La minimisation est généralement réalisée en utilisant des algorithmes numériques itératifs de type Newton-Raphson ou Levenberg-Marquardt. L'asservissement visuel virtuel (Marchand *et al.*, 2002; Comport *et al.*, 2006) est une variante de ces méthodes qui consiste à déplacer une caméra virtuelle, par asservissement visuel, afin de minimiser l'erreur entre les observations et la projection du modèle de la scène sur le plan image de cette caméra virtuelle. Le principal avantage de ces approches est la précision du résultat obtenu. En contrepartie, l'algorithme de minimisation est sensible aux minima locaux, et peut, dans certains cas critiques, diverger. C'est pourquoi, une bonne initialisation du vecteur de paramètres à estimer est souvent nécessaire.

Une solution très élégante est proposée dans (Dementhon *et al.*, 1995; Oberkampf *et al.*, 1996). Cette approche repose sur un algorithme itératif faisant initialement l'hypothèse d'un modèle de projection perspective à l'échelle (caméra para-perspective) et permet de se ramener progressivement vers un modèle de projection perspective pure. Cette méthode est extrêmement rapide en temps de calcul grâce à des précalculs judicieux et permet un calcul de la pose beaucoup plus précis que les approches analytiques ou linéaires.

L'estimation de la pose peut aussi se faire en utilisant des approches de filtrage récursives de type Kalman (Lowe, 1987; Gennery, 1992; Koller *et al.*, 1993). Le caractère non linéaire du problème rend nécessaire l'utilisation de filtres de Kalman étendus (EKF) ou même de techniques de filtrage particulaire.

2.3. *Du calcul de pose à la réalité augmentée*

Une fois le point de vue calculé pour chaque image, le processus d'augmentation est relativement simple. Étant donné un modèle de la scène virtuelle (dont les coordonnées sont exprimées dans le même repère $\mathcal{R}_o$ que le modèle de la scène réelle), il conviendra simplement de définir la caméra virtuelle dans le système de rendu (OpenGL, DirectX) avec les paramètres de la caméra réelle (c'est-à-dire avec la

pose ${}^{c}\mathbf{M}_{o}$ estimée et les paramètres intrinsèques préalablement calculés). En pratique, quelques transformations sont cependant nécessaires puisque les conventions utilisées pour la synthèse d'image sont légèrement différentes de celles utilisées en vision par ordinateur. Une explication complète est donnée dans (Li, 2001) en ce qui concerne OpenGL.

Nous venons de donner le principe général du processus calcul de point de vue. D'un point de vue pratique, le cas de scènes marquées ou de scènes naturelles va beaucoup influencer le choix des méthodes à utiliser.

3. Réalité augmentée à partir de marqueurs

Pour répondre à un soucis de fiabilité et de robustesse, la grande majorité des systèmes de réalité augmentée temps réel reposent actuellement sur l'utilisation de marqueurs. Le problème de la localisation 3D basée marqueurs peut sembler trivial mais il soulève cependant de nombreuses questions : la conception (Fiala, 2005; Kato *et al.*, 1999) et le choix de marqueurs (Davis *et al.*, 2004; Zhang *et al.*, 2002), leur extraction (Kato *et al.*, 1999) et identification, le marquage des scènes, etc.

Le système de ce type le plus utilisé actuellement est sans contexte le système ARToolKit (figure 2c). Le point de vue est calculé par des techniques simples de calcul de pose. Le problème est cependant simplifié puisque la scène considérée est supposée plane (Kato *et al.*, 1999). Le succès de cette bibliothèque est principalement dû à la simplicité de la mise en œuvre et au processus simple mais fiable et rapide de détection des marqueurs dans l'image. Des logiciels similaires ont été développés dans le cadre du projet Arvika[4].

D'autres approches reposant cette fois sur des techniques d'optimisation non linéaire existent aussi comme par exemple le logiciel Marker (*Marker-based* A*ugmented* R*eality* KER*nel*) reposant sur les méthodes d'asservissement visuel virtuel (Marchand *et al.*, 2002). Le point de vue peut être calculé à partir de primitives de type points ou cercles (figure 2a). Par ailleurs il existe des approches qui après une phase d'apprentissage reposant sur l'utilisation de marqueurs permettent de faire un suivi sans marqueur à partir de points d'intérêt (Friedrich, 2002; Genc *et al.*, 2002).

Si de telles approches sont très robustes et permettent le prototypage et la validation rapides de systèmes de réalité augmentée, la présence intrusive de marqueurs dans l'environnement nous semble être un frein majeur au déploiement de cette technologie en environnement opérationnel. Seul un nombre limité d'applications comme

4. Arvika est un vaste projet sponsorisé par le BMBF (ministère de la Recherche et de l'Éducation Allemand) visant à promouvoir la réalité augmentée en environnement industriel et à proposer et mettre en œuvre de tels systèmes (`http ://www.arvika.de`). Ce projet est sans doute la raison principale de la très forte implication des milieux industriels et académiques allemands dans le domaine de la réalité augmentée.

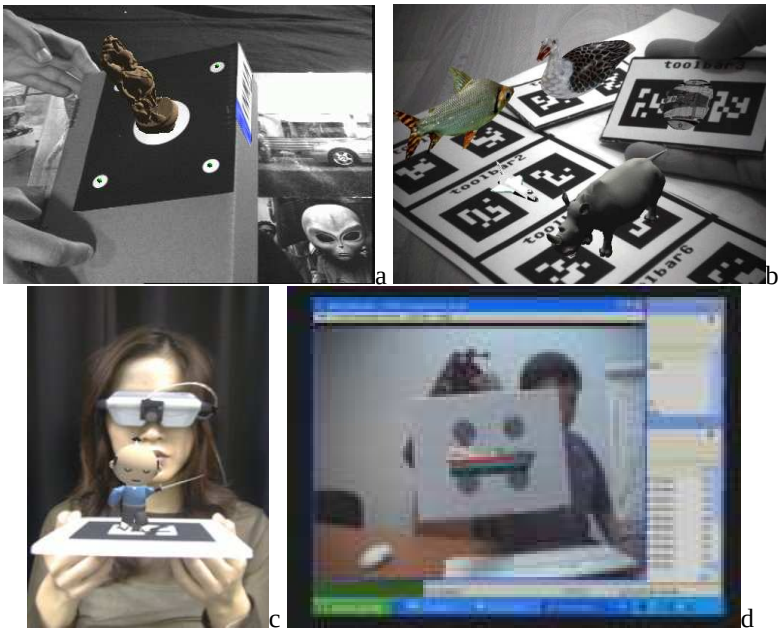


Figure 2. *Localisation 3D pour la réalité augmentée reposant sur un suivi de marqueurs (a) Marker (Marchand* et al.*, 2002) (b) ARtag (Fiala, 2005) (c) AR Toolkit (Kato* et al.*, 1999) (d) vidéo-conférence augmentée interactive* via *un réseau IP (Total-immersion)*

(sans être exhaustif) le design industriel, le jeu vidéo, la vidéo-conférence peuvent, à notre avis, tirer profit de ces méthodes.

4. Réalité augmentée à partir de primitives naturelles

L'utilisation de marqueurs en environnement naturel est cependant, en pratique, peu réaliste. Les méthodes de calcul de point de vue ont donc été adaptées au cas d'environnements quelconques. Dans le cas où un modèle complet de la scène est disponible, la méthodologie présentée dans la section 2.1 reste globalement valide (voir section 4.1). Mais les modèles 3D pouvant être complexes à obtenir, des solutions reposant sur l'utilisation de modèles simples ont vu le jour (voir section 4.2).

4.1. *Calcul de pose en environnement sans marqueur*

Concernant l'utilisation de techniques basées modèles sans marqueur pour la réalité augmentée l'approche la plus classique est une généralisation de la méthode évo-

quée dans le section 2.1 et repose sur une minimisation robuste de la distance entre le contour et la projection du modèle 3D dans l'image. La figure 3 montre quelques séquences augmentées réalisées en utilisant l'algorithme décrit dans (Comport *et al.*, 2006) (d'autres approches décrites dans (Simon *et al.*, 1998) (voir figure 5), (Drummond *et al.*, 2002) (voir figure 4a) ou même (Vacchetti *et al.*, 2004b) (voir figure 4b) reposent sur le même type de méthodes). L'un des avantages de ces méthodes est leur robustesse face aux occultations partielles, aux variations d'éclairage, aux mouvements relativement importants de la caméra, etc. Ceci est dû d'une part à des algorithmes efficaces pour gérer les appariements locaux et d'autre part à l'utilisation d'estimateurs robustes dans le processus de minimisation (Simon *et al.*, 1998; Drummond *et al.*, 2002; Comport *et al.*, 2006).

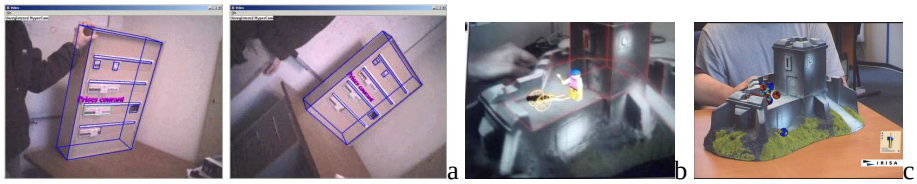


Figure 3. *Suivi par recalage 2D-3D, travaux réalisés à l'Inria Rennes, par (Comport* et al.*, 2006) : (a) suivi d'une armoire électrique, occultations et mouvements importants (b) suivi d'un château fort (Comport* et al.*, 2006), le mouvement (c) des billes incrustées est géré par un moteur physique d'animation intégré dans le logiciel D'Fusion de Total-Immersion*

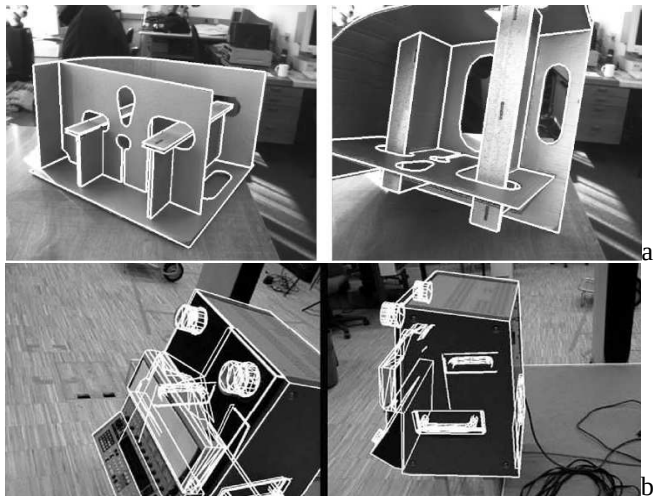


Figure 4. *Suivi par recalage 2D-3D : (a) travaux réalisés à Cambridge par (Drummond* et al.*, 2002) (b) travaux réalisés à l'EPFL par (Vacchetti* et al.*, 2004a)*

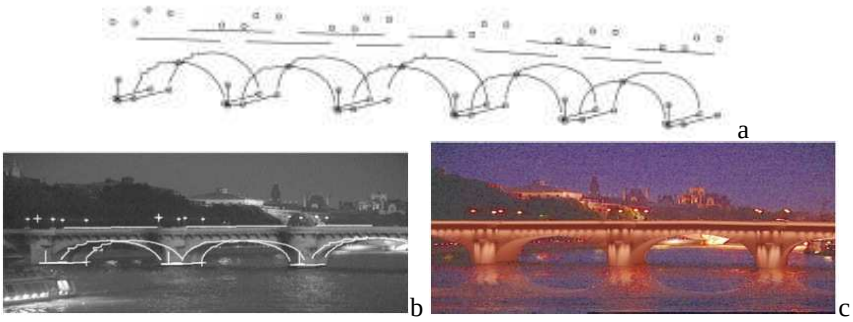


Figure 5. *Recalage 2D-3D en utilisant des modèles complexes (travaux réalisés au Loria par (Berger* et al.*, 1999)). (a) modèle 3D utilisé (b) modèle recalé sur l'image (c) image augmentée*

4.2. *Calcul de points de vue avec modèle simple de la scène*

L'utilisation de modèles 3D complexes n'est cependant pas toujours satisfaisante dans le sens où les scènes ne peuvent être augmentées que si elles contiennent des objets dont la géométrie est connue. Ceci est très restrictif et une autre approche peut être considérée. L'idée, bien répandue, suppose que la séquence vidéo est une suite d'images d'une même scène. L'exploitation de ce contenu commun entre images se succédant va permettre de calculer le déplacement de la caméra en utilisant les contraintes de la géométrie multivues. Autrement dit, ces méthodes se basent sur l'estimation du mouvement (ou du déplacement 3D) de la caméra entre deux images successives et non plus sur une seule image. Dans ce contexte l'ensemble de l'information lumineuse peut être considéré et non plus uniquement l'information située sur les forts gradients spatiaux (contour) de l'image comme c'est le cas pour les méthodes évoquées dans la section précédente. La pose de la caméra tout au long d'une séquence vidéo peut être obtenue par intégration du mouvement à partir du déplacement de la caméra et de sa pose initiale (Simon *et al.*, 2000; Simon *et al.*, 2002a).

4.2.1. *Principe général*

Le calcul du déplacement de la caméra entre deux positions peut se faire de façon assez similaire au calcul de pose. Considérons le cas d'une scène composée d'un certain nombre de primitives visuelles 2D (comme dans la section 2.1, nous considérons sans perte de généralité des primitives de type point). Pour estimer le déplacement de la caméra, une approche classique (Hartley *et al.*, 2001) est de minimiser la distance dans l'image entre la position de ces primitives mesurée dans l'image 2 ($^2\mathbf{p}$) et leur position dans l'image 1 ($^1\mathbf{p}$) transférée dans l'image 2 par une transformation particulière $^2tr_1(^1\mathbf{p})$.

Dans le cas général (scène non coplanaire et mouvements de caméra quelconques) le transfert de points peut s'effectuer en utilisant plusieurs images et en considérant la géométrie épipolaire et la matrice essentielle ou fondamentale (Hartley *et al.*, 2001). La plupart du temps les méthodes proposées se restreignent au cas moins général (scène plane ou à l'infinie, mouvement de rotation pure) ou le transfert de point peut être réalisé en utilisant une homographie. Dans ce cas, un point ${}^1\mathbf{p}$, exprimé en coordonnées homogènes ${}^1\mathbf{p} = ({}^1x, {}^1y, {}^1w)$, est transféré en ${}^2\mathbf{p}$ dans l'image 2 en utilisant la relation suivante :

$$ {}^2\mathbf{p} \propto {}^2tr_1({}^1\mathbf{p}) = {}^2\mathbf{H}_1 {}^1\mathbf{p} \quad [2] $$

où ${}^2\mathbf{H}_1$ est une homographie définie à un facteur d'échelle près. Quand un déplacement ${}^2\mathbf{M}_1$ de la caméra est généré, l'homographie ${}^2\mathbf{H}_1$ est donnée par (Hartley *et al.*, 2001) :

$$ {}^2\mathbf{H}_1 = ({}^2\mathbf{R}_1 + \frac{{}^2\mathbf{t}_1}{{}^1d} {}^1\mathbf{n}^T) \quad [3] $$

où ${}^1\mathbf{n}$ et ${}^1d$ sont la normale et la distance au plan de référence exprimées dans le repère de la caméra 1.

Si la transformation (homographie) présentée ici repose sur l'hypothèse d'une structure plane, certains travaux prennent néanmoins en compte un terme de parallaxe pour traiter des scènes plus complexes à partir des points mis en correspondance (Faugeras *et al.*, 1988; Boufama *et al.*, 1995).

4.2.2. *Méthodes de calcul du déplacement de la caméra*

Comme nous l'avons vu dans la section précédente, même si schématiquement l'objectif commun est l'estimation d'une transformation entre les deux images, ces techniques reposent d'une part sur la mise en correspondance de primitives géométriques et d'autre part sur l'utilisation directe de la luminance.

4.2.2.1. Estimation du déplacement à partir de la luminance

L'analyse des niveaux de gris de l'image permet d'estimer le mouvement d'une région dans l'image, c'est-à-dire la transformation qui permet d'effectuer le transfert des primitives visuelles d'une image à la suivante. L'avantage de telles approches est qu'elles ne requièrent pas de processus d'extraction ou de mise en correspondance d'indices visuels.

L'idée est donc d'estimer la transformation qui permet de recaler au mieux un modèle 2D des niveaux de gris, souvent appelé *template* $\mathcal{T}$, issu d'une image de référence avec les observations dans l'image courante. Dans le cas le plus simple, on peut partir de l'hypothèse de la conservation de la luminance (Horn, 1987) qui suppose que le signal lumineux du projeté d'un point 3D reste constant dans le temps $\forall \mathbf{p} \in \mathcal{T}, \mathbf{I}_2({}^2tr_1(\mathbf{p})) = \mathbf{I}_1(\mathbf{p})$, le critère de recalage se réécrit alors :

$$ \widehat{{}^2\mathbf{M}_1} = \underset{{}^2\mathbf{M}_1}{\operatorname{argmin}} \ \Delta \ \text{ avec } \ \Delta = \sum_{i=1}^{n} (\mathbf{I}_2({}^2tr_1({}^1\mathbf{p}_i)) - \mathbf{I}_1({}^1\mathbf{p}_i))^2. \quad [4] $$

Cette hypothèse de conservation de la luminance a été largement exploitée pour le suivi de points d'intérêt (petite fenêtre) (Lucas *et al.*, 1981; Shi *et al.*, 1994; Baker *et al.*, 2004) et de motif de grande taille (Hager *et al.*, 1998; Buenaposada *et al.*, 2002; Jurie *et al.*, 2002; Benhimane *et al.*, 2004). L'adoption d'une recherche exhaustive des transformations minimisant ce critère de corrélation étant très peu performante, il est possible de résoudre le problème *via* un processus de minimisation, linéaire ou non, permettant de prendre en compte des mouvements complexes (affines, homographiques, etc.). L'idée est de relier les variations de l'intensité lumineuse aux variations des paramètres de la transformation 2D à estimer. L'algorithme KLT proposé par (Lucas *et al.*, 1981; Shi *et al.*, 1994) et maintes fois repris (par exemple par (Baker *et al.*, 2004)) a été l'une des premières méthodes de suivi reposant sur ces approches différentielles.

Historiquement utilisé pour le suivi de points d'intérêt le KLT a été étendu et optimisé pour le suivi de motif de grande taille. Afin de réduire le coût calculatoire, (Hager *et al.*, 1998) propose une approche permettant d'alléger le calcul de la matrice jacobienne, reliant les variations de l'intensité lumineuse aux variations des paramètres à estimer, en exprimant le gradient de l'image courante en fonction de celui de l'image de référence ce qui permet d'effectuer hors-ligne le plus gros des calculs. (Jurie *et al.*, 2002) propose un cadre d'apprentissage de la pseudo-inverse de la matrice jacobienne. L'avantage est de minimiser les coûts de calcul en-ligne et surtout, l'apprentissage permet d'éviter le développement au premier ordre qui suppose que les niveaux de gris sont une fonction linéaire des paramètres de la transformation. Dans (Benhimane *et al.*, 2004), la minimisation basée sur un développement de premier ordre a été remplacée par une méthode très efficace utilisant une approximation du second ordre rapide, donnant des résultats précis et pouvant gérer de grands déplacements.

4.2.2.2. Estimation du déplacement à partir de primitives géométriques

Les méthodes précédentes permettent dans certains cas d'estimer directement la matrice d'homographie (et donc le déplacement 3D entre deux points de vue). Une alternative à cette approche repose sur la mise en correspondance de primitive géométrique (points d'intérêt, contour, etc.). Dans le cas de points d'intérêt les correspondances sont classiquement fournies par un suivi spatio-temporel (là encore on retrouve des approches différentielles comme le KLT (Lucas *et al.*, 1981; Shi *et al.*, 1994)) ou par des approches de mise en correspondance (ou reconnaissance) de points d'intérêt (Mikolajczyck *et al.*, 2005; Lowe, 2004; Lepetit *et al.*, 2006; Ozuysal *et al.*, 2007).

En considérant deux ensembles de points homologues dans deux images, l'estimation du déplacement peut donc se faire en résolvant le système suivant :

$$\widehat{{}^2\mathbf{M}_1} = \underset{{}^2\mathbf{M}_1}{\arg\min}\ \Delta \ \text{ avec } \ \Delta = \sum_{i=1}^{N} \left({}^2\mathbf{p}_i - {}^2tr_1({}^1\mathbf{p}_i)\right)^2 + \left({}^1\mathbf{p}_i - {}^1tr_2({}^2\mathbf{p}_i)\right)^2 \quad [5]$$

où N est le nombre de primitives visuelles considérées et ${}^{2_i}\mathbf{p} - {}^2tr_1({}^{1_i}\mathbf{p})$ est la distance signée entre les primitives 2D ${}^{2_i}\mathbf{p}$ et ${}^2tr_1({}^{1_i}\mathbf{p})$. Afin de prendre en compte des

erreurs dans l'extraction des primitives, il est souhaitable de minimiser les erreurs croisées dans les deux images. On considère donc les transformations directe (${}^2tr_1$) et indirecte (${}^1tr_2$).

Il est également possible d'exploiter les données obtenues par un processus d'extraction de contour bas-niveau. De manière similaire, il convient alors de chercher la transformation permettant de relier la courbe dans l'image de référence à la courbe extraite de l'image courante. Le problème de l'*aperture* peut alors être résolu en alternant une étape de mise en correspondance et une étape d'estimation de la transformation. On peut citer par exemple l'algorithme ICP (*Iterative Closest Point*) (Fitzgibbon, 2003) ou (Chesi *et al.*, 1999) qui propose un processus de mise en correspondance de points de contour situés sur une courbe couplée avec l'estimation d'une homographie.

4.2.3. *Calcul du déplacement et réalité augmentée*

La pose de la caméra tout au long d'une séquence vidéo peut donc être obtenue à partir de sa pose initiale et de l'intégration du déplacement de la caméra. Si ces méthodes ne requièrent plus de modèle 3D complexe, l'alignement initial du repère de la caméra et d'un repère lié à la scène reste à réaliser. Des solutions existent comme par exemple (Simon *et al.*, 2000) où la présence d'un quadrilatère est nécessaire dans la première image pour calculer la pose initiale. Toutefois il n'est pas nécessaire de le maintenir dans le champ de vision tout au long de la séquence.

Les premières approches reposaient sur la mise en correspondance robuste par corrélation de points d'intérêt (Simon *et al.*, 2000; Simon *et al.*, 2002b) et sur l'estimation d'une homographie plane (voir figure 6a). Plusieurs plans de la même scène peuvent aussi être considérés (Simon *et al.*, 2002a). Des contraintes sur la position relative de ces plans doivent alors être ajoutées dans le processus d'optimisation (voir figure 6a).

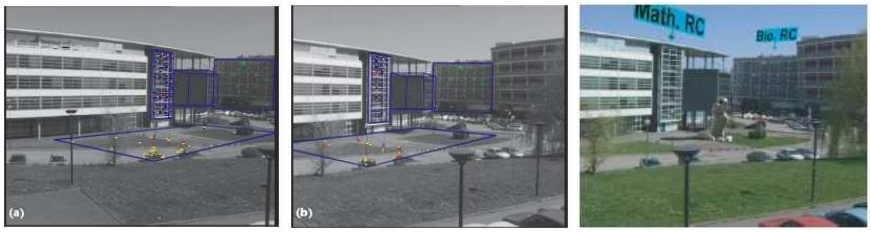


Figure 6. *Réalité augmentée sans modèle : estimation en ligne du déplacement de la caméra en utilisant plusieurs plans (travaux réalisés au Loria par (Simon* et al.*, 2002a))*

L'asservissement visuel virtuel se prête bien à l'estimation des homographies (Comport *et al.*, 2006). Contrairement à (Simon *et al.*, 2002b) les paramètres à estimer sont directement les paramètres du déplacement et non les paramètres de l'homographie. Le KLT est utilisé pour suivre les points d'intérêt entre deux images (voir figure 7).

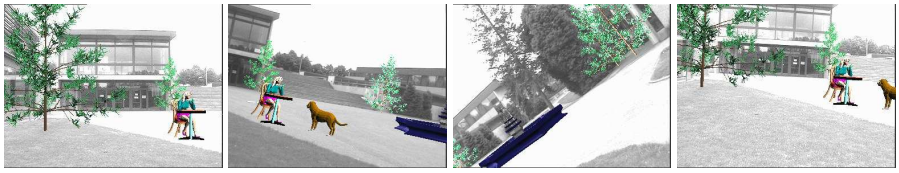


Figure 7. *Réalité augmentée sans modèle : estimation en ligne du déplacement de la caméra par asservissement visuel virtuel (Comport* et al.*, 2006). Des hypothèses sur le mouvement de la caméra (ici une rotation pure) doivent être faites*

L'algorithme ESM (*Efficient Second order Minimization*) (Benhimane *et al.*, 2004) a aussi été utilisé pour des applications de réalité augmentée. Contrairement aux méthodes précédentes l'estimation du déplacement (ici encore modélisé par une homographie) repose sur la mise en correspondance de patchs de grande taille. L'utilisation de plusieurs plans est ici aussi utilisée pour obtenir un suivi plus robuste (voir figure 8).

Figure 8. *Réalité augmentée sans modèle : estimation en ligne du déplacement de la caméra en utilisant un suivi basé sur l'apparence (travaux réalisés à l'Inria Sophia-Antipolis par (Benhimane* et al.*, 2004))*

Toute ces approches permettent de remonter à une estimation fiable du mouvement de la caméra sur des séquences relativement longues. Les effets de bougé (« jitter ») sont très faibles et si l'image de référence est régulièrement remise à jour, les dérives inhérentes au processus d'intégration du mouvement entre deux images successives, sont quasiment inexistantes[5].

4.3. *Suivi hybride multiprimitive*

4.3.1. *Contours ou textures ?*

Nous avons vu dans les sections précédentes deux approches faisant apparaître des avantages et des inconvénients complémentaires. Ainsi, le suivi basé sur les primitives de contour repose généralement uniquement sur des contraintes spatiales car exploitant uniquement les informations d'une seule image, tandis qu'avec les primitives ba-

5. Concernant ce problème de dérive, la remise à jour du template de référence est une étape critique qui a été étudiée dans (Matthews *et al.*, 2004).

sées sur la texture, on a un suivi spatio-temporel de l'objet car reposant sur plusieurs images liées entre elles par les contraintes imposées par la géométrie multivue.

L'avantage d'un suivi spatial est la haute précision des résultats si les contours sont correctement extraits de l'image. Néanmoins, les contours sont en général nombreux. Dès lors, les erreurs d'appariements locaux peuvent entraîner le suivi dans un minimum local ou à une divergence qu'il est difficile de détecter et de rectifier. Par contre, un suivi spatio-temporel fondé sur les niveaux de gris a une faculté de recouvrement d'erreur plus importante, mais peut être soumis à un problème de dérive dû à l'intégration du mouvement. Ce phénomène dégrade le suivi quand le motif « modèle » n'est plus assez représentatif du motif courant ou bien quand les erreurs de mise en correspondances de points d'intérêts s'accumulent. Par ailleurs, les algorithmes reposant sur le suivi de régions manquent souvent de précision pour estimer le mouvement aux frontières de celles-ci.

4.3.2. *Vers un suivi hybride : contours et textures*

L'idée a été, dans un premier temps, d'exploiter un suivi 2D pour initialiser un autre suivi 3D. Typiquement, l'estimation du mouvement d'une région peut servir d'étape préliminaire au suivi de contour (Bascle *et al.*, 1994). (Marchand *et al.*, 1999) propose d'initialiser une phase de calcul de pose fondée sur un recalage sur les gradients par l'estimation du mouvement affine dominant dans l'image. Ce principe a été repris par (Brox *et al.*, 2006) qui se base sur une initialisation par le flot optique.

Néanmoins, enchaîner séquentiellement les différents types de suivis permet certes d'améliorer le problème de l'initialisation mais n'est pas totalement satisfaisant. En effet, chacun des critères successifs n'est fonction que d'un type de primitives visuelles et donc l'estimation laisse de côté à chaque étape des informations pertinentes liées aux autres primitives visuelles. L'idée est alors d'utiliser simultanément plusieurs catégories de primitives visuelles dans un même critère, et non d'utiliser séquentiellement plusieurs critères. Les techniques d'estimation récursive de type Kalman permettent de fusionner dans un même filtre des mesures sur le mouvement de primitives, la couleur et l'orientation des contours pour obtenir un calcul de pose plus robuste (Taylor *et al.*, 2003). De même, (Kyrki *et al.*, 2005) intègre aussi des mesures basées sur les différences d'intensité et le contour dans un filtrage de Kalman. Récemment, plusieurs approches reposant sur une minimisation non linéaire ont introduit dans une même fonction objectif des primitives liées à un modèle de contour et de texture (Masson *et al.*, 2003; Pressigout *et al.*, 2006; Ladikos *et al.*, 2007) (voir figure 9).

(Vacchetti *et al.*, 2004b) utilisent une base d'images fournissant des correspondances 2D/2D de points d'intérêt avec l'image courante conjointement à un calcul de pose basé sur un recalage des contours de l'objet avec la projection d'un modèle 3D. La mise en correspondance de points reste cependant coûteuse en temps de calcul. Ensuite, la pose de l'objet correspondant aux images de la base étant connue, il est possible de fusionner l'ensemble des correspondances 2D/2D et 2D/3D dans un même critère à minimiser.

Figure 9. *Suivi hybride (Pressigout* et al.*, 2006). Images pour (a) : suivi 3D basé contours, (b) : suivi 2D basé sur l'apparence, (c) : suivi hybride. Seul le suivi hybride permet de suivre correctement l'objet sur toute la séquence malgré les spécularités et les modifications d'aspect de l'objet*

Figure 10. *Exemple de suivi multicapteur : deux images successives de la séquence avec un mouvement brusque de la caméra. L'objet incrusté est un cube visible sur les deux images. La deuxième image montre aussi les mises en correspondance de point qu'il a été possible de réaliser grâce à la fusion de l'information vidéo et de la centrale inertielle (travaux réalisés au Loria par (Aron* et al.*, 2007))*

5. Suivi hybride multicapteur

Quand les mouvements de la caméra sont rapides, la qualité du suivi (quelle que soit la méthode retenue) se dégrade rapidement. Un mouvement rapide implique en effet la présence d'un flou sur les forts gradients spatiaux de l'image qui rend difficile tout traitement d'image bas niveau. Il apparaît souhaitable dans ces conditions de coupler l'information en provenance de la caméra avec celle provenant d'un second capteur permettant de calculer de manière autonome le déplacement de la caméra.

Concernant la réalité augmentée le capteur externe le plus utilisé est la centrale inertielle (IMU) (Klein *et al.*, 2003; Ababsa *et al.*, 2007; Aron *et al.*, 2007) mais d'autres capteurs comme le GPS peuvent aussi être considérés (Reitmayr *et al.*, 2007). Dans la plupart des cas (Klein *et al.*, 2003; Aron *et al.*, 2007) l'intégration se fait par des techniques de prédiction/vérification. Cependant il est aussi possible d'utiliser des filtres de Kalman étendus (Foxlin, 1996) ou des filtres à particules (Ababsa *et al.*, 2007). Enfin précisons que pour pouvoir effectuer cette fusion une étape préalable de calibration (Aron *et al.*, 2007) est nécessaire pour connaître la position relative entre le repère de la centrale inertielle et du repère de la caméra.

La figure 10 illustre un suivi multicapteur reposant sur l'utilisation conjointe d'une centrale inertielle et d'une caméra (Aron *et al.*, 2007).

6. Conclusions et remarques

Dans cet article nous avons présenté quelques méthodes de suivi 2D ou 3D permettant de calculer à la cadence vidéo la position de la caméra par rapport à son environnement et donc de permettre l'incrustation d'objets virtuels dans des séquences d'images réelles. Nous nous sommes principalement focalisés sur deux grandes classes d'approches sans considérer un certain nombre d'éléments qui sont pourtant souvent très importants d'un point de vue pratique : les aspects de traitement d'image bas niveau et les aspects d'optimisation/minimisation.

Concernant les aspects bas niveau il est clair que, à part pour les approches reposant directement sur un recalage des niveaux de gris, une mise en correspondance spatio-temporelle d'indice visuel est nécessaire. Des solutions ont été évoquées (KLT...) mais actuellement une tendance forte semble être de considérer des algorithmes efficaces de « reconnaissance » de points d'intérêt (par exemple (Lowe, 2004; Ozuysal *et al.*, 2007; Lepetit *et al.*, 2006)[6]). Concernant les aspects d'optimisation, il convient de préciser que dans la plupart des cas, ces techniques utilisent des méthodes d'estimation robuste (RANSAC, M-Estimateur, etc.) ce qui permet de prendre en compte les inévitables et nombreuses données aberrantes issues des processus de traitement d'image bas niveau. Utilisé dans le contexte de la réalité augmentée elles permettent d'éviter (ou tout au moins de limiter) de manière efficace les effets de bougé (*jitter*) (Berger *et al.*, 1999; Comport *et al.*, 2006; Drummond *et al.*, 2002) ou les dérives (Simon *et al.*, 2002a).

Finalement nous n'avons considéré ici que les aspects de suivi d'objets rigides reposant sur des modèles plus ou moins complexes, des travaux remarquables ont récemment été proposés. Citons, par exemple, les travaux de (Pilet *et al.*, 2007; Gay-Bellile *et al.*, 2007) permettant de considérer des objets déformables ou les travaux

6. Ces algorithmes de mise en correspondance de points d'intérêt permettent par ailleurs de fournir une solution élégante au difficile problème de l'initialisation. En assurant que le système de réalité augmentée pourra se réinitialiser en cas d'échec du processus de suivi, ils lui confèrent une robustesse indispensable à un déploiement réaliste.

sur le SLAM (*Simultaneous Localisation and Mapping*) monoculaire permettant de se passer totalement des modèles de la scène (Skrypnyk *et al.*, 2004; Davison *et al.*, 2002).

7. Bibliographie

Ababsa F.-E., Mallem M., « Hybrid three-dimensional camera pose estimation using particle filter sensor fusion », *Advanced Robotics*, vol. 21, n° 1, p. 165-181, Janvier, 2007.

Aron M., Simon G., Berger M.-O., « Use of inertial sensors to support video tracking », *Computer Animation and Virtual Worlds*, vol. 18, n° 1, p. 57-68, Janvier, 2007.

Azuma R., « A Survey of Augmented Reality », *Presence : Teleoperators and Virtual Environments*, vol. 6, n° 4, p. 355-385, Août, 1997.

Azuma R., Baillot Y., Behringer R., Feiner S., Julier S., MacIntyre B., « Recent Advances in Augmented Reality », *IEEE Computer Graphics and Application*, vol. 21, n° 6, p. 34-47, Novembre, 2001.

Baker S., Matthews I., « Lucas-kanade 20 years on : A unifying framework », *Int. Journal of Computer Vision*, vol. 56, n° 3, p. 221-255, 2004.

Bascle B., Bouthemy P., Deriche N., Meyer F., « Tracking Complex Primitives in an image sequence », *Int. Conf. on Pattern Recognition, ICPR'94*, Jerusalem, p. 426-431, Octobre, 1994.

Benhimane S., Malis E., « Real-time image-based tracking of planes using efficient second-order minimization », *IEEE/RSJ Int. Conf. on Intelligent Robots Systems*, vol. 943-948, Sendai, Japan, p. 1, Octobre, 2004.

Berger M.-O., Wrobel-Dautcourt B., Petitjean S., Simon G., « Mixing Synthetic and Video Images of an Outdoor Urban Environment », *Machine Vision and Applications*, vol. 11, n° 3, p. 145-159, 1999.

Boufama B., Mohr R., « Epipole and fundamental matrix estimation using virtual parallax », *IEEE Int. Conf. on Computer Vision*, p. 1030-1036, 1995.

Brox T., Rosenhahn B., Cremers D., Seidel H.-P., « High accuracy optical flow serves 3-D pose tracking : exploiting contour and flow based constraints », *in* A. Leonardis, H. Bischof, A. Pinz (eds), *European Conf. on Computer Vision, ECCV'06*, vol. 3952 of *LNCS*, Springer, Graz, Austria, p. 98-111, May, 2006.

Brown D., « Close-Range Camera Calibration », *Photogrammetric Engineering*, vol. 4, n° 2, p. 127-140, Mars, 1971.

Buenaposada J., Baumela L., « Real-time tracking and estimation of plane pose », *IAP Int. Conf. on Pattern Recognition, ICPR'02*, vol. 2, Québec, Canada, p. 697-700, Août, 2002.

Chesi G., Malis E., Cipolla R., « Collineation estimation from two unmatched views of an unknown planar contour for visual servoing », *British Machine Vision Conference*, 1999.

Comport A., Marchand E., Pressigout M., Chaumette F., « Real-time markerless tracking for augmented reality : the virtual visual servoing framework », *IEEE Trans. on Visualization and Computer Graphics*, vol. 12, n° 4, p. 615-628, Juillet, 2006.

Davis L., Hamza-Lup F., Rolland J., « A Method for Designing Marker-Based Tracking Probes », *ACM/IEEE Int. Symp. on Mixed and Augmented Reality, ISMAR'04*, p. 120-129, 2004.

Davison A., Murray D., « Simultaneous Localization and Map-Building Using Active Vision », *IEEE Trans. on PAMI*, vol. 24, n° 7, p. 865-880, Juillet, 2002.

Dementhon D., Davis L., « Model-Based Object Pose in 25 Lines of Codes », *Int. J. of Computer Vision*, vol. 15, n° 1-2, p. 123-141, 1995.

Dhome M., Lapresté J.-T., Rives G., Richetin M., « Determination of the attitude of modelled objects of revolution in monocular perspective vision », *European Conf. on Computer Vision, ECCV'90*, vol. 427 of *Lecture Notes in Computer Science*, Antibes, p. 475-485, Avril, 1990.

Dhome M., Richetin M., Lapresté J.-T., Rives G., « Determination of the Attitude of 3D Objects from a Single Perspective View », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, n° 12, p. 1265-1278, Décembre, 1989.

Drummond T., Cipolla R., « Real-Time Visual Tracking of Complex Structures », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, n° 7, p. 932-946, Juillet, 2002.

Faugeras O., Lustman F., « Motion and structure from motion in a piecewise planar environment », *Int. Journal of Pattern Recognition and Artificial Intelligence*, vol. 2, n° 3, p. 485-508, 1988.

Faugeras O., Toscani G., « Camera calibration for 3D computer vision. », *Proc Int. Workshop on Machine Vision and Machine Intelligence*, Tokyo, p. 240-247, Février, 1987.

Fiala M., « ARTag, a fiducial marker system using digital techniques », *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'2005*, p. 590-596, June, 2005.

Fischler N., Bolles R., « Random Sample Consensus : A Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography », *Communication of the ACM*, vol. 24, n° 6, p. 381-395, Juin, 1981.

Fitzgibbon A., « Robust registration of 2D and 3D point sets », *Image and Vision Computing*, vol. 21, n° 12-13, p. 1145-1153, Décembre, 2003.

Foxlin E., « Inertial head-tracker sensor fusion by a complementaryseparate-bias Kalman filter », *IEEE Int. Symp. on Virtual Reality*, Santa Clara, CA, p. 185-194, Mars, 1996.

Friedrich W., « ARVIKA – Augmented Reality for Development, Production and Service », *IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR'02)*, Darmstadt, Germany, p. 295-304, Septembre, 2002.

Ganapathy S., « Decomposition of Transformation Matrices for Robot Vision », *Pattern Recognition Letter*, vol. 2, p. 401-412, 1984.

Gay-Bellile V., Bartoli A., Sayd P., « Deformable Surface Augmentation in spite of Self-Occlusions », *ACM/IEEE Int. Symp. on Mixed and Augmented Reality, ISMAR'07*, Nara, Japan, Novembre, 2007.

Genc Y., Riedel S., Souvannavong F., Akinlar C., Navab N., « Marker-less Tracking for AR : A Learning-Based Approach », *IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR'02)*, Darmstadt, Germany, p. 3-6, Septembre, 2002.

Gennery D., « Visual tracking of known three-dimensional objects », *Int. J. of Computer Vision*, vol. 7, n° 3, p. 243-270, 1992.

Hager G., Belhumeur P., « Efficient Region Tracking With Parametric Models of Geometry and Illumination », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, n° 10, p. 1025-1039, Octobre, 1998.

Haralick R., Joo H., Lee C., Zhuang X., Vaidya V., Kim M., « Pose estimation from corresponding point data », *IEEE Trans on Systems, Man and Cybernetics*, vol. 19, n° 6, p. 1426-1445, Novembre, 1989.

Hartley R., Zisserman A., *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2001.

Horaud R., Conio B., Leboulleux O., Lacolle B., « An analytic solution for the perspective 4-points problem », *Computer Vision, Graphics and Image Processing*, vol. 47, n° 1, p. 33-44, Juillet, 1989.

Horn B., *Robot Vision*, MIT Press, Cambridge, 1987.

Jurie F., Dhome M., « Hyperplane Approximation for Template Matching », *IEEE Trans. on PAMI*, vol. 24, n° 7, p. 996-1000, Juillet, 2002.

Kato H., Billinghurst M., « Marker Tracking and HMD Calibration for a video-based Augmented Reality Conferencing System », *ACM/IEEE Int. Workshop on Augmented Reality, IWAR'99*, San Francisco, CA, p. 85-95, Octobre, 1999.

Klein G., Drummond T., « Robust Visual Tracking for Non-Instrumented Augmented Reality », *ACM/IEEE Int. Symp. on Mixed and Augmented Reality, ISMAR'03*, Tokyo, Japan, p. 113-122, Octobre, 2003.

Koller D., Daniilidis K., Nagel H.-H., « Model-Based Object Tracking in Monocular Image Sequences of Road Traffic Scenes », *Int. Journal of Computer Vision*, vol. 10, n° 2, p. 257-281, Juin, 1993.

Kyrki V., Kragic D., « Integration of Model-based and Model-free Cues for Visual Object Tracking in 3D », *IEEE Int. Conf. on Robotics and Automation, ICRA'05*, Barcelona, Spain, p. 1566-1572, Avril, 2005.

Ladikos A., Benhimane S., Navab N., « A Real-Time Tracking System Combining Template-Based and Feature-Based Approaches », *Int. Conf. on Computer Vision Theory and Applications*, 2007.

Lavest J.-M., Rives G., *Étalonnage des capteurs de vision.*, Traité IC2, Hermès, chapter 1, Mai, 2003.

Lepetit V., Fua P., « Monocular Model-Based 3D Tracking of Rigid Objects : A Survey », *Foundations and Trends in Computer Graphics and Vision*, vol. 1, n° 1, p. 1-89, Octobre, 2005.

Lepetit V., Fua P., « Keypoint Recognition Using Randomized Trees », *IEEE Trans. on PAMI*, vol. 28, n° 9, p. 1465-1479, Septembre, 2006.

Li M., « Correspondence Analysis Between The Image Formation Pipelines of Graphics and Vision », *Spanish Symposium on Pattern Recognition and Image Analysis*, Benicasim, Castellón, Spain, p. 187-192, Mai, 2001.

Liu Y., Huang T., Faugeras O., « Determination of Camera Location from 2D to 3D Line and Point Correspondences », *IEEE Trans. on PAMI*, vol. 12, n° 1, p. 28-37, Janvier, 1990.

Lowe D., « Three-dimensional object recognition from single two-dimensional images », *Artificial Intelligence*, vol. 31, n° 3, p. 355-394, Mars, 1987.

Lowe D., « Fitting parameterized three-dimensional models to images », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 13, n° 5, p. 441-450, Mai, 1991.

Lowe D., « Distinctive image features from scale-invariant keypoints », *Int. Journal of Computer Vision*, vol. 60, n° 2, p. 91-110, 2004.

Lu C., Hager G., Mjolsness E., « Fast and Globally Convergent Pose Estimation from Video Images », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, n° 6, p. 610-622, Juin, 2000.

Lucas B., Kanade T., « An Iterative Image Registration Technique with an Application to Stereo Vision », *Int. Joint Conf. on Artificial Intelligence, IJCAI'81*, p. 674-679, 1981.

Marchand E., Bouthemy P., Chaumette F., Moreau V., « Robust real-time visual tracking using a 2D-3D model-based approach », *IEEE Int. Conf. on Computer Vision, ICCV'99*, vol. 1, Kerkira, Greece, p. 262-268, Septembre, 1999.

Marchand E., Chaumette F., « Virtual Visual Servoing : a framework for real-time augmented reality », *in* G. Drettakis, H.-P. Seidel (eds), *EUROGRAPHICS'02 Conf. Proceeding*, vol. 21(3) of *Computer Graphics Forum*, Saarebrücken, Germany, p. 289-298, Septembre, 2002.

Martin F., Horaud R., « Multiple Camera Tracking of Rigid Objects », *Int. Journal of Robotics Research*, vol. 21, n° 2, p. 97-113, Février, 2002. (INRIA RR-4268, september 2001).

Masson L., Jurie F., Dhome M., « Contour/Texture Approach for Visual Tracking », *13th Scandinavian Conf. on Image Analysis, SCIA 2003*, vol. 2749 of *Lecture Notes in Computer Science*, Springer, p. 661-668, 2003.

Matthews I., Ishikawa T., Baker S., « The Template Update Problem », *IEEE Trans. on PAMI*, vol. 26, n° 6, p. 810-815, Juin, 2004.

Mikolajczyck K., Schmid C., « A Performance Evaluation of Local Descriptors », *IEEE Trans. on PAMI*, vol. 27, n° 10, p. 1615-1630, Octobre, 2005.

Navab N., « Industrial Augmented Reality : Challenges in Design and Commercialization of Killer Apps », *IEEE Int. Symp. on Mixed and Augmented Reality, ISMAR'03*, Tokyo, Japan, p. 2-7, Octobre, 2003.

Oberkampf D., Dementhon D., Davis L., « Iterative Pose Estimation Using Coplanar Feature Points », *Computer Vision and Image Understanding*, vol. 63, n° 3, p. 495-511, Mai, 1996.

Ozuysal M., Fua P., Lepetit V., « Fast Keypoint Recognition in Ten Lines of Code », *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 2007.

Phong T.-Q., Horaud R., Yassine A., Tao P.-D., « Object Pose from a 2D to 3D point and line correspondance », *Int. Journal of Computer Vision*, vol. 15, n° 3, p. 225-243, Juillet, 1995.

Pilet J., Lepetit V., Fua P., « Fast Non-Rigid Surface Detection, Registration and Realistic Augmentation », *Int. Journal of Computer Vision*, vol. 76, n° 2, p. 109-122, Février, 2007.

Pressigout M., Marchand E., « Hybrid tracking algorithms for planar and non-planar structures subject to illumination changes », *ACM/IEEE Int. Symp. on Mixed and Augmented Reality, ISMAR'06*, Santa Barbara, CA, p. 52-55, October, 2006.

Reitmayr G., Drummond T., « Initialisation for Visual Tracking in Urban Environments », *IEEE/ACM Int. Symp. on Mixed and Augmented Reality, ISMAR'07*, p. 161-172, Nov, 2007.

Safaee-Rad R., Tchoukanov I., Benhabib B., Smith K., « Three dimentional location estimation of circular features for machine vision », *IEEE Trans. on Robotics and Automation*, vol. 8, n° 2, p. 624-639, Octobre, 1992.

Shi J., Tomasi C., « Good Features to Track », *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'94*, Seattle, Washington, p. 593-600, Juin, 1994.

Simon G., Berger M.-O., « A Two-Stage Robust Statistical Method for Temporal Registration from Features of Various Type », *Int. Conf. on Computer Vision, ICCV'98*, Bombay, India, p. 261-266, Janvier, 1998.

Simon G., Berger M.-O., « Pose Estimation for Planar Structures », *IEEE Computer Graphics and Applications*, vol. 22, n° 6, p. 46-53, Novembre, 2002.

Simon G., Berger M.-O., « Reconstructing While Registering : A Novel Approach for Markerless Augmented Reality », *IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR'02)*, Darmstadt, Germany, p. 285-294, Septembre, 2002b.

Simon G., Fitzgibbon A., Zisserman A., « Markerless Tracking using Planar Structures in the Scene », *IEEE/ACM Int. Symp. on Augmented Reality*, Munich, Germany, p. 120-128, Octobre, 2000.

Skrypnyk I., Lowe D., « Scene modelling, recognition and tracking with invariant image features », *ACM/IEEE Int. Symp. on Mixed and Augmented Reality, ISMAR'2004*, Arlington, VA, p. 110- 119, Novembre, 2004.

Taylor G., Kleeman L., « Fusion of multimodal visual cues for model-based object tracking », *Australasian Conference on Robotics and Automation (ACRA2003)*, Brisbane,Australia, December, 2003.

Vacchetti L., Lepetit V., Fua P., « Combining Edge and Texture Information for Real-Time Accurate 3D Camera Tracking », *ACM/IEEE Int. Symp. on Mixed and Augmented Reality, ISMAR'2004*, vol. 2, Arlington, Va, p. 48-57, Novembre, 2004.

Vacchetti L., Lepetit V., Fua P., « Stable Real-Time 3D Tracking Using Online and Offline Information », *IEEE Trans. on PAMI*, vol. 26, n° 10, p. 1385-1391, October, 2004.

Wuest H., Vial F., Stricker D., « Adaptive line tracking with multiple hypotheses for augmented reality », *ACM/IEEE Int. Symp. on Mixed and Augmented Reality, ISMAR'05*, Washington DC, p. 62-69, Novembre, 2005.

Zhang X., Fronz S., Navab N., « Visual marker detection and decoding in AR systems : a comparative study », *IEEE/ACM Int. Symp. on Mixed and Augmented Reality, ISMAR'2002*, p. 97-106, 2002.

Article reçu le 14 décembre 2007
Accepté après révisions le 24 juillet 2008

Éric Marchand *est chargé de recherche INRIA au centre de recherche INRIA Rennes-Bretagne Atlantique dans le projet Lagadic. Ses recherches sont axées sur la vision robotique, l'asservissement visuel. Depuis quelques années il s'interesse principalement au suivi d'objets (suivi robuste, précis, temps réel) pour des applications robotiques (environnements nucléaire, sous-marin, spatiaux) ou pour la réalité augmentée.*

Muriel Pressigout *est maître de conférences à l'INSA de Rennes. Ses travaux de thèse soutenue en 2006 à l'Université de Rennes 1 au sein du projet Lagadic à l'IRISA portent sur la vision par ordinateur et plus spécifiquement le suivi-temps réel pour des applications de réalité augmentée et de robotique. Actuellement, ses centres d'intérêt portent sur la vision par ordinateur dans le cadre du codage vidéo.*

ANNEXE POUR LE SERVICE FABRICATION

A FOURNIR PAR LES AUTEURS AVEC UN EXEMPLAIRE PAPIER
DE LEUR ARTICLE ET LE COPYRIGHT SIGNE PAR COURRIER
LE FICHIER PDF CORRESPONDANT SERA ENVOYE PAR E-MAIL

1. ARTICLE POUR LA REVUE :
   RSTI - TSI – 28/2009. Réalités virtuelle et augmentée

2. AUTEURS :
   Eric Marchand * *— Muriel Pressigout***

3. TITRE DE L'ARTICLE :
   Un panorama des techniques de suivi visuel temps réel pour la réalité augmentée

4. TITRE ABRÉGÉ POUR LE HAUT DE PAGE MOINS DE 40 SIGNES :
   Suivi temps réel pour la réalité augmentée

5. DATE DE CETTE VERSION :
   8 juillet 2009

6. COORDONNÉES DES AUTEURS :

– adresse postale :

* INRIA Rennes-Bretagne Atlantique, IRISA, F-35000 Rennes, France

marchand@irisa.fr

** INSA Rennes, IETR, UMR 6164, F-35000, Rennes

muriel.pressigout@insa-rennes.fr

– téléphone : 02 99 84 74 27

– télécopie : 02 99 84 71 71

– e-mail : marchand@irisa.fr

7. LOGICIEL UTILISÉ POUR LA PRÉPARATION DE CET ARTICLE :
   LaTeX, avec le fichier de style `article-hermes2.cls`,
   version 1.23 du 17/11/2005.

8. FORMULAIRE DE COPYRIGHT :
   Retourner le formulaire de copyright signé par les auteurs, téléchargé sur :
   `http://www.revuesonline.com`

SERVICE ÉDITORIAL – HERMES-LAVOISIER
14 rue de Provigny, F-94236 Cachan cedex
Tél. : 01-47-40-67-67
E-mail : revues@lavoisier.fr
Serveur web : http://www.revuesonline.com