REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université MENTOURI - Constantine -
Département d'Informatique

N° d'ordre : 38/TS/2008

Série : 04/inf/2008

THÈSE
Présentée par

BOUZENADA Mourad

Pour obtenir le titre de
Doctorat en sciences en Informatique

Incrustation d'objets virtuels dans des séquences vidéo pour la réalité augmentée temps réel

Directeur de thèse

Pr. BATOUCHE Mohamed Chawki

Composition du jury :

Pr. Benmohammed Mohamed	Université Mentouri - Constantine	Président
Dr. Chikhi Salim	Université Mentouri - Constantine	Rapporteur
Pr. Djedi Nourredine	Université Khider - Biskra	Examinateur
Dr. Babahennini Chawki	Université Khider - Biskra	Examinateur
Dr. Mostefai Sihem	Université Mentouri-Constantine	Examinateur
Dr. Hachouf Fella	Université Mentouri-Constantine	Examinateur

Thèse préparée au sein du laboratoire LIRE, Université Mentouri de Constantine

2008

ملخص

تعتبر الإضافة للحقيقة مجال جديد للبحث العلمي وهي عبارة عن نقطة تواصل بين عدة مجالات منها: معالجة الصور، النظر الاصطناعي، الحقيقة الخيالية، الخ. الهدف الرئيسي للإضافة للحقيقة هو الدمج بين الحقيقة والخيال. حيث تتمثل الحقيقة في عروض خاصة تنتمي لمكان معين نريد أن نضيف له شيئا ما. أما الخيال فهو عبارة عن عروض مصطنعة عن طريق الحاسوب أو عروض حقيقية مسجلة في ذاكرته. هذا الدمج يكون بإضافة أشياء خيالية لصور فيديو حقيقية. ليست هذه الإضافة عشوائية ولكن إضافة يراعى فيها من جهة الانسجام بين الأشياء الحقيقية التي تنتمي لصور الفيديو والأشياء الخيالية المضافة ومن جهة أخرى تقليص الوقت المستغرق للقيام بهذه العملية.نستطيع القول إذا، بأن أحسن نظام ينتمي للإضافة الحقيقية، هو نظام يستطيع أن يبقي في أي وقت من الأوقات الانسجام بين الأشياء الحقيقية والأشياء الخيالية. ولكي يتسنى ذلك لابد لهذا النظام أن يحتوي على متابعة دقيقة لمكان وتوجه الأشياء الحقيقية للعرض المصور.

تعتبر المتابعة للأشياء خطوة هامة في أي نظام من نظم الإضافة للحقيقة. لذا تناولت في مذكرتي هته، هذا الجانب بدراسة هذه الخطوة باعتبارها هامة من جهة ومن جهة أخرى الخطوة التي تستغرق الوقت الأكبر في عملية الدمج.

بعد دراسة معمقة لمختلف طرق المتابعة، استطعت أن أقدم طريقة جديدة ناجعة في متابعة أشياء تنتمي لصور فيديو والتي تمر بثلاث مراحل:

- مرحلة البداية وتتمثل في تعلم الحركات الممكنة للشيء الذي نريد إتباعه. هذا التعلم حققناه عن طريق استعمال شبكة خلايا عصبية اصطناعية.
- مرحلة المتابعة تعتمد على نتائج المرحلة السابقة لكي تتحصل على مكان الشيء المتبع.
- مرحلة الإضافة والتي تمكن من دمج الشيء الخيالي في المكان الذي تحصلنا عليه في المرحلة السابقة.

Abstract

Augmented reality (AR) combines real world and virtual objects which can be previously saved images or computer generated data. It enhances a user's view of the real environment by adding virtual objects in a realistic manner. Virtual objects display information that the user cannot directly detect with his senses. So, to look realistic, the virtual objects must be properly aligned with real-world objects. This issue is known as registration problem. It is one of the most basic problems currently limiting Augmented Reality applications.

The best AR system is the system which keeps at any time a good alignment between virtual and real objects. This required registration stability is obtained by accurate real-time trackers.

We are interested in this thesis in object tracking because of its importance in an AR system. Also, object tracking represents the most consuming part of time in an augmented reality application. So, we have proposed a global-based approach for object tracking to get accurate results using only information about target object. This information has been obtained, during an offline stage, over a set of output and input samples to determine the relationship between intensity and position variations of the target object. Then, the virtual object is projected at the position of the target object by applying a planar homography.

Résumé

La réalité augmentée (RA) est une discipline récente. Elle est située au carrefour de plusieurs disciplines. Elle a pour but l'intégration simultanée du réel et du virtuel. Le réel correspond aux scènes concrètes relatives à un environnement donné qu'on désire augmenter avec d'autres. Le virtuel correspond aux scènes synthétiques construites par ordinateur ou d'autres scènes réelles préenregistrées. L'augmentation consiste alors à les mixer de manière intelligente et en temps réel. Cela insinue la génération, parallèlement à la scène réelle, d'une scène finale où tout parait concret. Pour cela, il faut résoudre le problème essentiel de la coexistence (alignement ou 'registration') des objets réels et virtuels dans un même espace tridimensionnel.

Ainsi, un bon système de RA est un système qui permet de garder à tout moment un alignement correct entre les objets virtuels et les objets réels. Ceci est possible, grâce à un bon suivi de la position et de l'orientation des objets réels de la scène.

Le suivi d'objets constitue alors une phase importante dans un système de RA. Ainsi, nous nous sommes intéressés dans cette thèse au suivi d'objets (tracking) puisqu'en plus de son importance, il représente la partie la plus consommatrice de temps dans une application de réalité augmentée. Nous avons proposé une méthode de suivi d'objets basée pixels qui est constituée de trois étapes. L'étape d'initialisation (étape offline) consiste à faire un apprentissage sur le mouvement du motif à suivre. Cet apprentissage a été réalisé à l'aide d'un réseau de neurones artificiels. L'étape de suivi (étape online) se base sur les résultats de l'étape précédente pour déterminer la position de l'objet cible. Enfin, l'étape d'augmentation (étape online) permet d'incruster un objet virtuel à la position déterminée lors de la phase précédente en lui appliquant une homographie planaire.

Table des matières

Table des figures

A Mon défunt frère Mohsen Chiheb……

Remerciements

Les travaux de recherche décrits dans cette thèse ont été menés entre 2003 et 2007 principalement au laboratoire de recherche LIRE du département d'informatique à l'université Mentouri – Constantine. Par ailleurs, certains des travaux décrits ont été réalisés au laboratoire LIFL de Lille et à L'Irisa/Inria de Rennes.

Je voudrais commencer ce document en remerciant ceux avec qui j'ai travaillé directement ou indirectement. J'espère qu'ils en auront un souvenir aussi agréable que moi.

Merci au Pr. Batouche Mohamed chawki d'avoir été à l'écoute, d'avoir fait l'effort de me diriger durant toute cette période.

Merci au Dr. Chikhi Salim d'avoir accepté la charge de rapporter cette thèse, merci pour l'intérêt qu'il y a porté et pour ses remarques.

Merci au Pr. Benmohammed Mohamed de m'avoir fait l'honneur de présider mon jury de thèse.

Merci au Pr. Djedi Noureddine, au Dr. Babahennini Chawki, au Dr. Abassen Sihem et au Dr. Hachouf Fella d'avoir participé à mon Jury de thèse et pour l'intérêt porté à mon travail.

Merci au Pr. Bouatouch Kadi, pour son accueil chaleureux à L'Irisa/Inria de Rennes et pour l'intérêt porté à mon travail.

Merci à tous les amis, du labo et d'ailleurs, pour avoir fait que ces années de thèse soient des années géniales. Merci à: Reda Bahri, Hacene Belhadef, Mohamed Berkane, Ilhem Labed, Mohamed el-habib Laraba, Smaine Mazouzi, Sihem Meshoul, Brahim Nini, Djamel-eddine Saidouni, Hichem Talbi.

Merci à ma famille et spécialement, Mon père, Ma mère, Ma femme, Mes enfants. Merci d'avoir supporté mon indisponibilité et mon ingratitude sans le moindre reproche. Merci d'être toujours là, et surtout quand ça va mal.

CHAPITRE 1

Réalité augmentée : entre rêve et réalité

La convergence de la recherche vers la réalité augmentée paraît logique, suite à la puissance croissante des ordinateurs et la maîtrise presque totale du domaine de traitement d'images et des séquences vidéo. Il fallait exploiter cette évolution à d'autres fins. Dans un premier temps, la réalité virtuelle était déjà un grand succès d'intégration de l'informatique dans le milieu naturel de l'être humain. Elle a transformé les données numériques en une sensation naturelle sous forme d'interaction. Cependant tout est resté virtuel.

La réalité augmentée (RA) a pour but l'intégration simultanée du réel et du virtuel (Fig. I.1). Le réel correspond à des scènes concrètes relatives à un environnement donné et le virtuel à des scènes synthétiques construites par ordinateur et n'ayant pas d'existence réelle. Son problème est alors celui de mixer ces deux types de scènes de manière intelligente. Cette intelligence sous-entend la construction d'une scène finale où tout ‘parait’ réel (Fig. I.2). Elle tente de faire en sorte que l’utilisateur interagisse avec la scène réelle comme si les objets virtuels ajoutés y faisaient partie.

L'idée de la réalité augmentée tire son originalité de la Réalité Virtuelle (RV). Cette dernière qui fut beaucoup investie durant des années passées, possède une assise assez robuste et dénombre des résultats prolifiques. Cependant, elle s'est heurtée à la limite de ses domaines d'application qui restent restreints à ceux virtuels.

Fig. I.1 - Deux images d'une vidéo augmentée par un objet virtuel (véhicule rouge)

La réalité augmentée, quand à elle, effectue une incrustation du virtuel dans le réel, ou l'inverse, en ce sens que, la réalité virtuelle n'est pas une partie de la réalité augmenté, mais en est une continuité. Milgram [66] présente cette relation comme étant un continuum entre deux environnements, réel et virtuel. Ainsi, il est important de dégager la différence entre les deux:

- En Réalité virtuelle, l'utilisateur est complètement plongé dans un monde virtuel, reconstitué avec des données informatiques: il est coupé du monde réel.
- Par contre, en Réalité augmentée, l'utilisateur est maintenu au contact de son environnement réel. D'où la nécessité d'un nouveau cadre de définition et de conception de la réalité augmentée par rapport à la réalité virtuelle.

Fig. I.2 – Scènes augmentées où il est difficile de faire la différence entre objets réels et virtuels.

L'augmentation peut être de nature différente selon son objet et son contexte. D'ailleurs, il est possible de la scinder en deux types:

- L’exécution augmentée: un exemple consiste à porter un "badge" qui permet d’ouvrir une porte sans la toucher. Ce type d'augmentation est peu touché par la recherche.
- La perception augmentée: par exemple, affichage en temps réel de la position d'un outil chirurgical par rapport à la trajectoire planifiée, avant l’intervention.

Ainsi, son objectif majeur consiste à permettre une interaction (généralement visuelle) de l’utilisateur, en temps réel, avec des objets virtuels. Il peut prendre, faire déplacer ou glisser ou même détruire les objets. Cet objectif induit plusieurs contraintes à respecter, dont l'idée essentielle tourne autour de la résolution des problèmes de:

- La coexistence (alignement ou 'registration') des objets réels et virtuels dans un même espace tridimensionnel.
- Respect des lois naturelles par les objets virtuels.

La RA vise à définir des méthodes et des outils permettant d'augmenter la vision de l'utilisateur par l'ajout d'objets ou d'informations générées par un modèle informatique. Cette augmentation peut prendre la forme d'étiquettes, d'objets 3D de synthèse ou de modifications d'ombrage dans la scène. Les applications de ce concept concernent de nombreux domaines. Citons par exemple : la conception d'études d'impacts permettant de voir comment apparaîtrait un nouveau bâtiment dans une vidéo tournée sur le site d'implantation, la conception de simulateurs de tout genre permettant à un utilisateur d'apprendre des gestes fondamentaux en visualisant des données réelles, la conception d'effets spéciaux pour l'industrie cinématographique . . .

Dans un système de réalité augmentée, la vue réelle de la scène est augmentée en superposant des objets de synthèse sur les images de telle sorte qu'ils soient projetés dans la scène avec le même point de vue que celui adopté par la caméra lors du tournage. Le calcul du point de vue pour chaque image, c'est à dire la possibilité de calculer la position de la caméra ainsi que certaines caractéristiques intrinsèques de l'optique, est donc primordial pour une application de réalité augmentée.

Les prototypes déjà réalisés et les objectifs fixés, ainsi que la puissance des ordinateurs qui progresse en croissance, font de la réalité augmentée un nouvel axe de recherche très prometteur. Son importance réside dans le fait que les objets virtuels fournissent de l'information non détectable directement par les sens. En d'autres mots, l'imaginaire se fait percevoir comme réel. Elle peut également présenter une aide à l'utilisateur pour l'accomplissement de tâches réelles. C'est le cas, par exemple, des prototypes pour l'assemblage de câbles électriques d'un boeing ou de celui de la visualisation d'un fœtus virtuel à l'intérieur d'utérus d'une patiente enceinte.

Les méthodes à mettre en oeuvre pour le calcul du point de vue différent cependant selon les applications visées : dans le cas des études d'impacts ou des effets spéciaux, il s'agit de post production car le calcul des points de vue est réalisé une fois la séquence acquise et il est donc possible de calculer l'ensemble des points de vue de manière robuste en tenant compte de la redondance d'information présente dans les images. A l'opposé, les applications nécessitant une interactivité avec l'utilisateur (jeux vidéos, simulateurs) nécessitent de calculer le point de vue itérativement à chaque image perçue et ceci à la cadence vidéo. C'est ce dernier type d'activités que nous souhaitons aborder dans cette thèse.

Un bon système de RA est un système qui permet de garder à tout moment un alignement correct entre les objets réels et virtuels. Cet alignement correct ne se réalise que grâce à un suivi efficace et rigoureux de la position et de l'orientation des objets réels. Ainsi, nous nous sommes intéressés dans cette thèse au suivi d'objets puisqu'en plus de son importance dans l'alignement correct, il constitue la partie la plus consommatrice du temps de calcul dans un système de RA.

Nous avons d'abord pris comme exemple d'application : le tableau magique. En effet, un tableau magique n'est autre qu'un tableau blanc conventionnel amplifié par des services électroniques capables de contourner ses insuffisances intrinsèques, à savoir: la réorganisation spatiale des inscriptions, l'archivage, la diffusion et la collaboration synchrone à distance. Il fait partie des systèmes fortement couplés [27, 28, 42, 89, 90, 91] ou les représentations virtuelles et physiques sont parfaitement synchronisées. Les objets physiques sont suivis de manière continue en temps réel. Plus exactement, il

appartient aux systèmes de réalité augmentée qui renforcent le monde réel avec des objets virtuels.

Un système de tableau magique est composé alors d'un tableau blanc conventionnel sur lequel on écrit avec les feutres usuels à encre effaçable, d'une caméra vidéo mobile pour capturer les inscriptions se trouvant sur le tableau blanc, d'un ordinateur permettant de traiter le flux vidéo de la caméra et d'un projecteur pour afficher les retours d'information sur le tableau blanc.

L'une des principales fonctions du tableau magique est le suivi du doigt de l'utilisateur pour lui permettre d'indiquer au système des emplacements du tableau. La méthode choisie jusqu'ici est le suivi par corrélation. Cette méthode n'est applicable qu'aux translations effectuées dans un plan parallèle à l'image.

Pour pallier aux limites de cette méthode, nous avons proposé une approche basée sur l'étude réalisée par François Bérard [9, 11, 12] et l'algorithme de suivi de Jurie [50] pour doter la cible (doigt de l'utilisateur) d'une meilleure flexibilité dans le mouvement (translation + rotation) en temps réel. L'idée principale de cette contribution consiste à représenter la zone de recherche par un filet d'ellipses couvrant la surface de toute cette zone au lieu de parcourir toutes les sous parties de l'image ayant une même taille que le motif. Le processus complet passe en premier lieu par une phase de détection durant laquelle l'apparence de la cible est mémorisée dans le motif, puis par une phase de suivi composée de deux étapes: une étape d'apprentissage (offline) et une étape de suivi (online).

Ensuite, nous avons proposé une méthode de suivi d'objets [17] plus générale, basée pixels. Elle est constituée de deux étapes : Offline et Online. L'étape offline consiste à faire un apprentissage sur le mouvement du motif à suivre. Cet apprentissage a été réalisé à l'aide des réseaux de neurones. L'étape online se base sur les résultats de l'étape précédente pour déterminer la position de l'objet cible.

Par ailleurs, nous avons intégré cette méthode dans un processus d'augmentation de séquences vidéo [19].

La suite de cette thèse est organisée comme suit : Au chapitre 2, une présentation est faite sur le principe de l'augmentation, les outils, les systèmes existants et les domaines d'applications inhérents à la réalité augmentée temps réel. Dans le chapitre 3, l'accent est mis sur la particularité des systèmes de RA dans les environnements non préparés. Une synthèse des différentes méthodes de suivi d'objets dans des séquences vidéo est faite au chapitre 4. Le chapitre 5 est consacré à la présentation de la nouvelle méthode de suivi de doigt que nous avons proposée pour le tableau magique. Au chapitre 6, une description détaillée de la méthode de suivi proposée est faite. Cette méthode est considérée comme une extension de la méthode proposée au chapitre précédant dans le sens où elle permet de suivre un motif complexe quelconque. Enfin, au chapitre 7, une discussion critique est faite sur tout ce qui a pu être réalisé. Aussi, un recensement a été fait de tout ce qui peut être réalisé prochainement.

CHAPITRE 2

Réalité augmentée temps réel: Définitions, Problématique et outils

La RA en vision artificielle a pour but, donc, d'améliorer notre perception du monde réel par l'ajout d'objets qui ne sont pas à priori perceptibles par l'œil humain. Ainsi, la RA est définie comme un système capable de combiner des scènes réelles avec des objets virtuels [34, 68]. La composition doit être interactive et en temps réel où l'utilisateur perçoit des objets virtuels en même temps que l'environnement réel dans lequel il évolue. La contrainte temps réel [49] rend la RA indispensable pour effectuer une quelconque augmentation, tandis qu'en post production, plusieurs logiciels de traitement d'images permettent de modifier des séquences vidéo réelles. D'où, la nécessité d'introduire le caractère temps réel pour bien justifier l'utilité de la RA.

Dans l'augmentation d'une scène, l'effet réaliste est important. Toutefois, l'interaction de l'utilisateur, en temps réel, avec les objets virtuels est tout aussi importante. Il devrait pouvoir les prendre, les faire déplacer ou glisser et même les détruire en utilisant certains outils. Tout cela doit prendre en compte le recalage correct des objets réels et virtuels, ainsi que leur occultation.

Un bon alignement ('*registration*') des objets réels et virtuels ou recalage, constitue le point central de la recherche [39, 75]. Il consiste à incruster des objets virtuels, sans décalage, par rapport à leurs endroits prévus. Cette incrustation devrait respecter les principes d'occlusion, de zoom, d'éclairage, d'ombre, de rendu, etc. Cependant, sa maîtrise n'est pas simple à cause de la difficulté de contrôle d'un certain nombre de paramètres. Pour ce faire, des traitements supplémentaires sont nécessaires pour les compenser. Il en découle des retards entre le moment de la détermination de la position du point de vue, de la génération des images par le générateur de scènes et celui de leur affichage. La calibration de la caméra et la diversité des flux d'entrée sont à l'origine de ces paramètres.

Il existe deux catégories de paramètres qui causent cette difficulté. La première est celle qui correspond aux limites technologiques des outils utilisés, parfois insurmontables. Ils sont dits statiques. Les déformations optiques, qui induisent des images distordues, en sont un exemple. Elles sont inhérentes aux caméras et leur compensation est coûteuse en temps. Par analogie, la deuxième catégorie de paramètres dépend de la mobilité des objets du monde réel et de l'utilisateur (ou des caméras). Ils sont dits dynamiques. L'impact essentiel de cette difficulté de contrôle est d'engendrer des retards de génération et des défauts d'alignement des images virtuelles. Par conséquent, l'augmentation devient difficile à accomplir en temps réel.

2.1. Principe et problématique de l'augmentation

D'une manière générale, pour augmenter une scène, on doit disposer d'une caméra (ou plus, en stéréovision) aménagée relativement à un repère. Sa calibration consiste à déterminer, géométriquement, ses propriétés optiques ainsi que sa position et son orientation. La scène filmée dispose également de son propre repère, où devront être connues les positions de certains objets réels. Les traitements d'augmentation doivent prendre en compte les temps de latence [47], variables et grands, des outils qui composent le système. Leurs dissemblances (fiabilité, fréquence, nature, etc.) nécessitent des corrections spatiales et temporelles. En plus, l'imprécision inévitable des parties mécaniques est un autre point à considérer. C'est pour ces raisons que d'autres sondes, tels que des télémètres laser ou des capteurs magnétiques et même des

interventions humaines sont parfois nécessaires. Ils servent à fixer ou connaître la position ‘absolue’ de la caméra et des objets réels.

L’insertion des objets virtuels nécessite des indicateurs à identifier dans la scène réelle. Une fois recherchés et trouvés, ils seront utilisés comme repères d’incrustation. Le ‘tracking’ consiste à identifier leurs formes et leurs positions dans chaque image de la séquence relativement à un modèle numérique de chaque pattern utilisé.

L'incrustation des objets virtuels se fait alors selon des principes de projections géométriques. Le calcul de la matrice *H* (homographie planaire) est fait pour chaque image de la séquence. Elle correspond à la solution de l’équation $[x\ y\ w]^t = H\ [x'\ y'\ w]^t$, dans le cas d’une projection plane 2D, qui peut être résolue par la méthode des valeurs singulières. Dans cette équation (x’, y’, w) sont les coordonnées homogènes d’un point X de l’objet virtuel relativement à son repère et (x, y, w) ses coordonnées homogènes dans le repère de l’image. La résolution de cette équation nécessite la connaissance de quatre points dans le repère de l’image.

Ainsi, plusieurs problèmes se posent lorsqu’on tente d’incruster des objets virtuels dans des images réelles :

2.1.1. Alignement des caméras réelle et virtuelle

Le premier problème est de faire correspondre la perspective de l'objet virtuel avec celle de la scène réelle. Ce problème est connu sous le nom d'alignement des caméras réelle et virtuelle. Pour le résoudre, il faut d’abord retrouver les propriétés de la caméra réelle ayant donné lieu à l'observation, ensuite à calculer les images synthétiques en utilisant une caméra virtuelle reprenant ces propriétés. La figure II.1-b illustre un exemple de résultat obtenu lorsque l'image de la voiture respecte la perspective réelle. Par contre, la figure II.1-a montre le résultat lors d’une intégration quelconque de cette image. Le résultat obtenu en figure II.1-b ne suffit pas à obtenir une image réaliste : la voiture n'est pas correctement éclairée et l’arrière du véhicule devrait être occulté par le bâtiment photographié et non pas être projetée par-dessus l'édifice.

2.1.2. Cohérence spatio-temporelle

Les déplacements des objets virtuels dans la scène réelle et les occultations qui peuvent se produire entre objets virtuels et réels constituent le problème de cohérence spatio-temporelle. La figure II.1-c montre un résultat de composition où ce problème est pris en charge.

2.1.3. Cohérence photométrique

Enfin, la prise en compte des inter-réflexions lumineuses (ombres, reflets) entre les objets réels et virtuels est du ressort de la cohérence photométrique. Le résultat obtenu en figure II.1-d tient compte de ce problème.

a- Intégration quelconque

b- Prise en compte
de la perspective réelle

c- Prise en compte
des contraintes spatio-temporel

d- Prise en compte
des contraintes photométriques

Figure II.1 – Augmentations par le véhicule rouge

2.1.4. Discussion

Si ces problèmes sont communs à la post-production et à la réalité augmentée (temps réel), leur résolution sous la contrainte du temps réel impose de mettre en œuvre

des techniques particulières. Cette spécificité est bien sûr liée au fait que les calculs doivent être réalisés très rapidement, ce qui demande de faire preuve d'astuce au niveau logiciel et d'utiliser du matériel performant. Justement, nous allons passer en revue, dans la section suivante, les outils qui nous semblent nécessaires pour réaliser un système de RA.

2.2 Outils de la RA

Actuellement, deux moyens sont utilisés pour la réalisation des augmentations en temps réel. Des HMD (Head Mounted Display) [3, 4] avec des verres transparents (Fig. II.2) ou des HMD totalement opaques (Fig. II.4). Les applications optant pour ces derniers peuvent les remplacer par des écrans d'affichage ordinaires (Fig. II.6). Ces moyens contraignent à l'utilisation de techniques spécifiques, en l'occurrence, les systèmes optiques et les systèmes vidéo respectivement.

2.2.1. HMD optique

Pour les systèmes optiques (Fig. II.3), l'occlusion se fait par des verres transparents, partiellement transmissifs et partiellement réfléchissants. De cette façon, ils permettent à la lumière du monde réel de les traverser tout en visualisant des images projetées par un écran disposé au dessus. Ces images représentent des objets virtuels cumulés à la vue du monde réel.

Figure II.2- HMD optique

Le système optique présente des avantages par rapport au système vidéo pour certains aspects. L'implémentation est simple puisque le monde réel est perçu directement à travers les verres, donc non traité comme un second flux.

2.2.2. HMD vidéo

Pour les systèmes vidéo (Fig. II.5), le casque est totalement opaque et ne permet pas de voir directement le monde réel. Il est constitué d'un écran, sur lequel sont projetés la scène réelle filmée par deux caméras et les images virtuelles. Le principe consiste à combiner les vidéos obtenues des caméras avec celles créées par un générateur de scènes. L'ensemble fournit alors des vidéos augmentées.

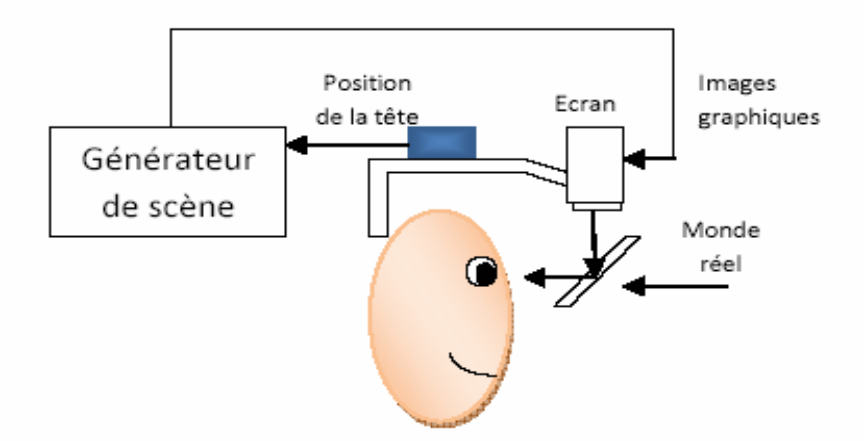


Figure II.3- Schéma HMD optique

Figure II.4 - HMD vidéo

Le système vidéo présente également des avantages par rapport au système optique pour d'autres aspects. Il est, particulièrement, plus flexible dans la composition pour l'occlusion entre objets virtuels et réels.

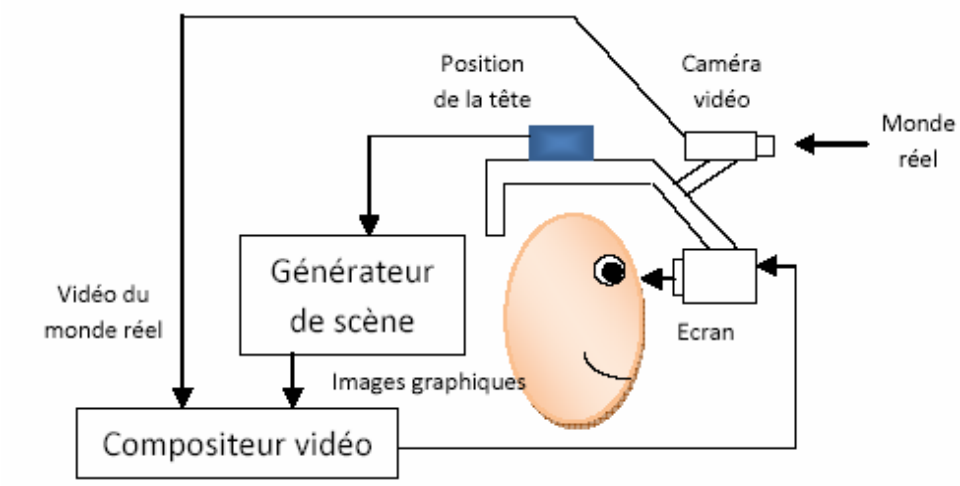


Figure II.5 - Schéma HMD vidéo

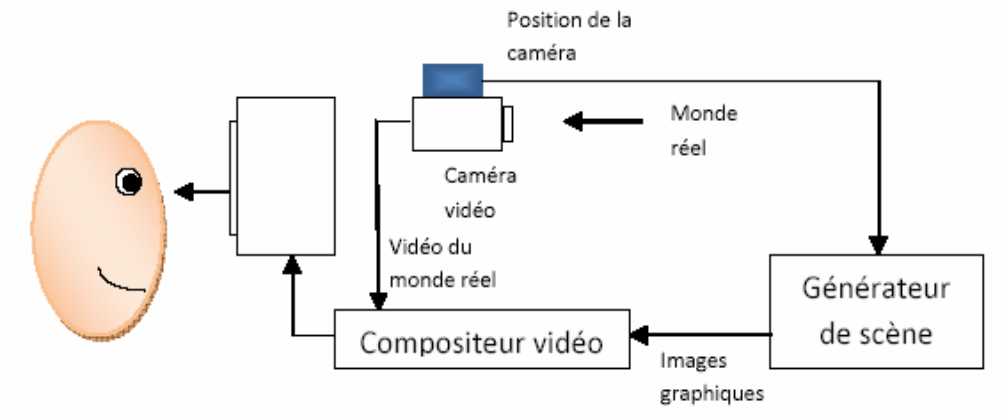


Figure II.6 - Schéma écran ordinaire

2.2.3. Outils annexes

L'augmentation n'est pas restreinte à une vue combinée de réel et de virtuel. Elle requiert également une interaction, particulièrement avec les objets virtuels. En effet, ces objets n'ont d'existence que dans le système informatique. Leur manipulation par l'utilisateur est conditionnée par des gestes que le système doit interpréter. Pour ce faire, des outils de maniement sont nécessaires. Ceux-ci varient en fonction des applications et des besoins. On cite à titre d'exemple, en plus des HMD, les tableaux d'affichage, les plates-formes utilisées essentiellement pour des augmentations collaboratives, les stylos attachés à des capteurs magnétiques, etc.

2.3. Systèmes de RA existants

Après avoir présenté le principe d'une augmentation et les outils nécessaires pour la réaliser, il nous est possible de classer, dans cette section, les systèmes de RA existants par catégorie. Avant cela, nous présentons l'architecture générale d'un tel système.

2.3.1. Architecture générale d'un système de RA

En effet, pour un système de RA de type post-production, trois parties distinctes sont nécessaires avant la composition (Fig. II.7) : la partie synthèse d'images qui consiste à modéliser les objets virtuels à incruster dans la séquence réelle et à les illuminer, la partie calcul des occultations qui consiste à déterminer les objets réels venant occulter un ou plusieurs objets virtuels afin d'assurer la cohérence géométrique de la scène, et enfin, le noyau du système représenté par la partie estimation du mouvement ou calibration de la caméra qui consiste à déterminer les paramètres de la caméra. Ces derniers sont utilisés à la fois pour le calcul des images virtuelles et la détermination des masques d'occultation.

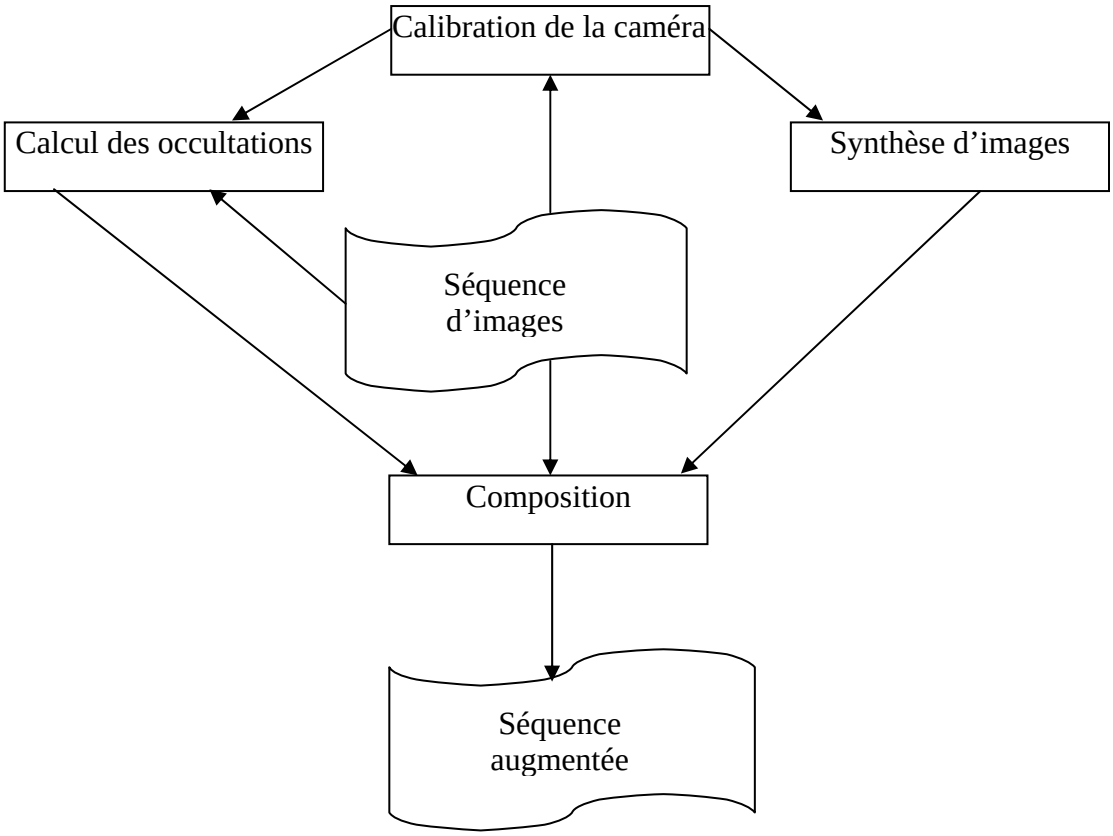


Figure II.7 – Architecture générale d'un système de RA

Par contre, dans une application en temps réel, les occultations ne sont généralement pas traitées et la partie synthèse se résume le plus souvent à la projection du modèle filaire des objets ou de simples annotations.

2.3.2. Calibration de la caméra

Cette étape consiste à extraire les informations nécessaires à l'augmentation à partir des images numérisées générées par la caméra. Cela revient à rechercher des indices ou marqueurs prédéfinis dans la séquence et à les utiliser comme repère d'incrustation des objets virtuels. Pour parvenir à cet objectif, il est important de définir les correspondances géométriques entre le monde réel et les images générées par la caméra. Le modèle de la Figure II.8, dit sténopé, présente l'avantage d'être simple tout en restant fidèle à la réalité.

Les objets sont positionnés dans un repère associé à la scène. L'objectif est de déterminer le point de vue [80] de la caméra (position et orientation) dans ce repère. La correspondance consiste à trouver les différentes transformations à utiliser pour faire passer les coordonnées de l'objet, exprimées dans un repère propre à l'objet ayant comme origine son centre, vers le repère du plan image. 'To' représente la transformation des coordonnées de l'objet de son propre repère vers le repère de la scène réelle. 'Tc' représente la correspondance entre les coordonnées du monde réel et ceux de la caméra. Enfin, 'Ti' représente le passage entre les coordonnées 3D de la caméra et les coordonnées 2D du plan image. L'augmentation nécessite une exactitude dans le calcul de ces trois transformations géométriques. Toute erreur à ce niveau engendre un défaut d'alignement.

La transformation 'Ti' nécessite la connaissance du modèle interne de la caméra (Fig. II.9). Tout point *p* de coordonnées (*x*, *y*, *z*) dans le système de la caméra se projette au point *p'* de coordonnées (*x'*, *y'*, *d*) sur le plan image perpendiculaire au plan $(\vec{i}, \vec{j})$ du système de coordonnées de la caméra. Le système de coordonnées de la caméra est centré sur son *objectif*. La distance *d* qui sépare l'objectif du plan image est appelée focale. La focale est positive pour éviter des raisonnements sur une image inversée. On démontre facilement que :

$$x' = d\frac{x}{z} \text{ et } y' = d\frac{y}{z},$$

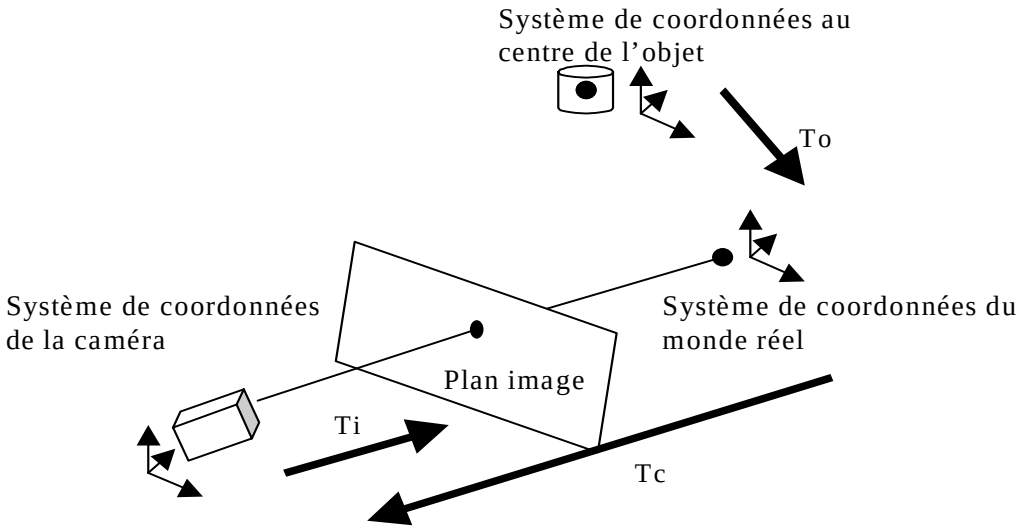


Figure II.8 - Transformations géométriques : modèle sténopé

Ce qui donne sous format matriciel en coordonnées homogènes :

$$\begin{bmatrix} x' \\ y' \\ z' \\ w \end{bmatrix} = \begin{bmatrix} d/z & 0 & 0 & 0 \\ 0 & d/z & 0 & 0 \\ 0 & 0 & d/z & 0 \\ 0 & 0 & 1/z & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Cette forme est générale et ne permet pas de déterminer les différents paramètres de la caméra de manière séparée et qui sont de deux types : intrinsèques et extrinsèques. Les paramètres intrinsèques sont ses caractéristiques géométriques et physiques. Ils représentent la longueur de la focale *d*, la position du centre de l'image (point principal) exprimé en pixels qui représente l'intersection de l'axe optique avec le plan image, les tailles des pixels et le coefficient de distorsion optique. Leur détermination constitue la transformation 'Ti'. Dans la plupart des cas d'augmentation, le coefficient de distorsion peut être ignoré.

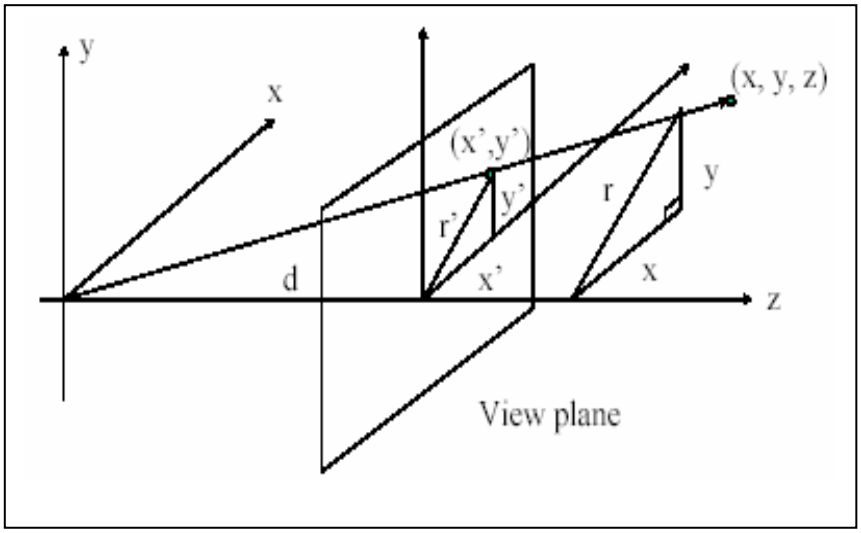


Figure II.9- Modèle de la caméra

Les relations qui lient les coordonnées image (x_{im}, y_{im}) à celles de la caméra sont $x' = -(x_{im} - o_x)s_x$ et $y' = -(y_{im} - o_y)s_y$. (o_x, o_y), désignent les coordonnées en pixels du point principal et (s_x, s_y) les tailles des pixels (en millimètres), relativement aux deux axes. Le signe '–' est utilisé pour exprimer les sens inversés des axes des coordonnées image par rapport aux coordonnées de la caméra.

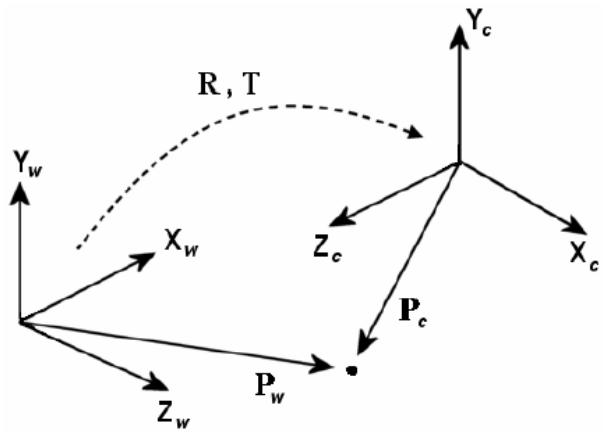


Figure II.10 - Transformation Tc

Les paramètres extrinsèques représentent la position et l'orientation de la caméra relativement aux coordonnées du monde réel (Fig. II.10). La transformation sur chaque point consiste en une rotation *R* et une translation *T*, dite point de vue de la caméra. En notation matricielle, cela donne : $\mathbf{P_c} = R\mathbf{P_w} + T$ ou encore :

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} = R \begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix} + T = \begin{bmatrix} R & T \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Y_W \\ 1 \end{bmatrix}.$$

Ainsi chaque point dans l'image aura un correspondant du monde réel par substitution des équations trouvées précédemment avec cette dernière. Cependant, la détermination simultanée des paramètres intrinsèques et extrinsèques est généralement peu précise. C'est alors que la finalité du calcul est d'avoir deux matrices, l'une représentant les paramètres intrinsèques (*f*, *o* et *s*) et l'autre, les paramètres extrinsèques (*R* et *T*). On obtient [43] :

$$M_{\text{int}} = \begin{bmatrix} f_u & 0 & o_x \\ 0 & f_v & o_y \\ 0 & 0 & 1 \end{bmatrix} \text{ et}$$

$$M_{ext} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix}$$

où $f_u = -d/s_x$, $f_v = -d/s_y$, $t_i = -R_i^T T$ avec R_i, i = 1, 2, 3 représente la *i*ème ligne de la matrice *R*. La projection globale s'exprime de la sorte :

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = M_{\text{int}} M_{ext} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = \begin{bmatrix} f_u r_{11} & f_u r_{12} & f_u r_{13} & f_u t_1 \\ f_v r_{21} & f_u r_{22} & f_v r_{23} & f_v t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix}$$

$$\text{avec } x_1/x_3 = x_{im} \text{ et } x_2/x_3 = y_{im}.$$

La matrice $M = M_{\text{int}} M_{ext}$ est dite matrice de projection perspective et représente le modèle sténopé de la caméra. Les valeurs de ces matrices étant inconnues, l'opération qui consiste à les déterminer est appelée calibration de la caméra. Pour effectuer cette calibration, il est nécessaire d'avoir au moins six points connus dans la scène. Ils

permettent de résoudre le système d'équations résultant contenant les 12 paramètres de la matrice *M*. Deux techniques sont utilisées pour calibrer une caméra : technique basée capteur [5, 81] et technique basée vision [43, 78]. Notons que certains systèmes utilisent une approche hybride. Notre intérêt est porté sur la technique basée vision qui comporte deux méthodes. Une première, dite calibration forte, tente de faire la recherche d'une mire 3D ou d'un pattern 2D [69, 72] ayant des caractéristiques connues dans la scène réelle. Cette recherche peut se faire de manière automatique ou manuelle. La seconde méthode tente d'estimer la matrice de projection perspective par extraction d'indices naturels en utilisant des méthodes statistiques à partir de deux ou plusieurs images. Elle est dite auto calibration ou calibration basée image [79].

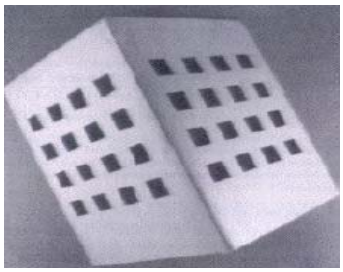


Figure II.11 - Mire 3D de calibration

Figure II.12- Pattern 2D de calibration

Selon que la partie calibration est réalisée d'une manière forte ou faible (autocalibration), on distingue deux catégories de systèmes de RA : les systèmes basés modèle et les systèmes basés image.

2.3.3. Les systèmes basés modèle

Le principe de la calibration dans ces systèmes est d'introduire une mire 3D (Fig. II.11) ou un pattern 2D (Fig. II.12) dans la scène pour calculer les paramètres de la caméra. Pour le cas d'une mire 3D, il suffit de connaître sa géométrie puis de la suivre tout au long de la séquence. Après détection de cette mire, il est nécessaire de déterminer les six points non coplanaires pour résoudre le système d'équations relatif à la matrice de projection perspective. La résolution du système permet d'obtenir les paramètres de la caméra qui seront utilisés par la suite pour l'augmentation. Toute insertion d'objets se fera suivant une projection selon les matrices *R* et *T*.

La calibration est plus difficile dans le cas où on utilise un pattern 2D au lieu d'une mire 3D. Dans un premier temps, il faut déterminer les nouvelles équations correspondant à un objet planaire (Z=0). Cette nouvelle hypothèse conduit à la forme suivante :

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{14} \\ p_{21} & p_{22} & p_{24} \\ p_{31} & p_{32} & p_{34} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix}$$

puisque $Z = 0$, soit

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \lambda . H \begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix}$$

Où λ est un facteur d'échelle et la matrice H est dite Homographie. Pour rechercher les éléments de la matrice H, quatre points sont nécessaires. Ces points permettront de résoudre le système d'équation avec la méthode de décomposition de valeur singulière. Pour déduire cette matrice H, il est nécessaire d'identifier le pattern 2D dans la séquence. Une fois déterminée, la matrice H permettra de faire des projections d'objets 2D dans l'image correspondante sans avoir à calculer les paramètres intrinsèques et extrinsèques de la caméra. Par contre, pour pouvoir augmenter la scène avec des objets 3D, la détermination de ces paramètres reste indispensable. Pour cela, il faut déterminer la relation qui existe entre les paramètres de la matrice H et ceux de M.

Dans un premier temps, il est clair que :

$$H = \begin{bmatrix} f_u r_{11} & f_u r_{12} & f_u t_1 \\ f_v r_{21} & f_v r_{22} & f_v t_2 \\ r_{31} & r_{32} & t_3 \end{bmatrix}.$$

A partir de cette équation, il est possible de déduire les éléments de *M* à partir de ceux de *H*. Pour cela, l'axe *Z* de la matrice de rotation doit être considéré orthogonal aux deux autres. Ainsi, les équations suivantes seront vérifiées :

$$r_{11}^2 + r_{21}^2 + r_{31}^2 = 1,\ r_{12}^2 + r_{22}^2 + r_{32}^2 = 1$$
$$\text{et } r_{11}r_{12} + r_{21}r_{22} + r_{31}r_{32} = 0.$$

Cela permet dans un premier temps de calculer f_u, f_v et λ, puis le reste des éléments de *R* et de *T* par de simples remplacements. Une fois déterminées, ces matrices permettront de faire des augmentations d'objets virtuels 3D positionnés relativement au système de coordonnées du pattern. En d'autres termes, l'homographie peut être utilisée pour 'auto-calibrer' la caméra. Les paramètres intrinsèques et extrinsèques sont calculés à base des éléments h_{ij} de la matrice *H* déduite à partir du pattern 2D. Ils sont respectivement déterminés comme suit :

$$f_u = \sqrt{\frac{h_{11}h_{12}(h_{21}^2 - h_{22}^2) - h_{21}h_{22}(h_{11}^2 - h_{12}^2)}{-h_{31}h_{32}(h_{21}^2 - h_{22}^2) + h_{21}h_{22}(h_{31}^2 - h_{32}^2)}},$$

$$f_v = \sqrt{\frac{h_{11}h_{12}(h_{21}^2 - h_{22}^2) - h_{21}h_{22}(h_{11}^2 - h_{12}^2)}{-h_{31}h_{32}(h_{11}^2 - h_{12}^2) + h_{11}h_{12}(h_{31}^2 - h_{32}^2)}},$$

$$\lambda = \frac{1}{\sqrt{h_{11}^2 / f_u^{\,2} + h_{21}^2 / f_v^{\,2} + h_{31}^2}}$$

$r_{11} = \lambda h_{11} / f_u$ $\quad r_{12} = \lambda h_{12} / f_u$ $\quad r_{13} = r_{21}r_{32} - r_{31}r_{22}$ $\quad t_1 = \lambda h_{13} / f_u$

$r_{21} = \lambda h_{21} / f_v$ $\quad r_{22} = \lambda h_{22} / f_v$ $\quad r_{23} = r_{31}r_{12} - r_{11}r_{32}$ $\quad t_2 = \lambda h_2 / f_v$

$r_{31} = \lambda h_{31}$ $\quad r_{32} = \lambda h_{32}$ $\quad r_{33} = r_{11}r_{22} - r_{21}r_{12}$ $\quad t_3 = \lambda h_{33}$

2.3.4. Les systèmes basés image

Ces systèmes se basent sur des algorithmes permettant de retrouver la structure de la scène en même temps que le mouvement de la caméra, à partir seulement du flot d'images de la séquence sans aucune connaissance à priori, ni sur la scène, ni la plupart du temps sur les paramètres intrinsèques de la caméra.

Toutefois, ces algorithmes sont loin d'être exploitables en temps réel. Ils reposent sur le principe de la géométrie projective (Fig. II.13). Leur principe consiste à calculer une matrice, dite fondamentale, qui impose des contraintes géométriques entre deux vues. Son raisonnement utilise le principe de deux caméras ou deux vues d'une même caméra.

Soit (ΔR Δt) le déplacement relatif (mouvement) de la seconde caméra par rapport à la première caméra, exprimé dans le repère de la seconde. Les trois vecteurs $\overrightarrow{CC'}=\Delta t$, $\overrightarrow{CM}$ et $\overrightarrow{C'M}$ étant coplanaires, il est possible d'écrire la relation : $M_{C'}^{T}\Delta t\Delta R M_C = 0$ où M_C est le vecteur $\overrightarrow{CM}$ exprimé dans le repère de la première caméra et $M_{C'}$ le vecteur $\overrightarrow{C'M}$ exprimé dans le repère de la seconde caméra. Ainsi, $\overrightarrow{CM}$ a pour coordonnées $\Delta R\, M_C$ dans le repère de la seconde caméra.

Si q et q' sont les coordonnées pixels respectives des points m et m', alors l'équation de projection perspective permet de donner : $q = M_{\text{int}} M_C$ et $q' = M'_{\text{int}} M'_{C'}$. Ces relations permettent de déduire l'équation de Longuet-Higgins :

$$q'^{T} F q = 0 \text{ où } F = \left(M'^{-1}_{\text{int}}\right)^{T} \Delta t \Delta R M^{-1}_{\text{int}}$$

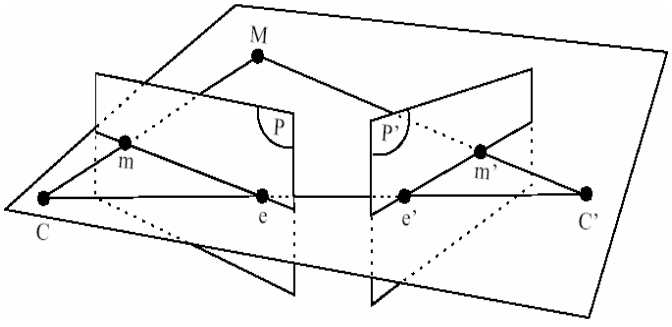


Figure II.13 - Géométrie épipolaire

L'équation citée ci-dessus représente la matrice fondamentale. Dans ce cas, huit appariements de points permettent de déterminer une solution unique à F. Mais comme généralement on dispose d'un nombre d'appariements plus grand, l'équation est résolue aux moindres carrées en minimisant le critère linéaire :

$$\min_F \sum_i (q_i'^T F q_i)^2 \ .$$

Comme ce critère est sensible au bruit, il est préférable d'utiliser un autre critère qui opère simultanément sur les deux images:

$$\min_F \sum_i (d^2(q_i', Fq_i) + d^2(q_i, F^T q_i'))$$

On dit qu'un couple de caméras est faiblement calibré si la matrice fondamentale est connue. On peut alors accéder à une reconstruction tridimensionnelle de la scène par une reconstruction homographique près. Ceci est possible sous condition de déterminer les paramètres intrinsèques de la caméra. Ils sont déterminés à partir de la matrice fondamentale qui impose deux contraintes polynomiales, dites équations de Krupa [79]. Ainsi, pour déterminer quatre paramètres, deux mouvements sont nécessaires. Cependant, il existe des mouvements pour lesquels les équations citées dégénèrent. Citons les translations pures, les rotations pures de faibles amplitudes, les mouvements plans, etc. l'auto calibrage est donc une méthode instable dans la pratique.

2.4. Domaines d'application

Parmi les domaines investis par les recherches en RA [4], on peut recenser les applications touchant : à la manipulation d'objets virtuels [70, 71, 74], à l'annonce informationnelle, aux systèmes collaboratifs, à la reconstruction visuelle de faits ou de structures historiques, etc.

La médecine est l'un des domaines les plus investis. L'idée principale consiste à produire, en temps réel, un ensemble de données sous forme d'images 3-D, représentant une partie invisible du patient, sans avoir à faire d'incision (Fig. II.14). Cette

construction utilise des données disponibles sous d'autres formes, telles que des images IRM, GT et ultrason, qu'il serait difficile à interpréter directement.

L'assistance pour la fabrication, la maintenance (Fig. II.15) ou la réparation est un autre domaine où la réalité augmentée est appréciée. Elle consiste à fournir des animations qui peuvent être superposées aux équipements, montrant leur manipulation. L'objectif, dans ce cadre, étant celui de remplacer les manuels (documentation). Et comme complémentaires à cette idée, l'annotation et la visualisation (Fig. II.16) sont utilisées. Elles consistent respectivement à ajouter une description textuelle ou graphique à des objets réels et à parfaire des images embrouillées obtenues durant des conditions de vues difficiles. Signalons, au passage, que l'augmentation nécessite la disponibilité de modèles des objets ou de l'environnement.

Figure II.14- Fœtus virtuel projeté sur le ventre d'une femme enceinte.

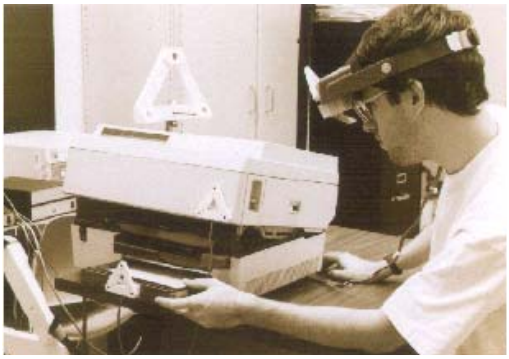


Figure II.15- Une application de RA pour la maintenance d'un photocopieur

La télé-opération est également ciblée, particulièrement pour le planning du déplacement d'un robot. Elle permet de manipuler une version virtuelle d'une opération

tout en enregistrant les actions. Une fois celles-ci testées et déterminées, le robot pourra exécuter le plan correspondant.

Dans le domaine des médias visuels [44], l'augmentation consiste à intégrer des objets, en temps réel et en 3D, dans des scènes après qu'elles soient filmées. L'occlusion et la disposition des objets sont les paramètres les plus importants à contrôler. Cette opération est particulièrement importante pour les objets virtuels ou ceux réels et dont l'intégration à une scène particulière est difficile ou coûteuse. Les publicités réalisées durant les matchs de football en sont un bon exemple.

Figure II.16- Une application de RA pour l'annotation dans un circuit de course de voiture.

Le domaine militaire reste celui le plus investi par la recherche et particulièrement l'aviation. Des informations supplémentaires sont ajoutées à celles déjà existantes sur le casque correspondant au monde réel. Elles permettent au pilote, par exemple, de viser et voir les directions des armes de son avion relativement aux cibles. L'exploration de terrains ennemis sans reconnaissance préliminaire est un autre cas d'usage.

2.5. Synthèse

Le domaine de la réalité augmentée est assez vaste et regroupe d'autres recherches qui sont considérées comme domaines en soit. C'est ce qui fait que plusieurs de ses résultats dépendent de ceux obtenus dans d'autres domaines, en y apportant des adaptations. Il existe même des axes repris de manière adaptative.

Afin que la recherche progresse dans ce domaine, plusieurs pivots sont traités. Pour satisfaire les augmentations en temps réel, les recherches sont axées sur la minimisation du temps de tracking et des retards de numérisation et de génération des objets virtuels ainsi que l'exploitation des systèmes temps réels. Pour simplifier les systèmes de réalité augmentée, des recherches dans le sens de la réduction des contraintes de calibration des caméras sont menées. En plus, la structure des interfaces nécessaires pour l'interaction, pouvant supporter ces nouveaux systèmes, est un facteur important de réussite. Rajoutées à celles-ci, des études psychophysiques sont nécessaires. Elles pourront étudier le niveau de détection des erreurs et de perception humaine, la tolérance aux erreurs d'alignement par les différentes applications, les effets des HMD après enlèvement, etc.

Aussi, les recherches sont de plus en plus orientées vers la conception des systèmes de RA pour des environnements non préparés, qui sont par nature délicats. Nous détaillerons ce point précis dans le prochain chapitre. C'est également le cas pour les environnements multi utilisateurs pour des applications de systèmes collaboratifs [40, 67, 73]. Ils permettent à différents utilisateurs d'interagir simultanément avec des objets réels et virtuels communs. Les mécanismes de communication nécessaires pour la coordination entre les participants sont également un aspect important. Cet aspect fait l'objet d'une autre thèse menée par un collègue au sein de notre équipe et dirigée par le professeur Batouche. Enfin, les effets photo réalistes des objets virtuels, telles que l'illumination et la réflectivité ainsi que le problème de la densité des données où les objets virtuels ne doivent pas dissimuler des objets réels ou des informations qui soient importantes, rentrent dans le cadre du réalisme de la scène.

Notons qu'il reste toujours certaines limites insurmontables. A titre d'exemple, si l'on considère une rotation de la tête (ou de la caméra) de 50° en 1 seconde, cela

exprime le fait que le système doit générer des images virtuelles avec un retard de 10ms ou moins afin de satisfaire une erreur de 0,5°. Cela exprime la génération d'une image chaque 10ms. Dans le cours des choses, juste l'affichage d'une image sur un écran à 60 Hz nécessite 16.67 ms. En d'autres termes, il n'est pas possible de satisfaire une augmentation qui atteigne une précision de 0,5°.

2.6. Conclusion

L'un des problèmes encore posés, en réalité augmentée, est la non fiabilité des méthodes de suivi (tracking). Elles sont toujours très gourmandes en temps de calcul et beaucoup de tentatives sont faites pour les améliorer. Ainsi, lorsqu'une simple rotation est enregistrée dans la scène, il est possible qu'une partie du contenu précédant de l'image virtuelle, déjà générée, soit utilisée pour la suivante. D'une manière générale, la prédiction permet de faire un traitement à priori prédictif, sans mesurer les valeurs réelles. Nous Consacrerons le chapitre 4, pour présenter les différents travaux réalisés jusqu'ici dans le suivi d'objets dans une séquence d'images vidéo.

CHAPITRE 3

Réalité augmentée pour les environnements non préparés.

L’idéal pour un système de RA est de fonctionner sans connaissances préalables sur la structure de la scène filmée et du mouvement de la caméra. Sachant que ces conditions ne sont pas réunies actuellement, tous les systèmes de RA existants demandent un degré de connaissance plus ou moins important sur la scène et les paramètres de la caméra.

La plupart des systèmes de RA disponibles nécessitent le positionnement de marqueurs facilement identifiables dans la scène. Ces systèmes sont généralement utilisés dans des environnements restreints à cause de l’ajout de ces marqueurs.

D’autres systèmes de RA, moins contraignants que les précédents, doivent disposer d'une base de données d'images de l'objet recalé, photographié sous plusieurs angles et plusieurs illuminations, ou encore doivent disposer d'un modèle complet de cet objet sous forme de facettes. Ces systèmes qui s'appuient sur la structure naturelle de la scène et non sur le positionnement de marqueurs artificiels peuvent être envisageable

pour fonctionner dans des scènes d'extérieur. On parle alors de réalité augmentée pour les environnements non préparés (Markerless augmented reality).

3.1 Evolution de la RA

C’est en 1979, qu’apparaît un des tous premiers travaux. Il consiste à faire une incrustation d'une image virtuelle dans une photographie après avoir été numérisée à partir d'un scanner à main. L'image composée est sortie sur un écran 320 X 240 ou une imprimante à aiguille.

Uno et al [82] réussissent à plaquer des textures réelles sur des objets virtuels, et même à incruster un immeuble virtuel dans une image réelle. Six ans plus tard, Maver et al [82] proposent un logiciel permettant d'évaluer l'impact d'un bâtiment dans un paysage par photomontage. Les paramètres de la caméra sont entrés par l'utilisateur.

En 1986, Nakamae et al [82] tentent la première composition automatique. Pour cela, Ils explorent quelques points clé de la RA, tels que: calcul des ombres projetées par les sources lumineuses réelles sur les objets virtuels, simulation de phénomènes atmosphériques, etc. Les paramètres de la caméra sont déterminés grâce à des appariements 3-D/2-D de points, mais ces points sont appariés à la main, et l'aspect séquence n'est pas du tout évoqué.

Les premiers systèmes manipulant des séquences d'images au lieu d'images isolées sont apparus au début des années 90. Ainsi, le point de vue est calculé à partir d'un objet de la scène dont le modèle 3-D est connu et qui est suivi dans la séquence, d’où le concept de recalage temporel. Plusieurs systèmes de RA basés modèle ont alors vu le jour.

L’étude du problème du calcul du point de vue à partir de données images a commencé au début des années 90 et l’aspect multi-images n'est apparu qu'à partir de 1996. Suite à cela, les systèmes basés images ont vu le jour. L'application de ces travaux à la RA est quant à elle relativement récente [94].

La plupart des recherches en RA s'orientent, de nos jours, de plus en plus vers les environnements non préparés [2, 6, 84]. Ainsi, les chercheurs visent à concevoir des systèmes pouvant évoluer dans un environnement libre plutôt que des systèmes à usage limité généralement ne dépassant pas la surface d'une petite pièce.

3.2 Classification des différentes approches

La connaissance de la géométrie de la scène est nécessaire pour y incruster des objets virtuels. La meilleure façon de déterminer cette géométrie, pour des exécutions temps réel, est de préparer cette scène au préalable en lui ajoutant des marqueurs facilement identifiables. Malheureusement, ceci n'est possible que pour des scènes d'intérieur (Indoor). Pour permettre une utilisation dans des environnements non préparés ou d'extérieur (Outdoor), plusieurs approches ont été proposées. Nous pouvons les regrouper en deux grandes classes : les approches qui utilisent exclusivement des techniques de vision artificielle et les méthodes hybrides qui en plus des techniques utilisées par les approches de la première classe font appel à d'autres technologies de capteurs (mécaniques, magnétiques, optiques etc.).

3.2.1 Approches basées sur les techniques de vision artificielle

Nous avons recensé un certain nombre d'approches appartenant à cette classe et nous les avons regroupé en trois sous classes.

3.2.1.1 Approches basées modèle

Le calcul du point de vue de la caméra se résout de façon séquentielle, et est connu sous le nom de recalage temporel, lorsque des primitives 3-D de la scène sont connues. Une étape préliminaire est, en général, utilisée pour calculer les paramètres intrinsèques de la caméra en utilisant les appariements connus dans la première image. Le problème se résume donc à calculer le point de vue de la caméra image après image, en utilisant un appariement 3-D/2-D.

Pour qu'un tel système soit autonome et précis, il faut maintenir au cours de la séquence un nombre suffisant de primitives images conformes aux primitives 3-D du modèle. Les erreurs de suivi peuvent en général être détectées par comparaison avec la re-projection du modèle, à condition que l'algorithme de calcul du point de vue soit

robuste face aux erreurs d'appariement. Le problème du maintien d'un nombre suffisant de primitives au cours de la séquence n'est pas simple puisque certaines primitives peuvent être mal suivies, ou encore disparaître du champ de vision de la caméra. Un système autonome doit donc être capable de mettre à jour l'ensemble des primitives suivies. En d'autres termes, il permet :

- De retrouver les primitives mal suivies dans l'image.
- D'intégrer les primitives 3-D entrant dans le champ de vision de la caméra au cours de la séquence (pour compenser la perte des primitives sorties) en recherchant leur homologue 2-D dans l'image (initialisation d'une primitive).

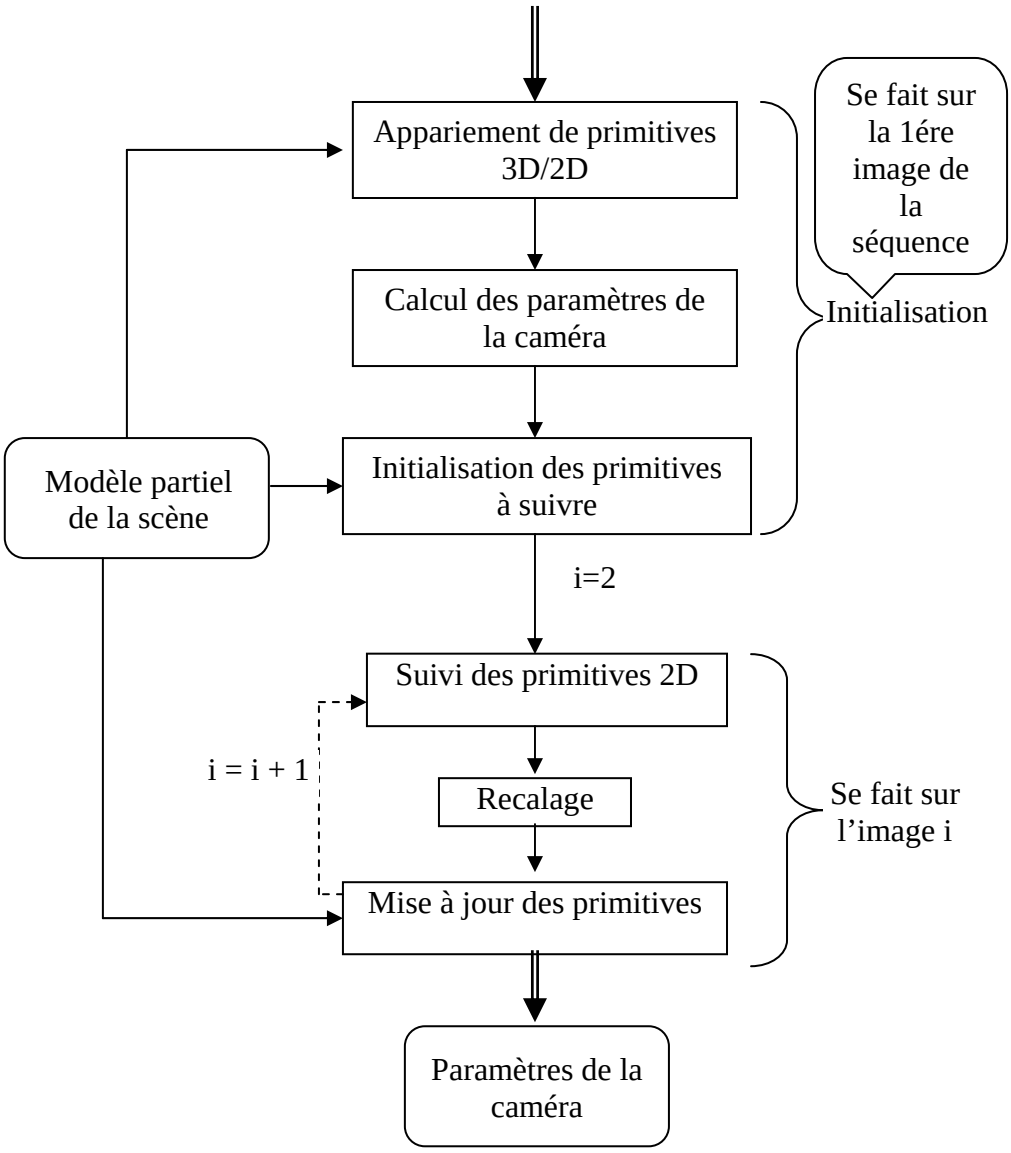


Figure III.1- La boucle de recalage temporel pour un système de RA basé modèle

La figure précédente (Fig. III.1) donne une vue globale sur la boucle de recalage temporel d'un système basé modèle. Les systèmes de RA, que nous allons présenter maintenant, implémentent plus ou moins partiellement cette boucle de recalage. Aussi, ils sont basés sur des primitives de types différents. Certains d'entre eux utilisent le signal intensité des images. D'autres se basent directement sur des points naturels de la scène. Le système de Gagalowitcz [82] prend en compte les facettes texturées du modèle qui sont suivies d'une image à l'autre de la séquence. D'autres algorithmes exploitent les contours de l'image au lieu du signal intensité. Les contours sont considérés comme des chaînes de points présentant un gradient fort du signal.

a- Prise en compte de points naturels

Les primitives utilisées dans ces systèmes sont des points de la scène identifiés en mesurant les niveaux de gris des points voisins. Nous prenons comme exemple, le système de recalage basé sur un suivi de points par corrélation proposé par Uenohara [82], qui fonctionne en temps réel et intègre toutes les étapes de la boucle présentée en Fig. III.1. Initialement, le système affiche le modèle filaire de l'objet à recaler dans la première image. L'opérateur doit déplacer, alors, la caméra par l'intermédiaire d'une interface graphique, jusqu'à ce que l'objet virtuel s'aligne sur l'objet réel. Quelques points-clé sont alors projetés dans l'image. Une recherche de leur correspondant 2-D est faite par la suite de ces projections. Pour améliorer la robustesse du recalage initial face aux changements d'illumination, ces systèmes font appel à des images de référence qui ont été capturées auparavant autour de chaque point-clé et sous diverses conditions d'illumination. Pour chaque point situé dans la zone de recherche, un score de corrélation normalisée est calculé avec toutes ces images de référence. Ainsi, le point correspondant 2-D recherché est celui qui à obtenu le meilleur score à condition que celui-ci soit supérieur à un certain seuil.

Avant l'exécution, une sélection d'un certain nombre de points faciles à suivre est faite. Ces points sont projetés dans l'image dés que la position initiale de la caméra est calculée, puis une fenêtre de corrélation est extraite autour de chaque point. Cette fenêtre sera utilisée comme référence pour le suivi dans l'image suivante. Le suivi est alors effectué par corrélation normalisée. Un invariant géométrique (basé sur cinq points coplanaires) est utilisé pour détecter les éventuels problèmes de suivi. Le point de vue

dans l’image courante est obtenu itérativement en appliquant l'algorithme de Newton à partir de la position obtenue dans l'image précédente.

Il est possible aussi d’utiliser l'invariant géométrique dans le cas où on veut effectuer une composition sur un objet plan dont le modèle 3-D n'est pas connu. En effet, lorsque cinq points coplanaires sont détectés dans une image, il suffit de suivre quatre parmi ces cinq points puisque la position du cinquième point peut être déduite. Cette propriété a été appliquée dans le domaine médical pour maintenir une épingle virtuelle sur un endroit particulier de la jambe d'un patient. Le médecin désigne la pointe de l'épingle dans une image. Sa position 2-D est ensuite calculée par rapport à quatre points d'intérêt de l'image. Ces quatre points sont suivis par corrélation et la pointe de l'épingle peut être repositionnée au bon endroit dés que la jambe du patient bouge.

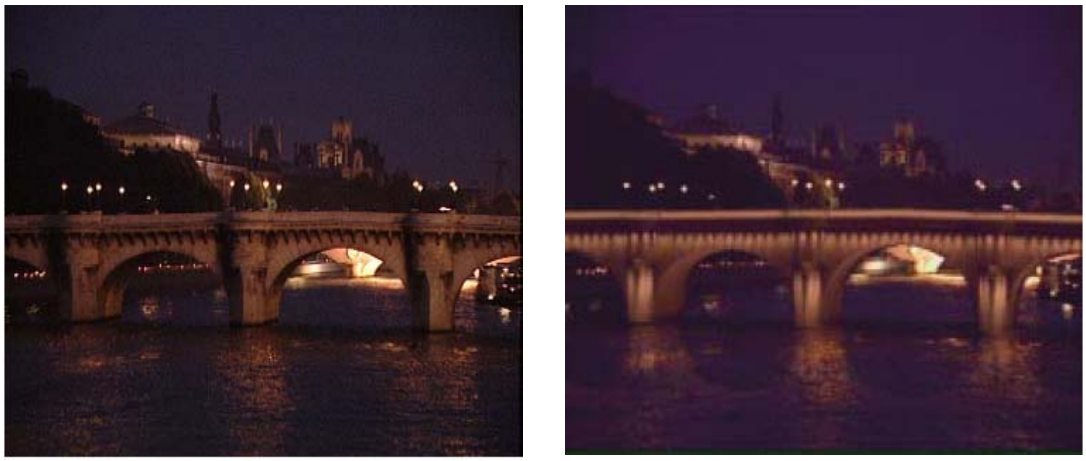


Figure III.2 – Illumination du pont neuf de Paris. A gauche image réelle et à droite image augmentée.

Le système qu’a proposé Ravela [82] est assez semblable à celui de Uenohara et Kanade [82]. Il est appliqué à la maintenance interactive. En effet, ce système permet à un technicien de réparation de regarder à travers un viseur des annotations virtuelles superposées sur l’objet à réparer. Ces annotations désignent certains éléments de l'objet en fonction des opérations à effectuer.

L’initialisation de la boucle de recalage temporel est faite à partir d'un ensemble de correspondances modèle-image de points, des paramètres intrinsèques de la caméra et d'une table d'aspect pré-compilée. Cette table associe des points de vue discrets avec

les primitives visibles depuis ces vues. L'opérateur spécifie les correspondances initiales. Ces correspondances sont alors utilisées pour estimer le point de vue dans la première image. La boucle temporelle, dans ce cas, se déroule en trois étapes :

1. L'information de point de vue est utilisée comme index dans la table d'aspect et une liste des primitives visibles est extraite,

2. les coordonnées 3-D de ces primitives sont projetées dans l'image suivante comme hypothèse de localisation du point image,

3. les fenêtres de corrélation sont localisées dans cette image (par corrélation normalisée avec les fenêtres de corrélation stockées dans la table d'aspect) et un nouveau point de vue est calculé.

Un algorithme itératif robuste est utilisé pour calculer les points de vue à partir de la position dans l'image précédente.

Les travaux de Gilles Simon [81, 82] rentrent dans la même catégorie et consistent à illuminer les ponts de Paris (Fig. III-2)

b- Utilisation des facettes texturées du modèle

Un système de recalage temporel basé sur les textures de l'objet 3-D a été proposé par Gagalowitcz [82]. Ce système exige la connaissance du modèle complet de l'objet sous forme de facettes. Pour calculer, les paramètres intrinsèques et extrinsèques de la caméra, il est nécessaire de faire des appariements 3-D/2-D de points définis par l'opérateur dans la première image. Des projections de polygones correspondant aux facettes de l'objet sont alors réalisées dans le plan image. A l'intérieur de chaque polygone projeté, les textures des facettes sont extraites à partir de l'image. Le modèle est recalé dans l'image suivante par une minimisation itérative de l'erreur entre les textures projetées et les textures observées. Les textures peuvent être réapprises toutes les p images et une approche multi-résolution est utilisée pour réduire les temps de calcul.

Le suivi de facettes texturées est sensible aux occultations et à la présence d'ombres portées. Lorsque les occultations ne couvrent pas une trop grande partie de l'objet, le suivi de facettes peut devenir robuste. Mais, ce type de suivi échoue totalement s'il y a une présence d'ombres portées.

Si une facette occultée par une partie de l'objet devient visible, elle n'est pas nécessairement prise en compte vu qu'elle est cachée relativement au point de vue de l'image précédente. Cette nouvelle facette ne peut donc pas être apprise. Pour pallier à ce problème, l'auteur propose l'utilisation d'un modèle de mouvement, qui restreint les applications possibles [84], ou le redémarrage du processus dans l'ordre inverse de la séquence, à partir d'une image postérieure à l'apparition de la facette, ce qui rend l'algorithme non linéaire.

c- Systèmes basés sur les contours de l'image

Des détecteurs de segments [54, 58] sont utilisés dans la plupart des algorithmes basés sur les contours de l'image. Une fois ces contours déterminés, ils seront appariés avec des primitives du modèle.

Prenant comme exemple, la méthode de recalage d'objets 3-D, dans un contexte d'assemblage de structures en orbite ou de réparation de satellites, proposée par Gennery [36]. Elle est basée sur les segments extraits de la carte des contours. A partir de la position des segments de l'objet 3-D dans l'image i-1, un filtre équivalent à celui de Kalman est utilisé pour prédire la position des ces segments dans une nouvelle image i. Pour cela, l'auteur considère un modèle de mouvement de la caméra à accélération aléatoire (bruit blanc). La valeur attendue de l'accélération, entre les images i - 1 et i, est nulle puisqu'elle est aléatoire. La vitesse prédite est donc celle obtenue entre les images i - 2 et i - 1. Les segments du modèle 3-D sont projetés à la position prédite de la caméra en i. Cette position a été obtenue à partir de sa position en i - 1 et de la vitesse prédite. Le segment le plus proche de la projection est choisi comme correspondant 2-D. Sa direction peut aussi être prise comme critère plus fin de sélection. De plus, un poids est associé à chaque segment en fonction de la précision supposée de sa détection et de sa potentialité à être détecté comme contour fort de l'image. Le point de vue ainsi que la vitesse de la caméra sont alors ajustés par des moindres carrés pondérés. Des résultats

précis ont été obtenus sur un prisme hexagonal, tournant autour d'un axe à vitesse constante. Le mouvement de la caméra doit vérifier l'hypothèse du filtre de Kalman au cas où ce filtre est utilisé. Ceci est approprié dans un cadre très contraint, où il est possible d'imposer des limites précises sur les changements d'accélération attendus. En plus, cette méthode n'est envisageable que si l'incertitude obtenue sur les primitives peut être estimée.

3.2.1.2 Approches utilisant une base d'images de référence

Au lieu d'utiliser un modèle 3D de la scène, Stricker [84] utilise une base d'images de référence de l'environnement. Cette méthode commence par une comparaison entre la vue de l'utilisateur c'est-à-dire l'image courante de la séquence vidéo et toutes les images de référence (Fig. III.3). Pour chaque comparaison un coefficient de corrélation est alors calculé. L'image de référence qui aura le meilleur score sera retenue pour évaluer la transformation 2D qui existe entre elle et l'image courante de la vidéo. L'objet virtuel pourra alors être inséré en utilisant la transformation 2D calculée précédemment. Le succès de cette méthode est fonction du choix de la méthode de comparaison entre images.

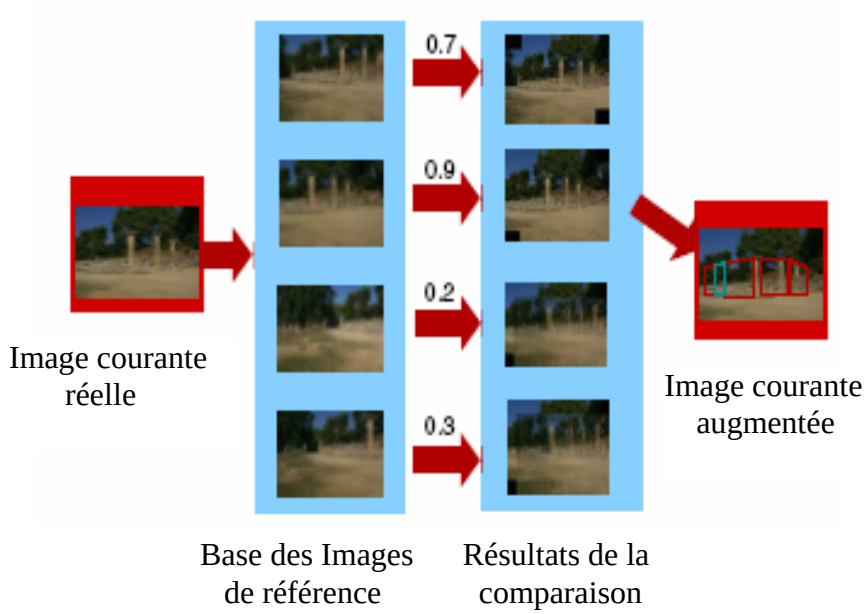


Figure III.3 – Processus de comparaison entre l'image courante et la base des images de référence

3.2.1.3 Approches exigeant une forte interaction

Ces approches n'utilisent aucun modèle de la scène. Pour cela, l'utilisateur commence par désigner des points qui seront suivis tout le long de la séquence. Il intervient ensuite à différents niveaux du processus pour corriger les éventuelles erreurs de suivi et prendre en compte, si nécessaire, de nouvelles primitives. Cette forte interaction de l'utilisateur fait que ces approches ne peuvent pas être appliqués dans un contexte temps réel.

3.2.2 Approches Hybrides

Les techniques de vision artificielle sont les seules habilitées à faire un bon suivi dans des séquences vidéo. Cependant, ces techniques souffrent du manque de robustesse et s'exécutent souvent avec des temps de calcul excessifs. Pour pallier à ces insuffisances, les approches hybrides proposent d'utiliser, en plus des techniques de vision artificielle, d'autres technologies sensorielles. Parmi ces technologies, nous citons à titre d'exemples : les capteurs optiques et les capteurs magnétiques.

3.3 RA et Mobilité

Azuma [2] souligne la mobilité inhérente de la réalité augmentée. Concevoir des systèmes de RA fonctionnant en extérieur est une étape naturelle du développement de la RA pour atteindre l'objectif final fixé : un fonctionnement de partout, quel que soit l'environnement. Ainsi la RA s'est orientée très tôt vers la mobilité et de nombreux systèmes sont déjà mobiles [35] ou le sont dans l'esprit. S'appuyant sur les dispositifs d'affichage portés sur la tête (*Head-Mounted Display*), sur la technologie sans fil et des créations maison comme l'intégration d'un PC à une veste faite sur-mesure, la RA a été et est encore à la pointe de la mobilité. Ainsi la RA sur supports mobiles (Fig. III.4) est souvent liée à l'informatique vestimentaire (*wearable computing*) et portable (*handheld computing*).

Il est néanmoins important de noter que tous les systèmes de RA ne sont pas mobiles. Aussi la mobilité est une caractéristique de classification des systèmes de RA :

- Les systèmes non mobiles : Ce sont les systèmes de RA où l'objet de la tâche est statique, de taille raisonnable et localisé en un seul lieu. Par exemple les

systèmes de GMCAO (Gestes Médicaux et Chirurgicaux Assistés par Ordinateur) conçus pour fonctionner dans le bloc opératoire.

- Les systèmes mobiles : Ce sont les systèmes qui visent à augmenter l'environnement dans lequel évolue l'utilisateur. Les objets de la tâche ne sont pas alors co-localisés. C'est par exemple le cas du campus augmenté ou du musée augmenté.

Figure III.4 – Un système de RA portable

La mobilité de l'utilisateur fait l'hypothèse que l'utilisateur et l'objet de la tâche sont co-localisés au moment de l'accomplissement de la tâche. L'utilisateur se déplace pour pouvoir réaliser la tâche. La co-localisation de l'utilisateur et de l'objet de la tâche définit une situation d'interaction de la RA. Il existe trois situations distinctes d'interaction:

- Téléprésence : L'utilisateur manipule l'objet de la tâche qui est distant grâce à son extension informatique. C'est par exemple le cas d'un bras manipulateur extension du bras de l'utilisateur qui manipule un objet à distance.

- Immersion : L'utilisateur manipule l'objet de la tâche qui est distant en effectuant des opérations sur une représentation de l'objet distant. Le terme Immersion se justifie par le fait qu'un objet distant est ramené dans l'environnement de l'utilisateur, créant ainsi un nouveau contexte d'interaction.

- Télépresence et immersion : Cette situation d'interaction est la fusion des deux précédentes. La télécommande d'une sonde spatiale sur Mars en est un bon exemple. Il convient aussi de s'interroger sur la distance entre les utilisateurs d'un système collaboratif de RA. La téléprésence est d'autant plus convaincante lorsqu'à distance autour de l'objet de la tâche nous avons d'autres utilisateurs. C'est le cas d'un groupe d'utilisateurs co-localisés avec l'objet de la tâche qui font appel à un expert distant. Une autre situation radicalement différente consiste à faire collaborer un groupe d'utilisateurs par le biais d'un objet unique partagé. Par exemple des utilisateurs collaborent à distance autour d'un tableau blanc. Chaque utilisateur ou groupe d'utilisateurs dispose d'un tableau (rond blanc). Une opération (par exemple un dessin) sur un tableau est automatiquement répercutée sur les autres tableaux. Ainsi les utilisateurs ont l'impression de travailler ensemble autour d'un tableau blanc commun. C'est un cas de téléprésence où l'objet de la tâche est partagé.

3.4 Conclusion

Les limites des systèmes de RA pour des environnements restreints, généralement ne dépassant pas une salle de quelques mètres carrés, n'ont pas permis une large utilisation de cette technique. En effet, pour permettre une meilleure évolution de la RA, il faut concevoir des systèmes destinés à une utilisation dans des environnements non préparés. Nous avons présenté dans ce chapitre les approches possibles permettant de construire de tels systèmes.

Afin d'intégrer notre approche, présentée au chapitre 6, dans un système de RA pour des environnements non préparés, nous l'avons doté d'une étape d'initialisation. Cette étape permet de sélectionner un objet pertinent faisant partie de la scène filmée. La sélection est faite sur la première image de la séquence vidéo. L'objet en question

sera suivi durant le reste des images de la séquence et servira comme repère pour faire d'éventuelles augmentations.

CHAPITRE 4

Suivi d'objets dans une séquence d'images vidéo

Le suivi d'objets dans une séquence d'images vidéo est un problème qui demande une extraction et un traitement d'informations provenant d'images complexes et incertaines dans beaucoup de cas. Ce problème devient de plus en plus difficile si la contrainte temps réel est exigée. Dans ce cas, en plus de la complexité et l'incertitude des informations à extraire, le temps de traitement doit être très court, pas plus de 40 ms.

Ce sujet a fait couler beaucoup d'encre et a suscité l'intérêt de beaucoup de chercheurs à travers le monde. Ainsi, un grand nombre de méthodes de suivi [23, 25, 46, 56, 57, 60, 61, 62, 63, 76, 77] existe de nos jours. Ce nombre est dû, d'une part au nombre important de problèmes à régler et d'autre part à la diversité des types d'applications concernées par le suivi. Ainsi, chacune des méthodes peut traiter certains aspects et échoue sur d'autres.

En effet, pour être capable de suivre un objet, il est nécessaire de connaître la relation entre un modèle de l'objet et son image courante. La mise en correspondance

spatio-temporelle se fera selon la nature des primitives visuelles retenues soit en analysant directement les caractéristiques du signal lumineux, soit en estimant une transformation qui permet de recaler le modèle de l'objet sur les données de l'image. Justement, nous allons discuter de la nature des primitives visuelles pour le suivi.

4.1. Primitive visuelle

Une primitive visuelle est une représentation d'une caractéristique particulière d'une image. Elle peut être : soit géométrique décrivant une partie de la structure de l'image (point, segment, cercle, etc.), soit liée directement à la texture de l'objet dans l'image (niveau de gris, couleurs, etc.). Beaucoup de primitives visuelles ont été utilisées en littérature. Nous pouvons citer :

- Black et Jepson [15] ont représenté l'objet comme un ensemble de pixels à suivre en estimant le flot optique.

- Les travaux qui s'intéressent au motif d'un objet. Dans ce cas, plusieurs représentations de la texture ont été utilisées. Par niveau de gris [41, 50, 51], par histogramme de couleur [24], etc.

- Pour les primitives géométriques, Lowe [59] a utilisé les points d'intérêts, Boukir et al [16] ont utilisé les droites, Vincze [88] a utilisé les ellipses, etc.

D'une manière générale, les primitives géométriques résistent bien au changement d'illumination mais à la présence d'ombres, aux occultations et aux motifs texturés. Elles sont donc bien adaptées aux scènes d'intérieur ou aux scènes dont le contenu est peu texturé. Par contre, les méthodes basées texture sont plus aptes à prendre en charges ces problèmes.

Maintenant, une fois notre primitive visuelle choisie, le suivi est défini comme un problème de recalage entre les données extraites des images courantes et le modèle de l'objet. La question qui se pose est quant aux transformations possibles à effectuer pour faire ce recalage. On en parlera dans la section suivante.

4.2. Transformation

Le recalage repose sur une transformation qui permet de faire le lien entre un modèle 2D ou 3D de l'objet et son déplacement et les données 2D. Le suivi revient donc à estimer la transformation qui minimise au mieux le critère de recalage. On distingue deux types de transformations :

- Les transformations 2D qui permettent d'effectuer un transfert des primitives visuelles d'une image à une autre. On parle dans ce cas d'un suivi 2D.
- Les transformations 3D qui permettent de projeter des primitives de l'espace tridimensionnel associées à la scène sur le plan image. On parle alors de suivi 3D.

Le point suivant sera consacré à une classification des méthodes de suivi selon les deux critères présentés en sections 1 et 2, à savoir : primitive visuelle et transformation.

4.3. Différentes approches de suivi

Nous allons, dans cette section, classer les méthodes existantes en deux grandes catégories (suivi 2D et suivi 3D) selon le type de transformation. Ensuite, nous allons répartir chaque catégorie en plusieurs groupes selon la primitive visuelle utilisée. Enfin, chaque groupe sera divisé en sous groupes selon la manière de procéder.

4.3.1. Méthodes de suivi 2D

C'est le groupe de méthodes qui se basent sur une transformation qui permet de faire un lien entre un modèle 2D de l'objet avec les observations dans l'image courante. Selon la primitive visuelle utilisée, nous pouvons recenser les deux groupes suivants :

4.3.1.1. Méthodes de suivi basées sur l'intensité lumineuse

Ces méthodes se basent sur la texture. Elles procèdent par l'analyse des niveaux de gris dans l'image. Le modèle le plus simple est celui de Horn qui suppose que le signal lumineux de la projection d'un point 3D reste constant dans le temps et se traduit par :

$$\forall X,\ I_2(f(X)) = I_1(X)$$

Où X est un point de l'image, I_1 le niveau de gris du point X dans l'image modèle de l'objet, I_2 le niveau de gris du point X dans l'image courante de la vidéo et f la fonction de transfert entre les deux images. Ainsi, le critère de recalage dans ce cas est :

$$\Delta = \Sigma\ (\ I_2(f(X)) - I_1(X))^2$$

a- Flot optique

Le flot optique cherche à estimer pour chacun des pixels de l'image le vecteur déplacement entre deux images successives en se basant sur un critère de corrélation. Le principe consiste à estimer le déplacement en chaque point. L'approximation du signal lumineux est un développement de Taylor du 1er ordre.

Ces approches sont très couteuses en temps de calcul pour pouvoir être couramment utilisées dans le cadre d'une application temps-réel. De plus, le mouvement est estimé en chaque point de l'image sans exploiter la contrainte imposée par la structure de la scène. Une information de si bas-niveau est difficilement exploitable pour remonter à la notion d'objets.

b- Suivi de région ou de motif

Ces méthodes sont dites globales puisqu'elles prennent en considération tous les pixels représentant le motif à suivre. Une analyse des niveaux de gris sera ensuite faite sur ces pixels. En effet, ceci permet d'estimer le mouvement d'une région dans l'image. L'avantage de ces méthodes est qu'elles ne demandent pas une extraction préalable de primitives visuelles. Par conséquent, beaucoup de travaux dans ce sens ont vu le jour.

Darell et al. [30], Brunelli et al. [21] proposent de maximiser un critère de corrélation entre un vecteur caractérisant le modèle de référence et le contenu de l'image. Les temps de calcul, significatifs dans ce cas, peuvent être réduits en travaillant dans des sous-espaces de la représentation initiale de l'image. La limitation principale de ces approches est leur manque de résistance au regard des occultations. Black and

Jepson [15] ont surmonté cette limitation en reconstruisant les parties occultées. Ils remplacent la norme quadratique généralement utilisée pour construire l'approximation de l'image dans l'espace propre par une norme d'erreur robuste. Cette reconstruction revient à une minimisation d'une fonction non linéaire, optimisée en utilisant une méthode de descente de gradient simple. Ils utilisent la même stratégie pour trouver la transformation paramétrique alignant le motif sur l'image. Des travaux similaires reposant sur l'utilisation d'espaces propres ont été réalisés par K. Deguchi [31] pour le suivi d'objets et du positionnement d'un robot par vision.

Plus récemment, de nouvelles méthodes efficaces de suivi ont été proposées : le problème du suivi est formulé comme un problème de recherche du meilleur ensemble de paramètres (au sens des moindres carrés) décrivant le mouvement et la déformation de la cible au cours de la séquence. Dans ce cas, les variations des paramètres sont écrites comme une fonction linéaire d'une image de différence (la différence entre l'image de référence et l'image courante). Cette approche est très efficace car le mouvement peut être facilement déduit de l'image de différence. Cootes et al [26] l'utilisent pour estimer dynamiquement les paramètres d'un modèle de visage en se basant sur l'apparence (modèle 2D). Hager et Belhumeur [41] l'utilisent dans un contexte général pour le suivi d'objet, pour des mouvements planaires affines. Leur but est de suivre un motif quelconque. Pour cela, ils estiment le paramètre μ d'une transformation 2D qui permet de suivre un motif de niveaux de gris d'image en image. La minimisation du critère de recalage repose sur le calcul d'une matrice Jacobienne J reliant la variation de l'intensité par rapport aux n paramètres μ du mouvement. Il existe de nombreuses variations autour de ces travaux, notamment sur le calcul de la matrice Jacobienne J. Son calcul dépendant des gradients de l'image courante, le cout de calcul peut être élevé si elle est ré-estimée à chaque étape du processus de minimisation itératif en chaque point. Certaines hypothèses dans [41] permettent d'alléger le calcul de la matrice Jacobienne en exprimant le gradient de l'image courante en fonction de celui de l'image de référence. L'hypothèse de conservation de la luminosité permet ainsi d'effectuer hors-ligne le plus gros des calculs, pour les transformations affines. Jurie et Dhome [50, 51, 52] proposent un cadre d'apprentissage de la pseudo-inverse de la matrice Jacobienne. L'avantage est de minimiser les coûts de calcul en ligne. L'apprentissage permet d'éviter le développement du premier ordre qui suppose que les niveaux de gris sont une fonction linéaire des paramètres de la transformation. La

matrice apprise peut ainsi tenir compte de non-linéarités et par conséquent a priori, représente plus fidèlement la relation entre niveaux de gris et paramètres du mouvement. En plus, cette méthode est une extension des travaux de hager [41] dans le sens où elle supporte les transformations projectives comme l'homographie par exemple. Elle comprend deux étapes. Une phase d'apprentissage hors ligne est dédiée au calcul de la matrice d'interaction qui lie les variations d'intensité lumineuse du motif de référence 2D de l'objet suivi dans une zone d'intérêt à son déplacement fronto parallèle. Par définition, un mouvement fronto parallèle est un mouvement tel que l'objet se déplace dans des plans parallèles au plan image. Sous l'hypothèse d'un tel mouvement, l'aspect apparent de l'objet suivi n'est pas modifié. Toutefois, sa position, son orientation planaire et sa taille peuvent changer. Une étape en ligne consiste à prédire la position de l'objet dans l'image (en position, échelle et orientation), à multiplier la différence entre le motif observé à l'endroit prédit et le motif de référence qui doit être suivi par la matrice d'interaction pour corriger les erreurs sur les mouvements fronto parallèles de l'objet dans l'image. Le problème du suivi du motif dans l'image se ramène alors à la correction des paramètres d'une transformation géométrique planaire par la détermination d'un vecteur d'offset. Compte tenu de la rapidité des traitements (multiplication d'une matrice par un vecteur) par rapport à la vitesse de déplacement des objets dans les séquences d'images, la méthode n'a pas besoin d'un algorithme de prédiction de mouvement. En effet, l'écart de position du motif entre deux images successives reste compatible avec les variations apprises lors de la phase d'apprentissage. Bouzenada et al. [17, 20] ont proposé d'utiliser un réseau de neurones artificiel (RDN) dans la phase d'apprentissage. En effet, l'avantage d'un RDN par rapport aux techniques classiques de modélisation non linéaire réside dans leur capacité à réaliser des modèles de précision équivalente avec moins de données expérimentales.

Le principal inconvénient de telles approches est lié à la mise à jour du motif de référence ou la prise en compte des changements d'illumination. Pour pallier a cet inconvénient, les travaux de Hager et Belhumeur [41] et de Black et Jepson [15] proposent d'utiliser une base d'images pour représenter un modèle d'illumination de la texture. L'image de référence en fait est la combinaison linéaire des images de la base, les parties occultées peuvent être reconstruites.

4.3.1.2. Méthodes de suivi basées sur les primitives géométriques

Ce type de suivi repose sur un processus bas-niveau qui permet d'extraire localement des points de contour. Sa recherche est généralement effectuée le long de la normale au contour. Aussi, le cadre probabiliste a été intensivement utilisé pour le suivi de primitives, comme la distance de Hausdorff, le filtre de kalman, etc.

a- suivi par modèle par silhouette

L'une des méthodes basées sur les techniques de transformation des distances est celle de Huttenlocher [45] qui consiste à retrouver la position d'un objet, rigide ou non, dans une séquence d'images à partir d'un modèle très simple. Le modèle utilisé pour effectuer la poursuite est simplement composé d'un ensemble de points d'arêtes. Il s'agit d'un modèle qui peut évoluer avec le temps à condition que la silhouette ne change pas trop rapidement d'une image à l'autre.

Soit M_t le modèle à l'instant t et M_0 le modèle de départ défini, dans la première image de la séquence, par l'utilisateur. Le modèle M_t permet de trouver la position de la cible dans l'image I_{t+1} à l'instant t+1. Pour cela, une détection d'arête de type Canny [22] est d'abord appliqué sur l'image pour qu'elle soit comparée au modèle M_t en utilisant la distance de Hausdorff. La position qui donne la distance la plus faible est considérée comme la position effective de l'objet recherché.

Afin d'améliorer cette méthode pour tolérer les changements brusques d'apparence, une base de modèles est générée. Cette base est enrichie au fur et à mesure par l'ajout d'un nouveau modèle jugé suffisamment différent des autres modèles déjà présents dans la base. Dans ce cas aussi, la distance de Hausdorff est utilisée pour apprécier la différence entre les modèles.

b- suivi par des contours actifs

Les contours actifs sont des courbes paramétriques qui peuvent être déformées sous l'influence de forces externes et internes utilisés pour le suivi dans [53]. L'énergie d'un contour actif [26] est définie par : une énergie externe provenant de l'image 2D et pouvant avoir différentes formes selon les caractéristiques qui doivent être capturées par le contour actif et une énergie interne qui a une influence sur l'évolution de la forme du

contour. Cette dernière possède deux paramètres : le premier régule l'élasticité de la courbe et le second sa flexibilité. Ces paramètres dépendent de la position de la courbe. L'évolution de la courbe se fait autour de la minimisation de la fonction de son énergie totale. Cette phase de minimisation peut être vue comme une évolution dynamique au cours de laquelle l'énergie dissipée est transformée en énergie cinétique et finalement le contour se trouve dans un état d'énergie plus stable.

Les travaux de Leymarie et Levine [57] sur la poursuite du déplacement des « fibroblastes » sur une surface plane se basent sur les contours actifs. Ils précisent la manière de choisir les différents paramètres (la tension, la rigidité, la masse, etc.), ainsi que les difficultés d'implémentation.

Les difficultés rencontrées avec les contours actifs résident dans le choix des nombreux paramètres à utiliser. En plus, le bruit peut attirer le contour vers d'autres objets hors l'objet poursuivi. Pour éviter ce problème, un filtrage des images s'avère nécessaire.

Une extension des contours actifs est présentée dans [85]. Cette approche propose l'intégration d'un filtre de Kalman pour faire converger le contour d'un point de vue temporel. Le filtre joue le rôle d'une mémoire qui permet de conserver approximativement la forme globale du contour.

Les contours actifs sont énormément utilisés pour le suivi d'objets déformables. Ils sont bien adaptés au suivi des cellules vivantes. Néanmoins, il faut pouvoir choisir correctement les termes de l'énergie interne et externe, ainsi que les constantes qui gouvernent l'évolution. Les méthodes se basant sur les contours actifs ont un caractère itératif couteux en temps de calcul. Par conséquent, elles sont mal adaptées pour des applications temps réel.

c- suivi par des modèles actifs de forme

Baumberg et Hogg [7] utilisent ce modèle pour le suivi d'une personne qui marche. Le modèle est formé d'un ensemble de points de contrôle soumis à une analyse par composantes principales pour réduire le nombre de degrés de liberté formant ainsi

un ensemble test. Le suivi est effectué à l'aide d'un filtre de Kalman permettant d'estimer les paramètres du modèle ainsi que ceux du mouvement. L'inconvénient de cette méthode est lié principalement à l'utilisation d'un modèle qui nécessite un entraînement. Autrement dit, le suivi ne pourra être correct que si certaines configurations figurent dans l'ensemble test.

d- suivi par des modèles de droites

Il s'agit de suivre des segments de droites appartenant à des contours extraits à partir d'une séquence d'images monoculaires. Différents algorithmes ont été implémentés pour résoudre ce problème en se basant sur un filtre de Kalman [37].

e- suivi avec utilisation des informations 3D

Koller [54] propose de suivre un véhicule sur une autoroute non pas pour calculer la pose mais plutôt pour analyser le trafic. Cette analyse comprend principalement le recensement du nombre de véhicules, l'estimation de leur vitesse et la détection de certains événements, à savoir : changement de voie, arrêt, etc.

Afin de traiter le problème d'occlusions qui apparaissent dans ce genre d'application, la distance du véhicule par rapport à la caméra est évaluée. La méthode se base principalement sur une combinaison entre un modèle complexe du background et un filtre de kalman. Ce filtre sert pour prédire l'évolution du modèle.

Cette méthode passe par trois étapes :

- Création d'un masque qui correspond aux objets en mouvement dans la scène. Il s'agit ici de détecter des véhicules qui entrent dans le champ de vision de la caméra.
- Approximation de chaque objet en mouvement par une spline cubique avec douze points de contrôle. Pour cela, le déplacement de chaque véhicule est décrit en utilisant un modèle de mouvement affine très simple avec une translation 2D et un facteur d'echelle.
- Le suivi est alors effectué sous forme d'une boucle de prédiction et de mesures. L'étape de prédiction est faite à l'aide de deux filtres de kalman : l'un pour la prédiction de la position du véhicule, l'autre pour la prédiction de sa forme.

Le suivi simultané de plusieurs véhicules nécessite l'ajout de traitements supplémentaires à cause des chevauchements (occlusions) qui apparaissent fréquemment dans une image 2D. Le raisonnement est simple, le véhicule le plus proche de la caméra peut cacher un autre un peu plus loin de celle-ci. Il est nécessaire, alors, d'estimer la distance qui sépare les véhicules et la caméra. Ceci ne pose aucun problème puisque la caméra est correctement calibrée.

Les travaux de Gil [38] sont similaires à celui de Koller [54]. La seule différence est au niveau de la modélisation du véhicule.

4.3.2. Méthodes de suivi 3D

Il s'agit, pour ces méthodes, d'estimer pour chaque image de la vidéo, les paramètres de position ou de déplacement de la caméra par rapport à l'objet.

4.3.2.1. Estimation à partir des transformations 2D

Les méthodes de suivi 2D peuvent servir pour déterminer la géométrie de la scène. Cette technique dite auto-calibration a été présentée en section 2.3.4 au chapitre 2. Ces approches ont souvent plusieurs solutions. L'incertitude peut être levée en utilisant soit une information a priori de la scène, soit au moins trois vues de la même scène. Pour simplifier ces approches, certains exploitent l'existence des structures planes dans la scène [83].

4.3.2.2. Méthodes basées modèle

Le principe des approches basées modèle, qui estiment la pose 3D à partir d'une séquence d'images, est le suivant:

- Obtenir une pose initiale de l'objet à suivre (phase initialisation). Cette phase est souvent faite interactivement.
- Projeter le modèle 3D de l'objet à suivre sur le plan image.
- Recaler la pose de l'objet avec le modèle projeté en se basant sur une mesure de comparaison.
- Prédire la pose dans l'image suivante à partir de la pose ajustée. Il est nécessaire, ici, de connaître certains paramètres (vitesse ou accélération).

Ces méthodes utilisent des approches mathématiques pour estimer les paramètres d'une transformation. Ces approches ont une influence importante sur le succès du suivi. Les solutions possibles sont nombreuses : estimation linéaire, non-linéaire (gauss-newton, levenberg-marquardt, filtre de kalman etendu), cadre probabiliste (filtres particulaires). Le système de prédiction entraine en ce qui concerne les filtres de kalman une erreur de traînage dans l'estimation des paramètres pouvant rompre le suivi en cas de discontinuité temporelle trop importante. De plus, ils reposent sur un bruit gaussien des observations, hypothèse souvent fausse. Les filtres particulaires, par contre, permettent de surmonter ce phénomène mais en gardant plusieurs jeux de paramètres en mémoire et par conséquent demandent plus de temps de calcul. Ainsi, ils ne s'adaptent pas bien aux applications temps réel. Nous allons, dans la suite, présenter brièvement et à titre d'exemples quelques méthodes:

a- Méthode de Gennery

Elle est parmi les premières méthodes comportant un système complet de suivi de cibles avec estimation de la pose 3D à partir d'images [36]. Elle consiste à suivre des objets polyédriques. La comparaison de la projection du modèle avec les données extraites de l'image est faite selon deux types de primitives : Les points d'intérêt et les segments de droite. Pour les points d'intérêt, la mesure prise est la différence entre les coordonnées de la projection du point d'intérêt et le point image correspondant. Pour les segments de droite, une discrétisation, en un ensemble de points distants de trois pixels environ, est appliquée sur chaque segment projeté du modèle. A chacun de ces points est associé le point d'arête de l'image dont le gradient est le plus proche. La mesure prise dépend ainsi de deux mesures : la distance orthogonale qui est la distance entre le point d'arête et le segment de droite et la distance parallèle qui est la distance entre la projection orthogonale du point d'arête et le point modèle.

b- Méthode de Baker

Cette méthode a été définie dans le cadre du projet européen ESPRIT. Elle propose de suivre des véhicules. La modélisation a été faite par un modèle en fil de fer. La mesure de comparaison entre les données extraites de l'image et la projection du modèle est faite par une fonction dite « évaluation iconic ». Il s'agit d'abord de projeter le modèle en éliminant les segments de droite qui ne sont pas visibles. Dans le voisinage

de chaque segment projeté, le gradient de l'image est calculé dans la direction perpendiculaire. Ces gradients sont ensuite moyennés parallèlement au segment et la valeur maximale est évaluée. La probabilité pour qu'un segment ait une telle valeur est estimée à l'aide de tables pré-calculées avec la technique de monté Carlo. Enfin, Ces probabilités sont utilisées pour tous les segments visibles du modèle dans l'image dans un test du χ^2. Cette fonction d'évaluation est à la base de trois approches différentes pour l'ajustement de la pose. Baker propose de trouver un maximum local de la fonction « évaluation iconic » pour ajuster la pose. Worral [93] propose un raffinement de la pose en utilisant des forces élémentaires. Worral [92] propose de remplacer l'évaluation de l'effet d'une force 3D, faite dans l'approche antérieure, par la recherche du minimum d'un terme d'erreur en appliquant un algorithme classique de minimisation par la méthode des moindres carrés.

c- Méthode de Koller

Elle permet aussi de suivre des véhicules modélisés en fil de fer [54]. La pose est définie par trois paramètres : deux définissent la position du véhicule dans le plan de la route et le troisième définit l'orientation par rapport à la perpendiculaire de la route. La projection du modèle est faite en éliminant les arêtes non visibles, l'extraction des segments est effectuée par la méthode de Korn [55] et la mise en correspondance est assurée par la méthode de Dériche et Faugeras [32].

d- Méthode de Lowe

La méthode d'estimation définissant la position de l'objet dans l'espace est proposée par Lowe [58] en se basant sur un modèle d'approximation polygonale. Des améliorations [59] ont été faites sur cette méthode pour qu'elle soit plus robuste en éliminant les faux appariements.

e- Méthode de Harris

Cette méthode propose de suivre un objet rigide en se basant sur un modèle simple. Ce modèle est représenté par les points d'intérêt se trouvant sur des arêtes bien marquées. Cette méthode présente de bonnes performances en poursuivant des objets ayant des vitesses de rotation allant jusqu'à 10 radian/s et une accélération angulaire qui peut aller jusqu'à 100 radian/s.

4.3.3. Méthodes hybrides 2D/3D

Ces méthodes reposent, à la fois, sur l'estimation du mouvement 2D de l'objet et sur le calcul de sa pose 3D. Le principal avantage de ces méthodes est d'éviter l'étape de prédiction de type Kalman en estimant le modèle de mouvement qui sera exploité pour fournir une initialisation correcte du calcul de pose. Notons, qu'il existe un certain nombre de travaux en ce sens. Marchand [25], par exemple, propose une méthode, dans le cadre d'un projet EDF pour des tâches de maintenance et de surveillance en milieu hostile, permettant un suivi robuste et rapide d'objets complexes pouvant être approximativement modélisés par une forme polyédrique. Un modèle de mouvement affine 2D est estimé, à partir des déplacements orthogonaux calculés le long des projections des arêtes du modèle dans l'image, grâce à un algorithme robuste. Puisque, le modèle de mouvement affine ne permet pas de représenter totalement le mouvement 3D de l'objet, une seconde étape est nécessaire pour recaler la projection du modèle de l'objet dans l'image. Cette étape consiste à calculer la pose de l'objet par rapport à la caméra. Il propose, pour cela, une minimisation itérative d'une fonction d'énergie non linéaire par rapport aux paramètres de la pose.

4.4. Conclusion

Les approches de suivi 2D basées sur les primitives géométriques ont l'avantage principal d'être simples lors de la mise en œuvre et rapide en temps d'exécution. Par contre, elles ne permettent pas le suivi de motifs complexes qui ne peuvent pas être modélisés à l'aide de primitives locales ou de contraintes 2D uniquement.

Pour les approches de suivi 3D, elles sont un bon moyen pour le calcul de la pose. Néanmoins, les limitations dues aux primitives basées contours souvent exploitées dans ce cadre entraînent un besoin d'améliorer ces méthodes.

Quant aux approches hybrides, elles apparaissent comme les mieux adaptées pour les applications faisant partie du domaine de la réalité augmentée.

Dans ce manuscrit, nous nous sommes intéressés aux approches 2D basées sur le suivi de région ou de motif puisque dans notre approche, nous n'avons pas eu besoin pour le moment d'un calcul de pose. Les deux chapitres suivants seront consacrés à la

description des améliorations que nous avons essayé de faire au niveau du suivi pour des applications de réalité augmentée.

CHAPITRE 5

Nouvelle approche de suivi de doigt :
Application au Tableau Magique.

Nous allons présenter ici, les améliorations que nous avons apportées à l'application « tableau magique » au niveau de la fonction du suivi de doigt. Ce travail constitue notre premier résultat [8] de recherche dans ce domaine ô combien complexe mais très plaisant. Pour ce faire, nous allons tout d'abord décrire d'une façon assez détaillée cette application.

5.1. Tableau magique

Le tableau magique [1, 9, 10, 11, 12, 13, 14] fait partie des applications appartenant au domaine de la réalité augmentée. Il n'est autre qu'un tableau blanc conventionnel amplifié par des services électroniques capables de contourner ses insuffisances intrinsèques, à savoir : la réorganisation spatiale des inscriptions, l'archivage, la diffusion et la collaboration synchrone à distance. Il fait partie des systèmes fortement couplés [27, 28, 42, 89, 90, 91] où les représentations virtuelles et physiques sont parfaitement synchronisées. Il est composé de :

- Un tableau blanc conventionnel sur lequel on écrit avec les feutres usuels à encre effaçable.
- Une caméra vidéo mobile pour capturer les inscriptions se trouvant sur le tableau blanc.
- Un ordinateur permettant de traiter le flux vidéo de la caméra.
- Un projecteur pour afficher les retours d'information sur le tableau blanc.

5.1.1. Motivations

Le tableau magique se justifie comme support à l'activité de brainstorming sur tableau blanc. Cette activité se traduit par la production et l'organisation d'idées entre individus œuvrant pour un projet commun. Le tableau blanc est un artefact largement répondu dans les bureaux et les écoles. Ces caractéristiques font de lui un outil adapté à l'activité de réflexion. Néanmoins, il présente certaines lacunes qui devront être prises en charge par les services électroniques offerts par le tableau magique.

5.1.1.1. Qualités du tableau blanc

Un tableau blanc offre trois propriétés favorables à l'activité de réflexion collective : disponibilité immédiate, facilité et rapidité d'utilisation, surface partagée servant de mémoire commune.

a- **Disponibilité immédiate** : Une réunion de travail peut être décidée à tout moment et des fois sans aucune programmation préalable. Chaque membre de cette réunion utilisera sûrement un tableau blanc si ces idées nécessitent un support écrit pour être bien expliquées aux autres membres. Cette utilisation est due au fait que ce tableau est disponible en permanence et sans délai de mise en route. Certains tableaux augmentés ne remplissent pas cette propriété. Citons : le liveboard, le softboard, le smartboard et le mimio. Tous ces systèmes enregistrent la trajectoire des outils de dessin que manipulent les utilisateurs. Mais les inscriptions produites avant la mise en marche sont perdues. L'utilisation de ces systèmes implique que le logiciel de gestion du tableau soit actif au préalable. D'autres systèmes tels que : le zombieboard et le brightboard, autorisent une activation à posteriori des services électroniques. Ces systèmes

sont fondés sur un tableau blanc conventionnel. Donc, ils satisfont la propriété de disponibilité immédiate.

b- **Facilité et rapidité d'utilisation** : Tout membre d'un groupe de réflexion et quelque soit son profil de compétence peut utiliser d'une manière aisée et rapide le tableau blanc conventionnel. Les feutres et l'encre ont des propriétés dont les participants peuvent tirer parti. En particulier, ils autorisent un bon contrôle de la forme et de l'épaisseur du trait, l'échange prompt entre feutres de différentes couleurs tenus dans une main, et des corrections rapides en effaçant l'encre au doigt.

c- **Surface collective de grande taille** : Le fait que le tableau blanc a une assez grande taille par rapport aux autres supports de dessin usuels, il favorise le travail en groupe sur sa surface. Le liveboard, le softboard et le smartboard ont des tailles limitées en raison du coût de fabrication. Par contre, le zombieboard et le brightboard ont été expérimentés sur des tableaux blancs couvrant des murs entiers. C'est le cas aussi du tableau magique.

5.1.1.2. Lacunes du tableau blanc

Les insuffisances du tableau blanc conventionnel sont de nature fonctionnelle et elles sont au nombre de trois classées par ordre d'importance:

a- **Réorganisation spatiale des inscriptions** : Puisque l'écriture est facile et rapide sur un tableau blanc, il permet facilement la concrétisation d'idées. Ces idées sont souvent transcrites sur le tableau blanc sans un ordre précis. La phase de transcription est toujours suivie d'une phase de réorganisation. L'organisation des idées est reflétée par les relations spatiales des inscriptions qui les concrétisent. Or, le tableau blanc ne facilite pas cette tâche de réorganisation. En pratique, cette dernière nécessite la recopie des inscriptions sur une partie vierge du tableau puis l'effacement des inscriptions originales. Il s'agit d'un processus lourd, sujet à erreur (suite aux recopies) et peu adapté au changement fréquent entre phases de génération et phases de réorganisation. Il est donc souhaitable d'augmenter le tableau afin de fournir les moyens de réorganiser les

inscriptions de façon efficace. Les LiveBoard, SoftBoard, SmartBoard et Mimio numérisent à la volée les inscriptions portées au tableau. La version numérique des inscriptions est immédiatement disponible pour être réorganisée à volonté. Dans le cas du LiveBoard, il n'existe pas d'inscription physique : on utilise des crayons optiques qui produisent immédiatement une inscription électronique. Cette approche, qui remplace l'outil usuel (le feutre à encre effaçable), ne satisfait pas le principe de conservation des outils naturels. L'usage de crayon optique limite, par exemple, les possibilités de contrôle de la forme du trait et l'effacement des inscriptions au moyen du doigt. Dans le cas des SoftBoard, SmartBoard et Mimio, les participants produisent des inscriptions physiques avec des feutres conventionnels dont la trajectoire est enregistrée numériquement (respectivement, de façon optique, tactile ou par ultrasons). Cette approche a l'avantage de conserver l'usage d'un feutre normal pour le dessin, mais nécessite que l'utilisateur efface l'encre physique lorsque ce sont les propriétés électroniques des inscriptions qui prévalent. Notons que l'effacement de l'encre physique n'est nécessaire qu'au premier déplacement. L'inscription existe ensuite uniquement au format électronique et peut ainsi être manipulée. La réorganisation spatiale nécessite que l'utilisateur puisse désigner des emplacements afin d'indiquer au système les informations à déplacer et leur destination. Dans le contexte du tableau magique, la désignation au doigt semble adaptée : en situation d'explication, l'index est utilisé pour souligner une information à l'adresse de l'auditoire. Par extension, la désignation au doigt servira également à la désignation des éléments de contrôle permettant d'exécuter les commandes du tableau magique.

b- **Archivage et diffusion** : Pour garder le contenu d'un tableau blanc, il faut impérativement le faire manuellement en prenant des notes. Cette prise de note est non seulement fastidieuse mais souvent infidèle, surtout avec des inscriptions comportant des schémas. Cette conservation peut servir : de point de départ pour une prochaine réunion, pour informer les personnes intéressées qui n'ont pas assisté à cette rencontre, etc. Tous les systèmes augmentés offrent cette fonction. Mais, la qualité visuelle du service varie en fonction des techniques adoptées.

c- **Collaboration synchrone à distance** : L'usage du réseau informatique permet d'envisager le déroulement de réunions entre personnes délocalisées. La collaboration sur des activités de réflexion entre personnes distantes représente un défi.

5.1.2. Fonctionnement

Pour réaliser le système du tableau magique, plusieurs fonctions doivent être prises en charge : la transformation entre repères, la capture des inscriptions [86], le suivi du doigt et l'interaction [29]. Dans la suite, nous allons donner une brève description des quatre fonctions citées ci-dessus et nous consacrerons la section suivante à la fonction du suivi de doigt qui est la partie qui nous intéresse dans ce mémoire.

5.1.2.1. Transformation entre repères

Le fait d'utiliser un projecteur vidéo et une caméra implique la présence de deux repères distincts : le repère de l'image projetée et le repère de l'image capturée. Il est donc nécessaire de transformer les coordonnées exprimées dans le repère capturé en coordonnées exprimées dans le repère projeté et réciproquement. Par exemple, la position du doigt, qui est extraite des images traitées par la vision par ordinateur est exprimée dans le repère capturé. En sortie, l'affichage d'un curseur à l'emplacement du doigt, nécessite de connaitre la position du doigt dans le repère projeté.

5.1.2.2. Capture des inscriptions

Le tableau magique inclut un service de capture à résolution variable utile à la mise en œuvre des fonctions d'archivage et de diffusion citées ci-dessus. Un certain nombre de techniques sont utilisées pour améliorer la qualité de la capture (seuillage adaptatif, mosaïque).

5.1.2.3. Suivi du doigt

La fonction « suivi du doigt » va permettre à l'utilisateur d'indiquer au système les emplacements du tableau pour que celui-ci réalise le service souhaité qui est lié à l'emplacement du doigt de l'utilisateur.

5.1.2.4. Interaction

Le doigt de l'utilisateur est le seul moyen d'interaction avec le système. Il servira à gérer une sélection (sélectionner, déplacer, indiquer, etc.) et à exécuter les commandes immédiates (sauvegarde, impression, copier, coller, etc.).

5.2. Fonction suivi du doigt

La fonction de suivi du doigt est réalisée en deux grandes phases : Une phase d'initialisation pour reconnaître le motif à suivre et une phase de suivi proprement dite.

5.2.1. Phase d'initialisation :

Dans cette phase, l'utilisateur est amené à maintenir pendant plus de 0,5 s son doigt dans une zone rectangulaire projetée sur le tableau blanc. Ce temps est nécessaire pour faire la différence entre une présentation volontaire du doigt dans la zone sensible et un passage involontaire du doigt par cette zone. La détection du doigt de l'utilisateur est faite par la technique de différence d'images. Une fois détectée, l'apparence du doigt est mémorisée dans un motif de (32x32 pixels).

5.2.2. Phase de suivi :

Le suivi adopté dans ce cas est le suivi par corrélation. Il consiste à rechercher la partie de l'image la plus ressemblante au motif. La recherche du motif dans une nouvelle image nécessite de corréler le motif et la partie de l'image à traiter. La corrélation de deux images permet d'évaluer leur similarité. Plusieurs formules existent pour le calcul de la similarité. La plus utilisée est NCC pour « Normalised Cross-Correlation » qui prend en compte la luminosité générale du motif et de l'image. Soit M un motif rectangulaire de taille u x v, et I l'image à traiter. L'écart entre le motif et la partie de l'image située aux coordonnées (x,y) est donnée par :

$$\mathrm{NCC}(x,y) = \frac{\sum_{u,v} M(u,v)\cdot I(x+u,y+v)}{\sqrt{\sum_{u,v} M^2(u,v)\cdot \sum_{u,v} I^2(x+y,y+v)}}$$

La valeur du NCC est comprise entre 0 et 1. Il est évident que cette valeur est maximale (égale à 1) lorsque le motif et l'image sont identiques à un coefficient de luminosité globale prés.

En plus, la recherche du motif n'est pas faite dans ce cas sur toute l'image mais sur une zone de recherche de (75 x 75 pixels) (Fig. V.1) jugée nécessaire pour contenir le motif recherché sachant que la vitesse de déplacement du doigt de l'utilisateur ne peut pas aller au delà de 200cm/s.

Le suivi par corrélation assure la stabilité statique de l'information extraite sous réserve que l'entité à suivre ne change pas d'apparence par rapport au motif qui sert de référentiel. Or, dans le contexte d'utilisation du tableau magique, l'apparence du doigt en déplacement varie. La raison essentielle est la suivante : lors d'un mouvement non contraint, l'index adopte l'orientation de l'avant-bras qui varie en fonction de la hauteur de l'inscription désignée sur le tableau (Fig. V.2).

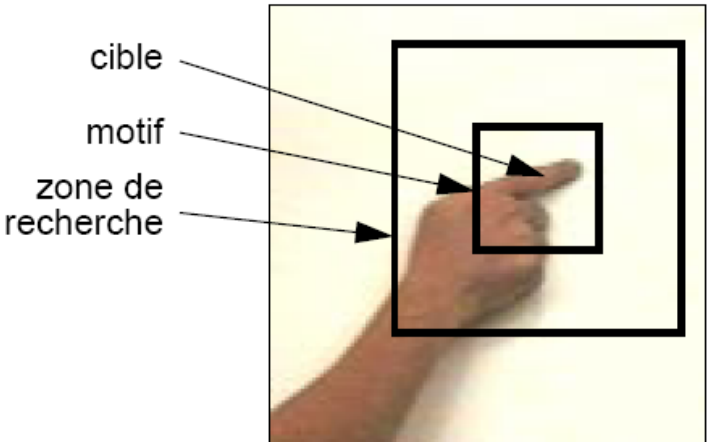


Figure V.1 - Motif et zone de recherche

En pratique, la méthode utilisée dans la fonction de suivi (suivi par corrélation) impose des contraintes pour le bon fonctionnement du système. Lorsque la main de l'utilisateur effectue une rotation significative, la stabilité statique n'est plus assurée. L'absence de la stabilité statique entraine une forte dégradation de l'interaction et par là il devient difficile de réaliser les tâches souhaitées. Il convient donc d'envisager de nouvelles solutions pour améliorer les performances du suivi de doigt.

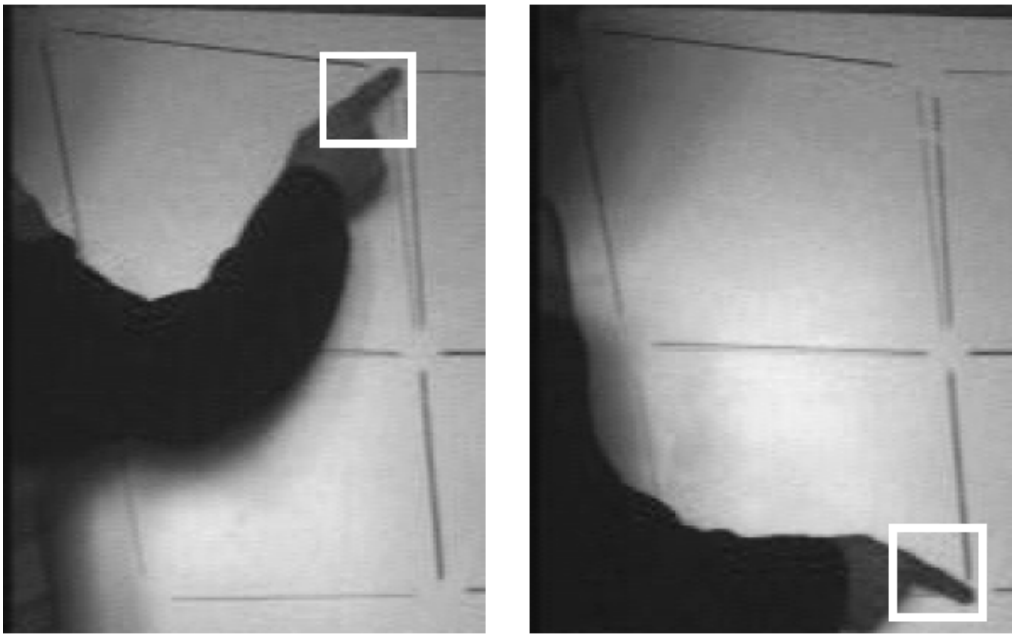


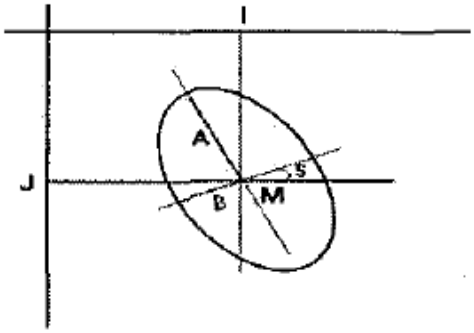
Figure V.2 - Variation d'apparence de l'index en fonction de l'orientation du bras

5.3. Méthode de suivi proposée

La méthode de suivi que nous avons proposé est une méthode de suivi basée pixels et qui prend en considération les mêmes conditions que celles prises lors du suivi par corrélation en section 3. Ainsi, notre zone sensible est de forme elliptique et de taille 11 pixels (3,3 cm) pour le grand axe A de l'ellipse et 7 pixels (2,1 cm) pour le petit axe B de l'ellipse. Soit une surface de A x B x π = 242 pixels (le quart de la zone proposée lors du suivi par corrélation). La zone de recherche aura une taille de (84 x 98) pixels donc un peu plus large que celle du suivi par corrélation (Fig. V.1). Le but de cette méthode est de permettre une utilisation plus confortable du tableau magique en permettant une certaine liberté de bouger le bras. Nous avons défini dans cette méthode trois phases essentielles : initialisation, détection et suivi.

5.3.1. La phase d'initialisation

Dés la mise en marche, le système projette une zone sensible sous forme d'une ellipse dans un endroit précis du tableau blanc pour détecter le motif à suivre (doigt de l'utilisateur). Cette zone elliptique sera donc identifiée par cinq paramètres (Fig. V.3).



FigureV.3- Les cinq paramètres de l'ellipse contenant le motif.

Avec M centre de l'ellipse ayant (I , J) comme coordonnées, S angle de rotation, A et B respectivement grand axe et petit axe. On peut constituer à ce niveau ce que nous avons appelé « le vecteur de paramètres de la zone sensible » qui est égal à (X_{init}, Y_{init}, A, B, S_{init}) où (X_{init}, Y_{init}) représente les coordonnées du centre de l'ellipse, S_{init} l'angle de rotation et A et B respectivement le grand et le petit axe de l'ellipse. Puisque A et B sont invariants dans notre cas, le vecteur de paramètres de la zone sensible n'est représenté que par le triplet (X_{init}, Y_{init}, S_{init}).

Autour de cette zone sensible, un filet d'ellipses (Fig. V.4) est placé virtuellement pour couvrir toute la zone de recherche. On construit alors ce que nous avons appelé la matrice des paramètres du filet contenant trois colonnes (Xi, Yi) centre de l'ellipse et S son angle de rotation et k lignes représentant les k ellipses couvrant toute la zone de recherche autour de la zone sensible.

X_1	Y_1	*S*
⋮	⋮	⋮
X_i	Y_i	*S*
⋮	⋮	⋮
X_K	Y_K	*S*

La zone sensible sera ensuite échantillonnée (Fig. V.5) pour préparer la phase suivante. Les points où sont réalisés les échantillonnages sont répartis sur un ensemble d'ellipses concentriques déduites par homothétie de l'ellipse englobante. Ces ellipses

sont échantillonnées de la plus petite à la plus grande. Le nombre de points sur chacune d'elles est prédéfini pour représenter un pas de parcours quasi constant. L'échantillonnage débute à partir de l'orientation du demi grand axe supérieur de l'ellipse. L'ensemble des valeurs échantillonnées est ainsi stocké toujours dans le même ordre.

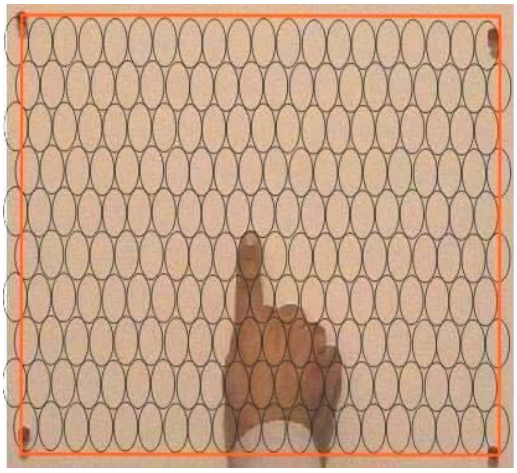


Figure V.4 - Le filet d'ellipse autour de la zone sensible contenant le doigt de l'utilisateur

Figure V.5- Echantillonnage à l'intérieur d'une zone elliptique. Les points représentent les endroits où sont mémorisées les valeurs des niveaux de gris

L'ensemble des ellipses couvrant toute la zone de recherche sera échantillonné de la même façon que la zone sensible. Ainsi, on aura ce que nous avons appelé la matrice de forme du filet qui est une matrice ayant comme nombre de lignes, le nombre d'ellipses couvrant la zone de recherche et comme nombre de colonnes, le nombre de points résultant de l'échantillonnage de chaque ellipse. Cette matrice se présente comme suit :

X_{11}	Y_{11}	Z_{11}	……	X_{1j}	Y_{1j}	Z_{1j}	……	X_{1N}	Y_{1N}	Z_{1N}
.				.				.		
.				.				.		
X_{i1}	Y_{i1}	Z_{i1}	……	X_{ij}	Y_{ij}	Z_{ij}	……	X_{iN}	Y_{iN}	Z_{iN}
.				.				.		
.				.				.		
X_{K1}	Y_{K1}	Z_{K1}	……	X_{Kj}	Y_{Kj}	Z_{Kj}	……	X_{KN}	Y_{KN}	Z_{KN}

Chaque élément (X_{ij} , Y_{ij} , Z_{ij}) de la matrice représente un point appartenant à la zone de recherche. (X_{ij} , Y_{ij}) représente ces coordonnées et Z_{ij} son niveau de gris. L'indice i designe que le point appartient à l'ellipse i et l'indice j désigne qu'il est le jème point de l'échantillon.

5.3.2. La phase de détection

Cette phase est lancée dés que l'utilisateur maintient son doigt pendant plus d'une demi-seconde dans la zone sensible. Le motif à suivre (doigt de l'utilisateur) sera détecté par la **technique de différence d'images** et mémorisé.

5.3.2.1. Calcul du vecteur de forme référentiel

Une fois le motif mémorisé, l'étape de l'estimation du vecteur de forme commence. Elle consiste à déterminer pour chaque point de l'échantillonnage, sa composante manquante et qui est la valeur du niveau de gris. Le motif sera représenté alors par un vecteur de taille N (nombre de points échantillonnés dans la zone sensible). Chaque composante de ce vecteur à trois éléments : Xi et Yi les coordonnées du ième point de l'échantillon et Zi son niveau de gris. D'où le Vecteur de forme qui est égal à ((X_1, Y_1, Z_1), ……., (Xi, Yi, Zi), ……,(X_N, Y_N, Z_N)). Xi et Yi ont été déterminés lors de la phase d'initialisation et Zi qui est déterminé dans cette étape.

La représentation ainsi faite du motif sera sensiblement la même puisque les valeurs enregistrées sont positionnées dans un repère lié à l'ellipse et donc au motif. De

plus, pour garantir une certaine insensibilité aux changements des conditions d'éclairement de la scène, le vecteur de forme sera centré et normé.

5.3.2.2. Calcul de la matrice d'interaction

Pour estimer cette matrice, nous effectuons un apprentissage sur un ensemble de jeux de données. Nous pratiquons un ensemble de perturbations (Rotations et translations) sur la zone sensible et nous effectuons pour chaque perturbation une différence d'image entre le motif original et la zone perturbée.

Ainsi, si nous prenons M mesures de ce type, N étant la taille du vecteur représentant le motif, il est possible d'estimer la matrice d'interaction A si M > N. Cela revient à résoudre M systèmes d'équations à N inconnues. La matrice A est estimée par une minimisation au sens des moindres carrés [50, 51, 52], en utilisant un algorithme basé sur une décomposition en valeurs singulières. En réalité, la résolution d'un seul système linéaire ou plus exactement le calcul d'une seule matrice pseudo inverse est nécessaire. En effet, si nous notons $\Delta I^j = (\Delta i_1^j, \Delta i_2^j, \ldots\ldots, \Delta i_N^j)$ le vecteur de différence entre le motif de référence et le motif correspondant à la j^{ieme} perturbation et $\Delta\theta^j$ la variation d'orientation entre les ellipses permettant de calculer ces deux vecteurs. Pour obtenir la ligne Aθi de la matrice d'interaction relative à l'orientation de l'ellipse, nous arrivons au système linéaire suivant :

Δi_1^1	Δi_2^1	.	Δi_N^1
.	.	.	.
Δi_1^j	Δi_2^j	.	Δi_N^j
.	.	.	.
Δi_1^M	Δi_2^M	.	Δi_N^M

X

$A\theta_1$
.
$A\theta j$
.
$A\theta_N$

=

$\Delta\theta^1$
.
$\Delta\theta^j$
.
$\Delta\theta^M$

Noté sous forme matricielle comme suit : $M_{\Delta I} \times A\theta = \Delta\theta$

La solution est alors obtenue par : $A\theta = (M^t_{\Delta I} \times M_{\Delta I})^{-1} \times M^t_{\Delta I} \times \Delta\theta = M^+_{\Delta I} \times \Delta\theta$

La matrice $M^{+}_{\Delta I}$ est qualifiée de pseudo inverse de $M_{\Delta I}$. Il est évident que le calcul des quatre autres lignes de la matrice d'interaction A utilise le produit de la même matrice avec des vecteurs de perturbations différents (ΔX_C, ΔY_C, ΔR_1, ΔR_2).

5.3.3. La phase de suivi

Le principe de base est de limiter la recherche de la cible à une zone réduite centrée autour de la dernière position connue de l'entité à suivre, mais avec une représentation différente à celle utilisée dans le suivi par corrélation. Au lieu de parcourir toutes les sous parties de l'image ayant la même taille que le motif pour calculer le maximum des pics de corrélation. Nous représentons la zone de recherche par un filet d'ellipse couvrant la surface de toute cette zone (Fig. V.4). Cette phase est un processus itératif qui commence par : faire une différence entre l'image courante de la séquence et l'image précédente, ensuite, calculer la nouvelle position de la cible dans l'image courante et enfin, repositionner la zone de recherche selon la nouvelle position de la cible.

5.3.3.1. Effectuer une différence d'images

Dans chaque itération, une différence d'images est faite entre l'image courante et l'image précédente (où la position de la cible est connue) en ne gardant que les pixels ayant subi une variation de leur niveau de gris (Fig. V.6).

Chaque ligne de la matrice de forme du filet correspond à un vecteur de forme d'une ellipse faisant partie du filet couvrant la zone de recherche de notre cible, ainsi la matrice de forme du filet aura : K lignes représentant le nombre total d'ellipse couvrant la zone de recherche et N colonnes représentant les points échantillon de chaque ellipse. Chaque élément de cette matrice est composé de trois champs : X abscisse, Y ordonnée et Z niveau de gris. Ainsi, Zij représente le niveau de gris du point qui se trouve sur le ième point de la jème ellipse. Ce point a comme coordonnées (Xij, Yij).

Le résultat de la différence d'images est une matrice obtenue à partir de la différence entre la matrice de forme du filet estimée pour l'image courante et la matrice de forme du filet de l'image précédente contenant la dernière position connue de la cible.

Si le niveau de gris (z1) du pixel situé aux coordonnées (x, y) dans l'image courante est similaire à celui du pixel situé aux mêmes coordonnées dans l'image

précédente (z2), alors la différence des niveaux de gris (z3) sera égale à zéro. Si le niveau de gris (z1) du pixel situé aux coordonnées (x, y) dans l'image courante est par contre différent à celui du pixel situé aux mêmes coordonnées dans l'image précédente (z2), alors la différence des niveaux de gris (z3) sera égale à (z1).

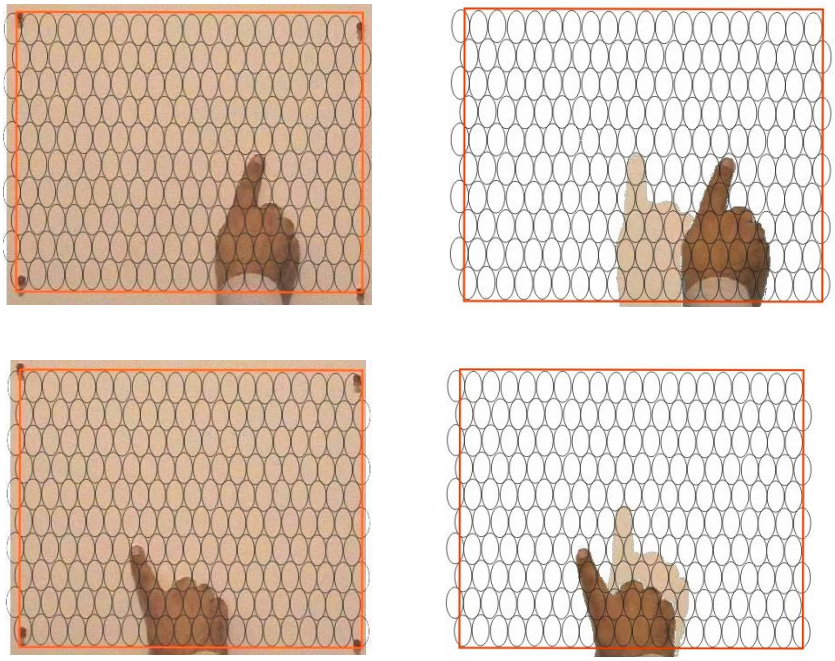


Figure V.6 : Deux exemples sur les résultats de la différence entre l'image contenant le motif référentiel et des images contenant plusieurs cas de déplacement de doigt.

5.3.3.2. Calculer la nouvelle position de la cible

Le résultat de la différence d'images est utilisé pour calculer la nouvelle position de la cible. En effet, pour chaque ligne (Vj / j : 1..K) de la matrice de différence, obtenue lors de l'étape précédente, ayant au moins la moitié des valeurs de niveau de gris (z) non nulles, nous calculons le produit de la matrice d'interaction A par la différence entre le vecteur de forme référentiel Vref et Vj. Il est à noter que les opérations effectuées ici ne concernent que la composante z (niveau de gris) des éléments du vecteur. D'où le vecteur de paramètres correctif Ec égal à :

$$Ec = A \times (Vref - Vj)$$

Le vecteur de paramètres correctif Ec obtenu permet de prédire une nouvelle position de la cible (vecteur de paramètres Epre) à partir du vecteur de paramètres Ej (se trouvant dans la matrice des paramètres du filet / j : 1..K) de l'ellipse correspondant au vecteur de forme Vj, comme suit : Epre = Ec + Ej.

Les éléments du vecteur correctif Ec seront utilisés pour calculer les nouvelles coordonnées (Xi, Yi) des pixels formant le vecteur de forme prédit Vpre en translatant les coordonnées (Xij, Yij) des éléments du vecteur Vj par des équations géométriques (Fig. V.8). Dans ce cas, les opérations seront effectuées sur les champs x et y correspondant aux coordonnées des pixels et non sur la composante z correspondant au niveau de gris.

Le résultat obtenu (Vpre) va être estimé par l'évaluation des niveaux de gris z et comparer au vecteur de forme referentiel. Nous utiliserons pour cette comparaison la formule de corrélation normalisée (NCC) pour évaluer la similarité des deux vecteurs (Vpre et Vref) comme suit :

$$NCC(V_{Pre}) = \frac{\sum_i V_{Pre}(Z_i) \times V_{ref}(Z_i)}{\sqrt{\sum_i V_{Pre}(Z_i)^2 \times V_{ref}(Z_i)^2}} \qquad \text{i :1…N}$$

Le vecteur de paramètres Epre correspondant à la nouvelle position de la cible sera mémorisé pour l'affichage du curseur ainsi que la réinitialisation de la matrice de forme et la matrice des paramètres du filet d'ellipses couvrant la nouvelle zone de recherche.

L'objectif de la différence d'images est de réduire le temps de calcul de la nouvelle position. En effet, la recherche est limitée aux zones de l'image où quelque chose a bougé (pixels en mouvement).

5.3.3.3. Repositionner la zone de recherche

Une fois la nouvelle position de la cible déterminée, la zone de recherche sera repositionnée en assurant toujours que la cible se trouve au niveau de son centre. Pour

cela, il faut réinitialiser les deux matrices du filet d'ellipses: vecteurs de forme et paramètres.

Cette réinitialisation est basée sur des calculs géométriques (Fig. V.7 et V.8) sachant que les mouvements possibles entre l'ancienne et la nouvelle position de la cible sont :

- Soit une rotation d'angle S (résultat obtenu dans le vecteur de paramètres correctif Ec) par rapport au centre M (i, j) de l'ellipse contenant la dernière position de la cible).

- Soit une translation de rayon MM' (M'(i', j') étant le nouveau centre de l'ellipse contenant la cible)

- Soit les deux en même temps.

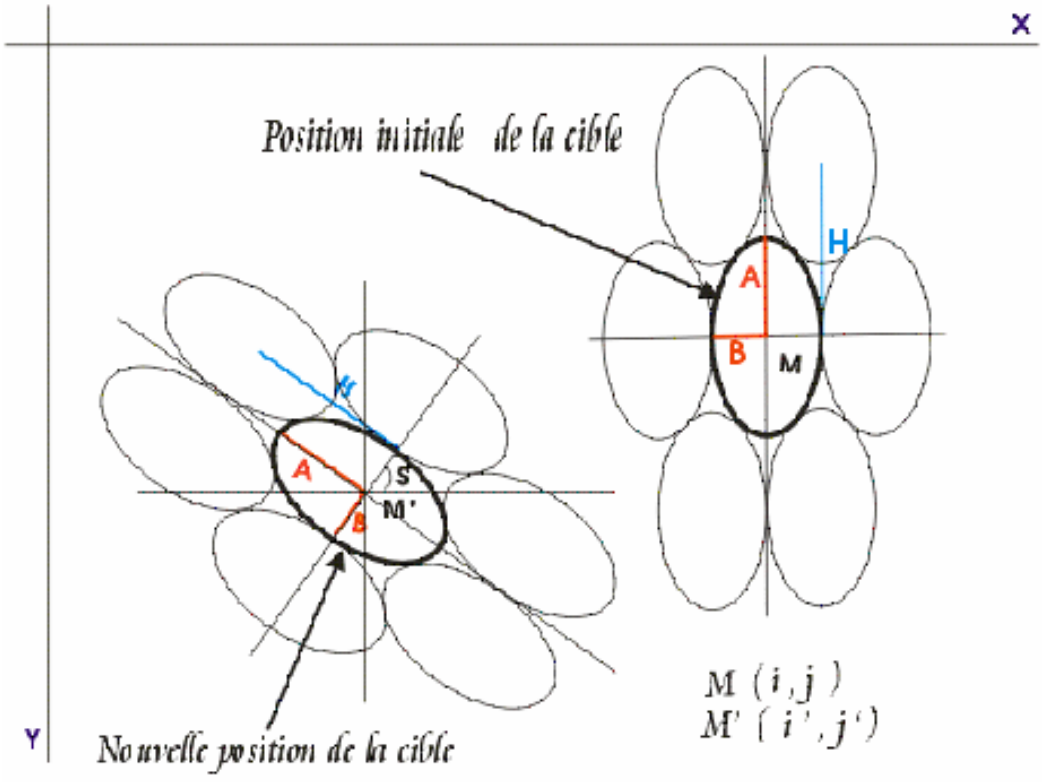
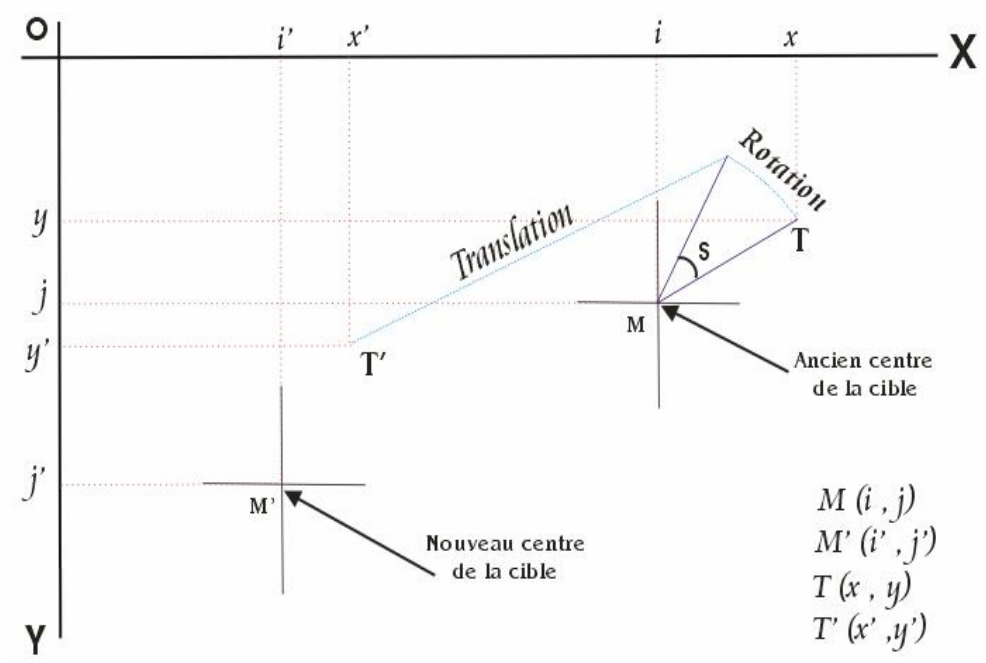


Figure V.7 - Schéma représentatif de la nouvelle zone de recherche suivant les nouveaux paramètres de la position de la cible.



$$x' = (x - i)\cos(s) - (j - y)\sin(s) + i'$$
$$y' = j' - (x - i)\sin(s) - (j - y)\cos(s)$$

Figure V.8- Les calculs géométriques nécessaires pour la réinitialisation des matrices caractérisant la zone de recherche.

5.4. Conclusion

Nous avons présenté, dans ce chapitre, notre première contribution dans le domaine de suivi d'un motif plan dans une application de réalité augmentée. En effet, notre contribution permet une flexibilité dans l'utilisation du tableau magique. L'utilisateur de ce système aura une plus grande liberté à faire bouger son bras en faisant des mouvements de rotation, chose qu'il n'avait pas le droit de faire auparavant. Cette méthode de suivi que nous avons proposée s'apprête bien pour une utilisation temps réel puisqu'elle ne demande pas beaucoup de temps de calcul. Néanmoins, elle nécessite comme la plupart des méthodes basées pixels une étape offline d'initialisation qui demande un temps de calcul plus lent.

Cette première contribution, nous a ouvert le chemin pour définir une extension de cette méthode pour qu'elle soit générale et qu'on peut appliquer au suivi d'un motif plan dans n'importe qu'elle séquence d'images vidéo. Cette méthode sera détaillée au chapitre suivant.

CHAPITRE 6

Approche proposée pour le suivi d'un motif plan : Application au processus d'augmentation d'une séquence vidéo réelle.

Le suivi d'objets constitue une phase essentielle dans le processus d'augmentation d'une séquence vidéo. L'augmentation d'une vidéo consiste à incruster un objet n'appartenant pas à la scène réelle filmée. Cette augmentation n'est possible que si cet objet virtuel est inséré dans chacune des images de la vidéo. Le processus d'augmentation commence alors par préciser la position dans laquelle l'objet virtuel va être inséré dans la première image. Le suivi de cette position dans les images suivantes de la vidéo, phase que nous avons qualifié auparavant d'essentielle, viens par la suite. A chaque fois que la position est déterminée dans une image, l'objet virtuel est inséré dans cet endroit précis de l'image.

Dans ce chapitre, nous allons présenter une approche de suivi d'un motif plan à base d'un réseau de neurones artificiel (RNA). Cette approche s'inscrit parmi les approches 2D basées sur le suivi de régions. Ainsi, elle prend en considération tous les pixels de la région où se trouve le motif à suivre. Elle permet de suivre en temps réel un motif vu le coût très réduit du temps de calcul qu'elle offre.

6.1 Notations utilisées

Soit I(p) la valeur du niveau de gris d'un point p se trouvant dans l'image I à la position donnée par les coordonnées (x,y) et soit Rref = (X_1, X_2, ..., X_n) l'ensemble des n positions des points Xi (i=1 à n) représentant la région contenant le motif à suivre (Fig. VI.1). Ainsi, I(Rref) = (I(X_1), I(X_2),...., I(X_n)) représente le vecteur intensité lumineuse de la région cible contenant le motif recherché.

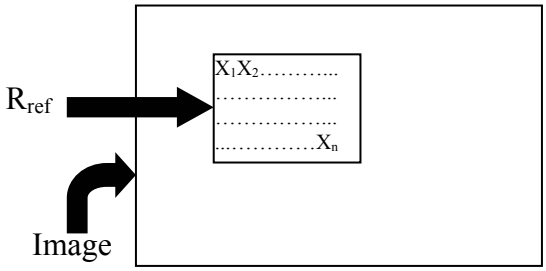


Figure VI.1. La région de réference contenant le motif à suivre dans la 1ére image de la vidéo.

Notons par μ le vecteur de déplacement d'une région. Les perturbations (déformations) de la région Rref (Fig. VI.2) seront obtenues en appliquant N déplacements (μ1, μ2, ..., μn). Un déplacement, dans notre cas, peut être une translation selon l'axe des X, une translation selon l'axe des Y, une translation selon les deux axes X et Y ou une rotation. Pour chaque déplacement μi, une différence δIi est faite entre le vecteur I(Rref) et le vecteur I(Rcur) (avec Rcur = (X'1, X'2, ..., X'n)). I(Rref) représente le vecteur niveau de gris de la région cible et I(Rcur) represente le vecteur niveau de gris de la région courante obtenue après un déplacement μi.

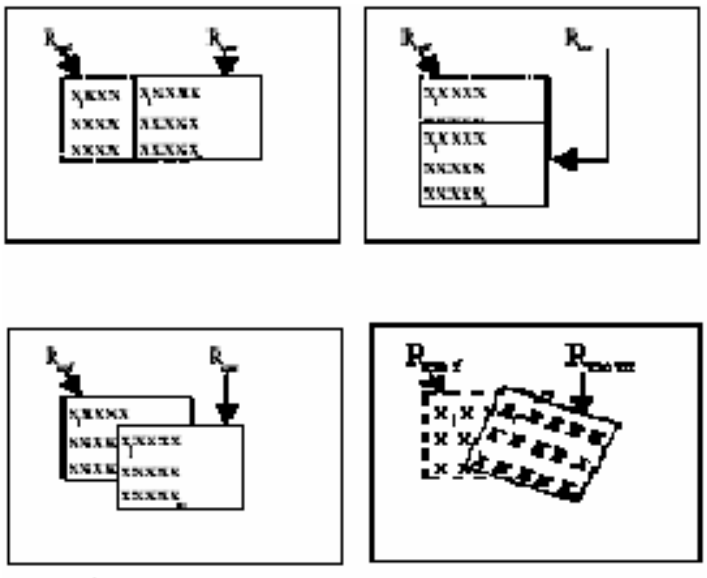


Figure VI.2. Les perturbations effectuées

Ainsi, nous voulons estimer, dans cette approche, la relation entre les n déplacements (μ1, μ2, ..., μn) et les n différences (δI1, δI2, ..., δIn) effectuées après chaque déplacement μi.

6.2 Vue générale de l'approche proposée

Notre approche s'articule autour de trois phases (Fig. VI.3). La première phase est une phase offline qui permet de faire un apprentissage à l'aide d'un réseau de neurones sur la relation entre les déplacements possibles de la région contenant le motif cible et les changements de l'intensité lumineuse engendrés par ces déplacements. Cette phase est coûteuse en temps de calcul. La seconde et la troisième phases sont des phases online, itératives et liées. La phase de suivi consiste à déterminer, pour chaque image de la vidéo, la position où l'objet virtuel doit être inséré. Enfin, la phase d'augmentation prend, pour chaque image de la vidéo, la position calculée lors de la phase précédente pour incruster l'objet virtuel.

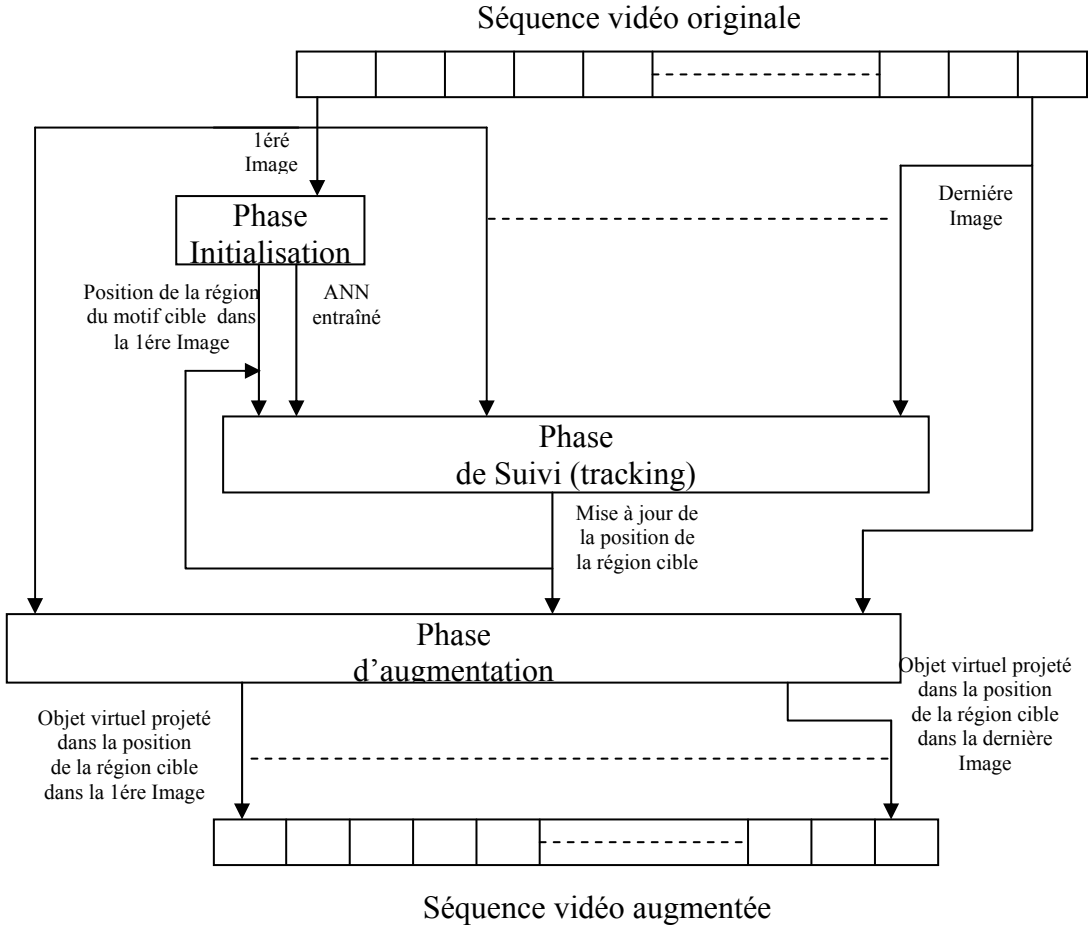


Figure VI.3. Articulation des phases

6.3 Phase d'initialisation

Après la sélection de la position d'insertion de l'objet virtuel dans la première image de la séquence vidéo, cette phase procède à l'entraînement d'un réseau de neurones sur deux ensembles donnés l'un représentant ses entrées et l'autre représentant ses sorties. Nous représentons cette position par le quadruplet [X_{LL}, Y_{LL}, W, H], où le couple (X_{LL}, Y_{LL}) désigne les coordonnées du coin bas à gauche de la région contenant le motif à suivre Rref et le couple (W, H) désigne respectivement sa longueur et sa largeur. En effet, pour procéder à cet entraînement il est nécessaire de savoir quel est le type de réseau de neurones que nous devons utiliser et quelle est l'information disponible pour entraîner ce réseau.

6.3.1 Information disponible pour le RNA

Nous disposons dans notre cas de l'ensemble des perturbations Y = (μ_1, μ_2, ..., μ_n) appliquées à Rref. Ces perturbations représentent la sortie de notre RNA. Aussi, nous pouvons obtenir les différences de niveaux de gris entre l'image de référence Rref et les images obtenues après avoir effectué une perturbation μ_i. L'ensemble de ces différences noté H = (δI_1, δI_2, ..., δI_n) représente les entrée de notre RNA. Chaque perturbations μ_i, est représentée par un vecteur de trois éléments [dX, dY, θ]. dX est un déplacement selon l'axe des X, dY est un déplacement selon l'axe des Y et θ est l'angle de rotation. Il est nécessaire à ce niveau de discuter du type et du nombre de perturbations à appliquer à Rref choisie dans la première image de la séquence vidéo. Nous avons choisi trois types de perturbations:

- Des translations horizontales et verticales ne dépassant pas les 15% respectivement de la largeur et la longueur de Rref. En d'autres termes, si [X_{LL}, Y_{LL}, W, H] est la position de Rref, alors [X_{LL}+dx, Y_{LL}, W, H] et [X_{LL}, Y_{LL}+dy, W, H] constituent respectivement les translations horizontales et verticales avec max(|dx|) = 15% de W et max(|dy|) = 15% de H. Nous nous sommes limité à ce pourcentage, pour constituer l'ensemble Y et H nécessaire pour l'entraînement du RNA, parce qu'il y a un léger changement entre les images successives d'une séquence vidéo.
- Des rotations avec un angle de (+/-) 35°.

- Des combinaisons de translations horizontales, de translations vertical et de rotations.

Pour ce qui est du nombre de perturbations, 1500 est un bon compromis entre la précision du résultat de la recherche et le temps d'exécution. Ce nombre peut être révisé à la hausse pour une meilleure précision ou la baisse pour une exécution plus rapide.

6.3.2 Choix et Entraînement du RNA

Avant de présenter notre choix du type et de la méthode d'entraînement de notre RNA, il est utile d'introduire ce concept.

6.3.2.1 Brève description des RNA

Un RNA est un groupe de neurones artificiels interconnectés qui utilisent un modèle mathématique pour traiter une information. Ce modèle est basé sur une approche connexionniste de calcul.

a- Neurone artificiel

Le neurone artificiel est aussi appelé nœud. Il constitue l'unité de base d'un RNA. Il simule la fonction d'un neurone biologique. Mathématiquement, il est défini comme une fonction allant d'un ensemble à n dimensions vers un ensemble à une dimension. Il reçoit alors une ou plusieurs entrées pour produire une seule sortie après un certain traitement. Ce traitement consiste d'abord à réaliser une somme pondérée pour la soumettre après comme entrée d'une fonction non linéaire donnée dite fonction d'activation ou de transfert (fig. VI.4). Cette fonction prend, dans la plupart des cas, la forme d'une fonction sigmoïde mais peut prendre la forme d'autres fonctions non linéaires. Généralement, la fonction transfert est monotone croissante.

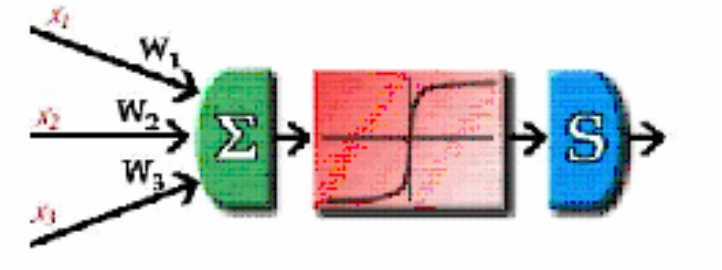


Figure VI.4 Un neurone artificiel avec trois entrées

Avec :

- (X1, X2, X3) : les entrées du neurone.
- (W1, W2, W3) : les poids associés aux entrées.
- Σ : la somme pondérée
- f : la fonction de transfert ou d'activation
- S : la sortie.

b- Réseau de Neurones Artificiel (RNA)

Un RNA peut être vu comme un graphe direct pondéré. Les neurones artificiels forment les noeuds de ce graphe et les liaisons entre les entrées et les sorties des neurones les connexions entre les nœuds de ce graphe. Etant basés sur un modèle de connexion, les RNA peuvent être groupés en deux catégories [48]:

- Les RNA Feed-forward, leur graphe ne présente pas de boucles. Cette catégorie regroupe trois familles : Les perceptrons monocouche, les perceptrons multicouches et les réseaux à fonctions radiales.
- Les RNA Recurrent (or feedback), leur graphe admet des boucles. Cette catégorie regroupe quatre familles : les réseaux compétitives, les réseaux de Hopfield, les réseaux de Kohonen et les modèles ART.

L'habilité d'apprendre est un signe fondamental de l'intelligence. Quoiqu'il est difficile de formuler d'une manière précise l'apprentissage, nous pouvons le ramener dans les RNA à un problème de mise à jour de leurs architectures et des poids des connexions. Ainsi, un RNA peut traiter d'une manière efficace une tâche spécifique. Le RNA doit souvent actualiser la valeur des poids de connexion à partir du jeu disponible de données d'entraînement. La performance des résultats donnés par le RNA s'accroît au fur et à mesure par une mise à jour itérative des poids des connexions.

L'habilité des RNA d'apprendre automatiquement à partir des exemples les rendent attractives. Pour comprendre ou concevoir un processus d'apprentissage pour un RNA, il est nécessaire d'avoir un modèle de l'environnement dans lequel il évolue. En d'autres termes, il faut savoir : quelle est l'information disponible pour entraîner le

RNA et comment il procède pour mettre à jour les poids des connexions. Un algorithme d'apprentissage utilise des règles d'apprentissage pour mettre à jour ces poids.

Deux paradigmes d'apprentissage existent: supervisé et non supervisé. Dans l'apprentissage supervisé, ou apprentissage en présence d'un enseignant, Le RNA dispose d'une réponse (sortie) correcte pour chaque modèle d'entrées dans l'ensemble des données d'entraînement. Les poids des connexions sont déterminés pour permettre au RNA de produire des réponses aussi proches que possible des réponses correctes disponibles, tandis que, l'apprentissage non supervisé, ou l'apprentissage sans enseignant, ne demande pas d'avoir une réponse correcte associée aux modèles des entrées dans l'ensemble des données d'entraînement. Le RNA, dans ce cas, explore la structure des données d'entraînement, ou mesure les corrélations entre les modèles relatifs à ces données et les regroupent en catégories.

Dans la théorie de l'apprentissage par l'exemple, trois points fondamentaux doivent être discutés: la capacité, la complexité des exemples et la complexité de calcul. La capacité nous renseigne sur le nombre de modèles pouvant être stockés. La complexité des exemples nous informe sur le nombre de modèles d'entraînement nécessaires pour mener un bon apprentissage du RNA.

La complexité de calcul concerne le temps nécessaire à un algorithme d'apprentissage pour estimer une solution à partir des modèles d'entraînement. Plusieurs algorithmes existants ont une complexité de calcul élevée. Ainsi, concevoir des algorithmes avec de faible temps de calcul reste toujours d'actualité et fait l'objet de plusieurs recherches en cours. Il existe quatre types de règles d'apprentissage: correction d'erreur, Boltzmann, Hebbian et apprentissage compétitif.

Chaque algorithme d'apprentissage est conçu pour l'apprentissage dans un RNA ayant une architecture spécifique. Il permet de bien traiter un ensemble réduit de tâches [48] (Fig.VI.5).

Paradigme	Règle d'apprentissage	Architecture	Algorithme apprentissage	Tache
Supervisé	Correction d'erreur	Perceptron monocouche ou Multicouche	- Perceptron, -Back-propagation -Adaline, -Madaline	- Classification, - Approximation de fonction, - Prédiction,
	Boltzmann	Récurrente	Algorithme d'apprentissage de Boltzmann	Classification
	Hebbian	Feed-Forward Multicouches	Analyse Linéaire Discriminante	- Analyse de données, - Classification
	Compétitive	Compétitive	Compétitive	Compression de données.
		ART	ARTMap	Classification
Non supervise	Correction d'erreur	Feed-Forward Multicouche	Projection de Sammon	Analyse de données
	Hebbian	Feed-Forward ou compétitive	Analyse en Composante principale	- Analyse de données - Compression de données
		Réseau de Hopfield	Apprentissage avec Mémoire Associative	Mémoire Associative
	Compétitive	Compétitive	Compétitive	- Catégorisation, - Analyse de données
	Somme de Kohonen	Somme de Kohonen	Somme de Kohonen	- Catégorisation, - Analyse de données

Figure VI.5 Les algorithmes d'apprentissage les plus connus

6.3.2.2 Justification de nos choix

Nous avons pu constituer en §6.1 un jeu de données d'apprentissage. Ce jeu est formé d'une part des perturbations μi effectuées sur la région Rref. L'ensemble des μi constitue la sortie de notre RNA. D'autre part des différences des niveaux de gris δIi entre l'image Rref et les images Rcur sont obtenues après avoir effectué une perturbation μi. Ces différences constituent les entrées de notre RNA.

Sachant que nous avons la différence correcte δIi (entrée du RNA) pour chaque perturbation μi (sortie du RNA), la généralisation de se résultat à n'importe quelle différence δIi revient à résoudre un problème d'approximation de fonction. Alors,

l'apprentissage supervisé est le seul valide dans ce cas. Ainsi, nous avons choisi pour notre RNA une architecture de perceptron multicouches avec son algorithme d'apprentissage. D'après les expérimentations que nous avons menées, nous avons choisi pour les paramètres du RNA les valeurs suivantes :

- Le nombre de couches pour le RNA est de 2 (1 couche cachée et 1 couche de sortie).
- Le nombre de neurones par couche est de 200 pour la couche cachée et 3 pour la couche de sortie.
- La fonction d'activation (une fonction sigmoïde pour les neurones de la couche cachée et une fonction linéaire pour les neurones de la couche de sortie).
- La fonction d'entraînement est une fonction trainscg.
- L'erreur tolérée ou but est de 10^{-5}.

Pour plus de performance les données d'entraînement H et Y doivent être normalisées. Ainsi à ce niveau l'entraînement de notre RNA peut commencer. Il suit les étapes suivantes:

- Initialisation et seuillage des poids de connexion à de petites valeurs aléatoires.
- Présentation du vecteur H comme entrée du RNA et évaluation du résultat noté D en sortie.
- Mettre à jour les poids en minimisant la différence entre la sortie évaluée D et la sortie souhaitée Y.

6.4 Phase de suivi (Tracking)

Dans cette phase notre RNA est déjà entraîné et prêt à être utilisé pour le suivi du motif cible dans la suite des images de la séquence vidéo. Cette phase est itérative et s'exécute selon le schéma suivant :

Etant donné NB le nombre d'images dans notre séquence vidéo et i l'indice de l'image courante de cette séquence.

Pour I allant de 2 jusqu'à NB faire

1- Initialiser la position prédite de la région courante Rcur dans l'image i à la même position de la région de référence Rref dans l'image i-1.

2- Faire une différence des niveaux de gris entre la région prédite et la région cible. Ainsi, $\delta I = I(Rcur) - I(Rref) = (I(X_1), I(X_2),...., I(X_n))^t - (I(X'_1), I(X'_2),...., I(X'_n))^t$.
3- Présenter δI comme entrée au RNA entraîné. Nous obtenons alors le vecteur correctif [dX, dY, θ] comme sortie.
4- Mettre à jour la position de la région prédite en utilisant le vecteur correctif.
5- Prendre cette nouvelle position de la région Rcur comme nouvelle position de la région de référence Rref pour la prochaine itération (prochaine image de la séquence vidéo).

Finalement, la position de la région contenant le motif recherché est déterminée dans toutes les images de la séquence vidéo.

6.5 Phase d'augmentation

L'objet virtuel est projeté, en appliquant une homographie planaire [64, 65], dans la position déterminée dans chaque image de la séquence vidéo lors de la phase de suivi.

Puisque la région de référence sélectionnée, en première image, pour contenir l'objet virtuel, est planaire, nous pouvons suppose que la composante Z=0. Ainsi, nous pouvons associé l'image de notre objet virtuel à la région détermine dans chaque image de la séquence vidéo en appliquant une projection 2D-2D. En d'autres termes, si x = (x, y, 1) représente les coordonnées homogènes d'un point dans l'espace image de l'objet virtuel et x´ = (x', y', 1) les coordonnées correspondantes du même point dans l'espace image de la séquence vidéo, $x \leftrightarrow x'$ définit une correspondance donnée par :

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = H \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \qquad (1)$$

Où H représente une matrice 3x3 appelée homographie planaire.

Ainsi, nous pouvons réécrire la relation (1) comme suit:

$$\begin{aligned} x'(h_{31}x + h_{32}y + h_{33}) &= h_{11}x + h_{12}y + h_{13} \\ y'(h_{31}x + h_{32}y + h_{33}) &= h_{21}x + h_{22}y + h_{23} \end{aligned} \tag{2}$$

Où hij est un élément de la matrice H placé en ligne i et en colonne j.

Avec quatre correspondances de points (correspondances entre les coins de la région suivi et les coins de l'image de l'objet virtuel), nous pouvons obtenir les éléments de la matrice H en résolvant le système de huit équations linéaires donné ci-dessous :

$$\begin{pmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1'x_1 & -x_1'y_1 & -x_1' \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -y_1'x_1 & -y_1'y_1 & -y_1' \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2'x_2 & -x_2'y_2 & -x_2' \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -y_2'x_2 & -y_2'y_2 & -y_2' \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -x_3'x_3 & -x_3'y_3 & -x_3' \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -y_3'x_3 & -y_3'y_3 & -y_3' \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -x_4'x_4 & -x_4'y_4 & -x_4' \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -y_4'x_4 & -y_4'y_4 & -y_4' \end{pmatrix} \mathbf{h} = \mathbf{0} \tag{3}$$

Où h est un vecteur à neufs éléments contenant les éléments hij. $h = [h_{11}\ h_{12}\ h_{13}\ h_{21}\ h_{22}\ h_{23}\ h_{31}\ h_{32}\ h_{33}]^T$

Puisque la forme matricielle de cette équation est donnée par Ah=0, la solution est l'espace nul de A. La résolution peut alors être faite en utilisant la méthode de la décomposition en valeurs singulières [79, 87].

6.6 Résultats Expérimentaux

Notre algorithme a été appliqué à plusieurs séquences vidéo. Nous allons présenter quelques une de ces applications. Les séquences utilisées ont toutes les caractéristiques suivantes: 375 images de 320 x 240 pixels chacune. L'objet suivi est bordé d'un rectangle bleu.

Dans le premier exemple (Fig. VI.6), nous avons essayé de suivre un objet fixe. Le RNA est entraîné avec les valeurs mentionnées en §6.3. L'algorithme a été efficace dans ce cas.

Image 1

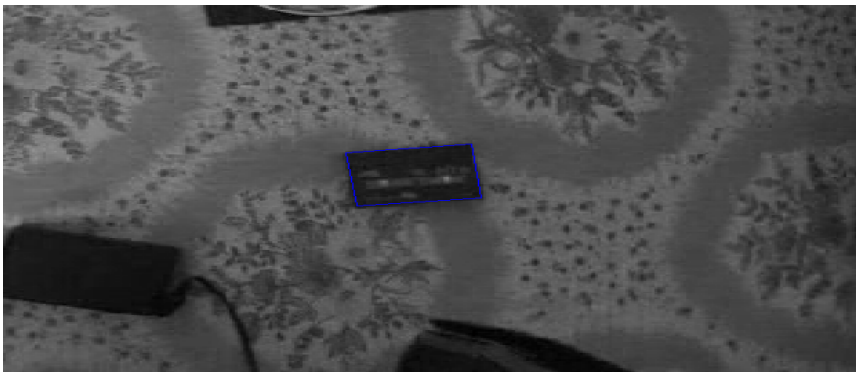


Image 36

Image 215

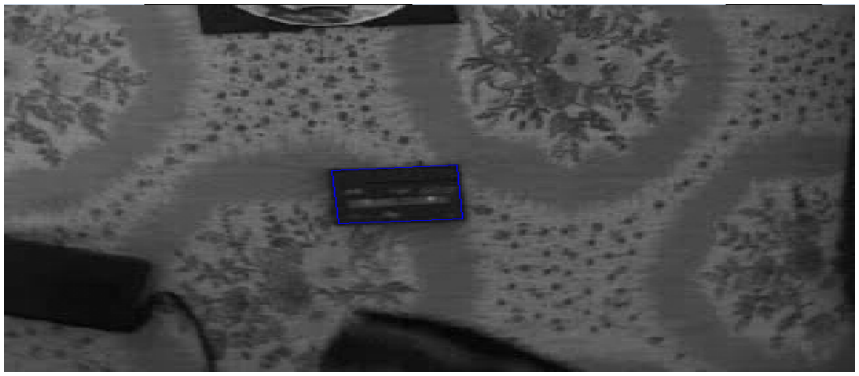


Image 315

Figure VI.6- Suivi d'un objet fixe

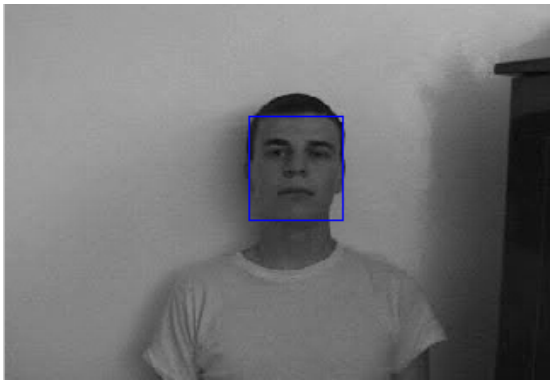


Image 1

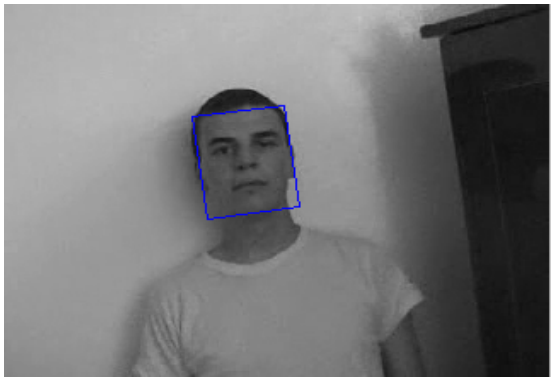


Image 60

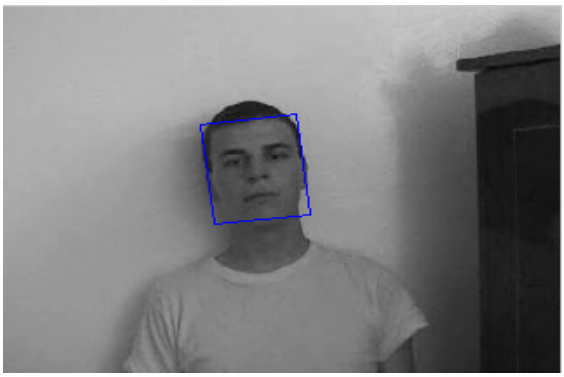


Image 160

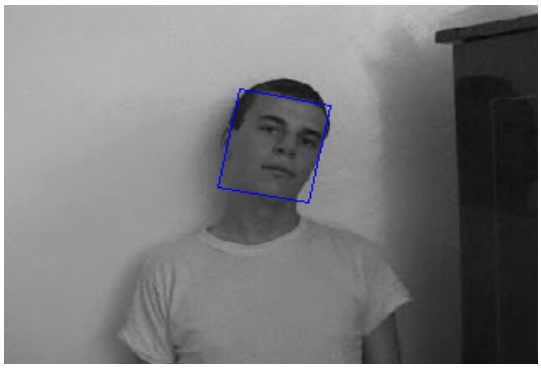


Image 300

Figure VI.7- Suivi d'un visage humain en mouvement

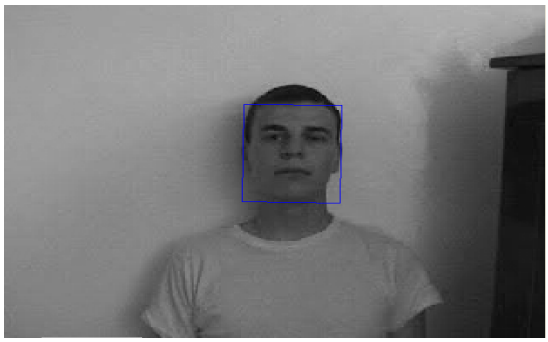


Image 1

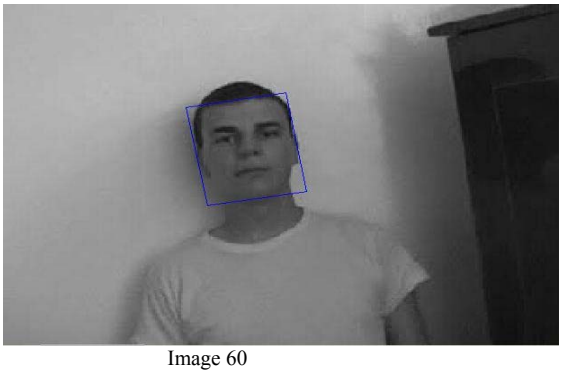


Image 60

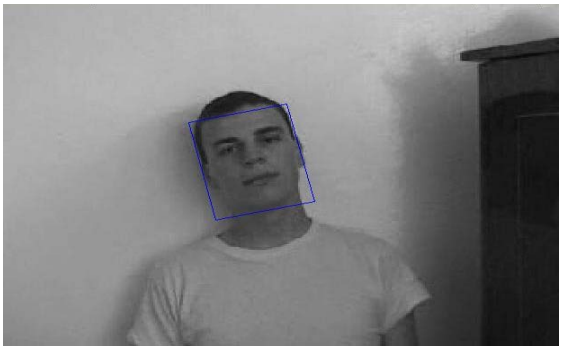


Image 180

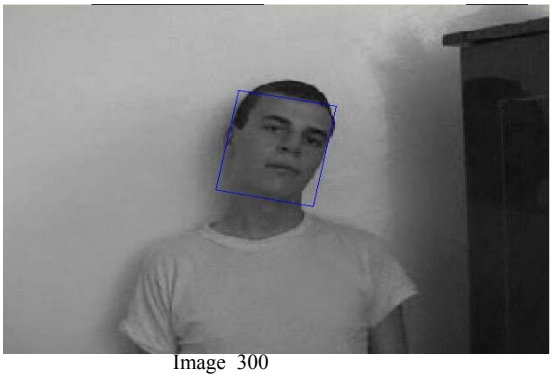


Image 300

Figure VI.8- Même séquence que la Fig. VI.7 avec modification des valeurs pour entraîner le RNA

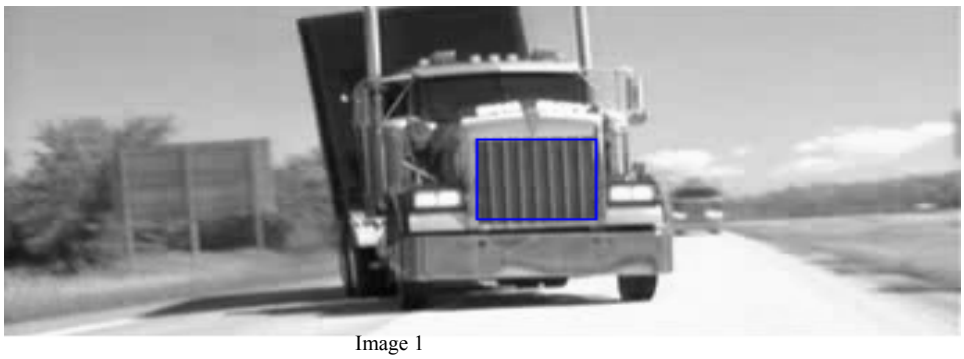


Image 1

Image 30

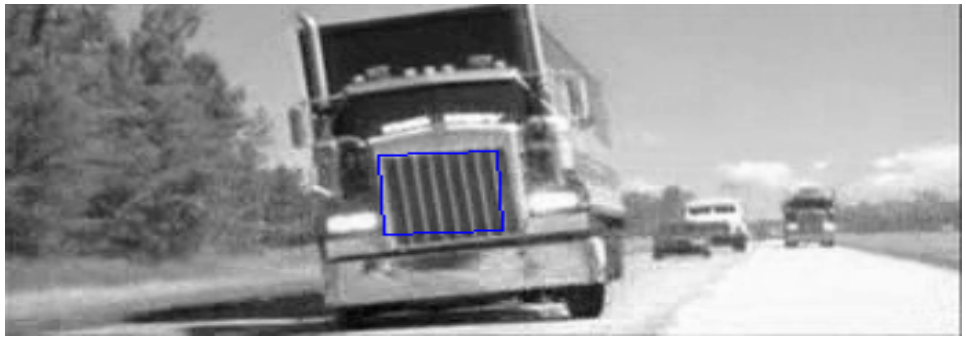


Image 80

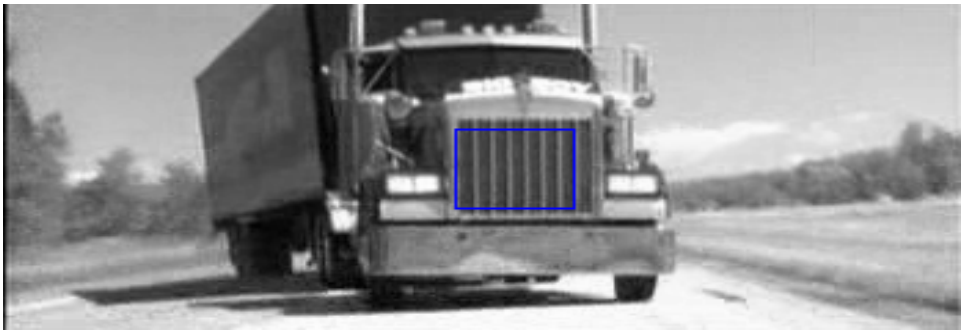


Image 105

Figure VI.9. Suivi d'un objet très dynamique

Dans le second exemple, nous avons essayé de suivre le visage d'une personne en mouvement. Les mêmes valeurs sont utilisées pour entraîner le RNA. L'algorithme est toujours efficace (Fig. VI.7).

En gardant la même séquence de l'exemple précèdent mais en changeant certaines valeurs pour entraîner le RNA (2000 perturbations, 2 couches cachées, 300 neurones pour chaque couche cachée et un but de 10^{-7}). L'algorithme donne les mêmes résultats qu'avec les anciennes valeurs (Fig. VI.8).

Dans le dernier exemple, nous avons suivi la calendre d'un camion effectuant des manœuvres brusques. Les mêmes valeurs, du premier exemple, sont utilisées pour entraîner le RNA. Les résultats de l'algorithme reste acceptables (Fig. VI.9).

6.7 Conclusion

Nous avons développé un algorithme de suivi efficace. Pour cela, nous avons utilisé un RNA dans une phase offline. Le RNA permet de déterminer la relation entre la variation du niveau de gris relative à un déplacement donné. Ainsi, Le RNA une fois entraîné devient capable de nous fournir rapidement la correction à apporter à la position prédite pour trouver la position exacte de l'objet cible. D'où la possibilité d'utiliser cet algorithme dans une application RA temps réel. L'objet suivi peut alors être remplacé par un objet virtuel en appliquant une homographie planaire.

CHAPITRE 7

Bilan & Perspectives

Le domaine de la réalité augmentée constitue le croisement de plusieurs domaines, dont la réalité virtuelle, la vision artificielle, l'infographie, les mathématiques et l'optique sont les piliers. C'est ce qui fait que plusieurs de ses résultats dépendent de ceux obtenus dans d'autres domaines, en y apportant des adaptations. Il existe même des axes repris de manière adaptative.

Afin que la recherche progresse dans ce domaine, plusieurs pivots sont traités. Pour satisfaire les augmentations en temps réel, les recherches sont axées sur la minimisation du temps de tracking et des retards de numérisation et de génération des objets virtuels ainsi que l'exploitation des systèmes temps réels. Pour simplifier les systèmes de réalité augmentée, des recherches dans le sens de la réduction des contraintes de calibration des caméras sont menées. En plus, la structure des interfaces nécessaires pour l'interaction, pouvant supporter ces nouveaux systèmes, est un facteur important de réussite. Rajoutées à celles-ci, des études psychophysiques sont nécessaires. Elles pourront étudier le niveau de détection des erreurs et de perception

humaine, la tolérance aux erreurs d'alignement par les différentes applications, les effets des HMD après enlèvement, etc.

Notons qu'il reste toujours certaines limites insurmontables. A titre d'exemple, si l'on considère une rotation de la tête (ou de la caméra) de 50° en 1 seconde, cela exprime le fait que le système doit générer des images virtuelles avec un retard de 10ms ou moins afin de satisfaire une erreur de 0,5°. Cela exprime la génération d'une image chaque 10ms. Dans le cours des choses, le simple affichage d'une image sur un écran à 60 Hz nécessite 16.67 ms. En d'autres termes, il n'est pas possible de satisfaire une augmentation qui atteigne une précision de 0,5°.

En effet, un bon système de RA est un système qui permet de garder à tout moment un alignement correct entre les objets virtuels et les objets réels de la scène. Pour chaque mouvement accompli par un utilisateur du système les objets virtuels doivent suivre la position et l'orientation des objets réels de la scène. Cela est possible, grâce à un suivi efficace des objets réels dans chaque point de vue. Le suivi d'objet constitue, alors une phase importante dans le processus d'augmentation. Ainsi, nous nous sommes intéressés au suivi d'objets (tracking) puisqu'en plus de son importance, il représente la partie la plus consommatrice de temps dans une application de réalité augmentée.

Nous avons amélioré, dans un premier temps, la fonction de suivi du doigt de l'utilisateur dans le tableau magique. En effet, la méthode utilisée jusqu'ici est le suivi par corrélation. Cette méthode n'est applicable qu'aux translations effectuées dans un plan parallèle à l'image. Le but de l'approche proposée est de doter la cible (doigt de l'utilisateur) d'une meilleure flexibilité dans le mouvement (translation + rotation) en temps réel.

Nous avons, ensuite, développé une méthode de suivi d'objets complexes. Cette méthode passe par deux étapes. L'étape offline consiste à faire un apprentissage sur le mouvement du motif à suivre. Cet apprentissage a été réalisé à l'aide des réseaux de neurones. L'étape online se base sur les résultats de l'étape précédente pour déterminer la position de l'objet cible. Cette méthode a été, enfin, intégrée dans un processus d'augmentation de séquences vidéo.

Nous nous intéressons actuellement à l'extension de notre méthode pour incruster des objets 3D complexes. Pour cela, nous utilisons la synthèse de vue pour représenter l'objet virtuel à insérer dans la séquence réelle vue dans [33]. Le but est de pouvoir remplacer la sphère de vue [18] ayant un nombre très important de vues par un nombre très restreint de vues pouvant générer toutes les vues possibles.

Bibliographie

[1] Annedouche, S. Loup, B. et Prodhomme, M. "*Le Tableau Magique*", rapport de projet de 3ème année de l'École Nationale Supérieure en Informatique et Mathématiques Appliquées de Grenoble (ENSIMAG), juin 1999.

[2] Azuma R., "The Challenge of Making Augmented Reality Work Outdoors", in Mixed Reality: Merging Real and Virtual Works, Springer-Verlag, 1999, Chp 21 pp. 379-390, ISBN 3-540-65623-5.

[3] Azuma R., "Tracking Requirements for Augmented Reality". Communications of the ACM, juillet 1993, pp. 50-51.

[4] Azuma R., "A Survey of Augmented Reality," Presence: Teleoperators and Virtual Environments. vol. 6, no. 4, Aug. 1997, pp. 355-385.

[5] Azuma R., "Recent advances in Augmented Reality", 0272-1716/01/ IEEE Nov-dec 2001.

[6] Azuma R., YOU S and Neumann U., "Orientation Tracking For Outdoor Augmented Reality Registration", IEEE Computer Graphics and Applications, Nov/Dec 1999.

[7] Baumberg A.M., Hogg D.C., "An efficient method for contour tracking using active shape models", 1994

[8] Bencheikh El-Hocine M., Bouzenada M. and Batouche M.C., 2004, "A new method of finger tracking applied to the magic board", Proceedings of IEEE ICIT 04, pp 1046-1051, Tunisia, 2004.

[9] Bérard F., J. Coutaz et J.L. Crowley* " Le tableau Magique : un outil pour l'activité de réflexion " Actes de la conférence ErgoIHM'2000 (3-6 octobre 2000, Biarritz, France), pp. 33-40.

[10] Bérard, F. (1994) *Vision par Ordinateur pour la Realité Augmentée : Application au Bureau Numérique*, Rapport de DEA en Informatique (Master's thesis), Université Joseph Fourier, Grenoble, France

[11] Bérard, F. (2001) Augmentation d'un tableau blanc par des techniques de Vision par Ordinateur, Vidéo présentée à la journée AFIHM de la conférence ASTI'2001.

[12] Bérard, F. (2003) The Magic Table: Computer-Vision Based Augmentation of a Whiteboard for Creative Meetings, in IEEE workshop on Projector-Camera Systems (PROCAM), Nice, France.

[13] Bérard, F. “Vision par ordinateur pour l’interaction homme-machine fortement couplée”, Thèse de doctorat de l’Université Joseph Fourier, Grenoble, 1999.

[14] Bérard, F. Coutaz, J. and Crowley, J.L. (1994) Suivi du doigt en Vision par Ordinateur : Application au Bureau Numérique, actes des 6ème journées de l'Ingénierie des Interfaces Homme-Machine, Lille, France.

[15] Black M.J. and Jepson A.D., 1998, “Eigen tracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation”, Int'l J. Computer Vision, vol. 26, no. 1, pp. 63-84.

[16] Boukir S., Bouthemy P., Chaumette F., Juvin D., "A local method for contour matching and its parallel implementation", Machine Vision and Application, 10(5/6):321-330, avril 1998.

[17] Bouzenada M., M.C. Batouche and Z. Telli. Neural network for object tracking. Information Technology Journal, ISSN 1812-5638, 2007, pp 526-533.

[18] Bouzenada M., B. Nini, M. Berkane, « Proposition d’une méthode de reconnaissance d’objets basée sur les graphes d’aspect », VVATI’03, JIJEL.

[19] Bouzenada M., M.C. Batouche, "An Object Tracker for Markerless Augmented Reality", IPCV’07, 25-28 Juin 2007, Las Vegas, USA, pp 584-589.

[20] Bouzenada M., M.C. Batouche, S. Boussemroun, H. Boushaba, R. Filali, « Une nouvelle méthode de suivi basée sur les réseaux de neurones. », Dans les actes de la Conférence Internationale sur la Productique (CIP 2005), Tlemcen, Algérie, December 2005.

[21] Brunelli R. and Poggio T., 1995, “Template Matching: Matched Spatial Filter and Beyond”, A.I. Memo 1549, Massachusetts Inst. of Technology.

[22] Canny J.F., "A Computational approach to edge detection", IEEE trans.Pat. Anal. and Mach. Intel., Vol 8(6), pp.34-43, 1986.

[23] Collins R.T. and Liu Y., “On-line selection of discriminative tracking features”. IEEE Trans. PAMI, vol. 27, no. 10, pp. 1631-1643, 2005.

[24] Comaniciu D., Ramesh V. and Meer P., “Kernel-based object tracking”. IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI), vol. 25, 564–577, 2003.

[25] Comport Andrew I., Éric Marchand, François Chaumette. A real-time tracker for markerless augmented reality. Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR ’03), 2003.

[26] Cootes T.F., Edwards G.J. and Taylor C.J., 1998, “Active Appearance Models”, Proc. European Conf. Computer Vision, pp. 484-498.

[27] Coutaz, J. “*Interfaces Homme-Machine : le Futur ne Manque pas d'Avenir*”, Conférence invitée, Proc. ERGO-IA'98, Ed. ESTIA/ILS, p. 43- 55, 1998.

[28] Coutaz, J. Lachenal, C. Bérard, F. Barralon, N. (2002) *Quand les surfaces deviennent interactives*, (in french), in Les cahiers du numérique, Lavoisier, Vol. 3, Numéro 4-2002, pp.101-126.

[29] Crowley, J.L. Bérard, F. and Coutaz, J. (1995) *Finger Tracking as an Input Device for Augmented Reality*, in Procs. of International Workshop on Automatic Face and Gesture Recognition, Zurich, Switzerland.

[30] Darrell T., Essa I.A. and Pentland A.P., 1996, “Task-Specific Gesture Analysis in Real-Time Using Interpolated Views” IEEE Trans. PAMI, vol. 18, no. 12, pp. 1236-1242.

[31] Deguchi K., "A Direct Interpretation of Dynamic Images with Camera and Object Motions for Guided Robot Control", Int. Journal of Computer Vision, 37(1):7-20, Juin 2000.

[32] Deriche R. and Faugeras O., "Tracking Line Segments", Image Vision Computation, Vol. 8, 1990, pp 261-270.

[33] Dib A., M.Bouzenada, M.C. Batouche, "Une nouvelle approche de reconstruction d’objets 3D à partir d’images utilisant enveloppe visuelle / Stérervision", Dans les proceedings de la CIIA 06 14,15 et 16 Mai 2006, SAIDA, Algérie.

[34] Dubois Emmanuel , Laurence Nigay et Jocelyne Troccaz, “Combinons le monde virtuel et le monde réel”, Acte des Rencontres Jeunes Chercheurs en IHM, Île de Berder, France, Mai 2000, p. 31-35.

[35] Feiner Steven, “A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment”, In Proc ISWC, October 13–14, 1997, pages 74–81.

[36] Gennery D.B., "Tracking Known Three-Dimensional Objects", 2nd National Conference on Artificial Intelligence, Pittsburg, PA, August, 1982, pp 13-17.

[37] Giai-Checa B., Deriche R., Viéville Th., et Faugeras O., "Suivi de segments dans une séquence d'images monoculaire", INRIA, RR 2113, 1993.

[38] Gil S., Milanese R. and Pun T., "Feature selection for object tracking in traffic scenes", 1994.

[39] Gleicher M., “Projective Registration with Difference Decomposition” Proc.CVPR '97, 1997, pp. 331-337.

[40] Grasset Raphaël , Jean-Dominique Gascuel, “Environnement de Réalité Augmentée Collaboratif : Manipulation d’Objets Réels et Virtuels.”AFIG '01 (Actes des 14èmes journéees de l'AFIG) pages 101-112 , Novembre 2001

[41] Hager G.D. and Belhumeur P.N., “Efficient Region Tracking with Parametric Models of Geometry and Illumination”, IEEE Trans. PAMI, vol. 20, no. 10, 1998, pp. 1025-1039.

[42] Hardenberg, C. Bérard, F. (2001) *Bare-Hand Humant-Computer Interaction*, in ACM workshop on Perceptive User Interfaces (PUI 2001), Orlando, Florida.

[43] Hirokazu Kato, Mark Billinghurst. “Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System”. Proceedings of 2nd IEEE and ACM International Workshop on Augmented Reality IWAR ‘99, 1999. pp.85-94.

[44] Höllerer Tobias, “Situated Documentaries: Embedding Multimedia Presentations in the Real World”, Proceedings of ISWC ’99, IEEE October 18–19, 1999, pp. 79–86.

[45] Huttenlocher D.P., Noh J.J., Rucklidge W.J., "Tacking non-rigid objects in complex scenes", TR92-1320, Computer science departement, cornell university, dec. 1992.

[46] Isard M. and Blake A., “Condensation- conditional density propagation for visual tracking”, Int. J. Computer Vision, 29(1), 1998, pp-5-28.

[47] Jacons Marco C., “Managing Latency in Complex Augmented Reality Systems”. Proceedings of the 1997 symposium on Interactive 3D graphics, Providence, Rhode Island, United States.

[48] JAIN A.K. and MAO J., 1996, “Artificial neural networks: A tutorial”, IEEE. Computer, vol. 29, no. 3, pp. 31-44.

[49] Jethwa M., A. Zisserman, A. Fitzgibbon. "Real-time Panoramic Mosaics and Augmented Reality". On-Line Proceedings of the Ninth British Machine Vision Conference. 1998. UK.

[50] Jurie F. and M. Dhome. Hyperplane approximation for template matching. IEEE Trans. Pattern Analysis and Machine Intelligence, 24(7), pages 996-1000, 2002.

[51] Jurie F. and M. Dhome. Real time tracking of 3d objects with occultations. In International Conference on Image Processing, pages (I)413--416, Thessaloniki, Greece, October 2001.

[52] Jurie F. and M. Dhome. Un algorithme efficace de suivi d'objets dans des séquences d'images. In Conf. reconnaissnace des Formes et Intelligence Artificielle, pages 537--546, Paris, Fev. 2000.

[53] Kass M., Witkin A.P.and Terzopoulos D, "Snakes: Active Contour Model", International Journal of Computer Vision, Vol. 1, No. 4, jan 1998, pp. 321-331

[54] Koller D., Weber J. and Malik J., "Towards realtime visual based tracking in cluttered traffic scenes", Proc. Of the Intelligent Vehicles Symposium, october 1994.

[55] Korn A.F., "Towards a symbolic representaion of intensity changes in images", IEEE TPAMI, Vol. 10, 1998, pp 610-625.

[56] La Cascia M., Sclaroff S. and Athitsos V., “Fast, Reliable Head Tracking under Varying Illumination: An Approach Based on Registration of Textured-Mapped 3D Models”, IEEE Trans. PAMI, vol. 22, no. 4, 2000, pp. 322-336.

[57] Leymarie F., Levine M.D., "Tracking Deformable Objects in the Plane Using an Active Contour Model", IEEE PAMI, 1993, pp. 617-634.

[58] Lowe D.G., "Fitting Parameterized Three-Dimentional Models to Images", IEEE TPAMI, Vol. 13, No. 5, May 1991, pp 441-450.

[59] Lowe D.G., "Robust Model-Based Motion Tracking Through the Integration of Search and Estimation", International Journal of Computer Vision (8), No. 2, August 1992, pp 113-122.

[60] Martin J. , “ Suivi et interpretation de geste : Application de la vision par ordinateur à l’interaction Homme-Machine ». DEA, Université Joseph- Institut national Polytechnique de Grenoble, 1995.

[61] Martin Jérôme, "Reconnaissance de geste en vision par ordinateur", Thèse de doctorat dans le cadre de l’école doctorale « *Mathématique et informatique*», Juillet 2000.

[62] MartinJ. , « Techniques visuelle de détection et de suivi de mouvements », Magistère informatique, université Joseph Fourier , Septembre 1994.

[63] Masson L., Dhome M. and Jurie F., “Robust Real time tracking of 3D objects”, International Conference on Pattern Recognition, pages (IV)252-255, Cambridge, UK, 2004.

[64] Meshoul S., M. Batouche et M. Bouzenada, "Insertion réaliste d'objets 2D dans des séquences vidéos pour des applications de réalité augmentée", Conférence Internationale sur les Sciences Electroniques Technologies de l'information et des Télécommunications, SETIT 2004, Sousse, Tunisie, 15-20 Mars 2004, ISBN : 9973-41-902-2.

[65] Meshoul S., M. Bouzenada, M.Baddache, Berrandjia, « Une méthode robuste pour l'insertion réaliste d'objets virtuels dans une séquence vidéo », dans les actes de la Conférence Internationale sur la Productique, CIP'03, CDTA - Alger, 2003.

[66] Milgram, P., Takemura, H., Utsumi, A., Kishino, F., "Augmented Reality: A Class of Displays on the Reality-Virtuality Continuum", SPIE Vol. 2351 Telemanipulator and Telepresence Technologies, 1994.

[67] Nini B., M. Berkane, M. Bouzenada, " Manipulation d'objets virtuels dans un cadre collaboratif ", dans proceedings du 4ème séminaire national en informatique SNIB'2004,, 4-6 Mai, 2004.

[68] Nini B., M. Bouzenada et M. Batouche, « La réalité augmentée temps réel », JSTA'03, Guelma.

[69] Nini B., M. Bouzenada et M. Batouche, « Augmentation 3D à base de pattern planaire », Dans les proceedings de la 3ème Conférence sur le génie électrique 15-16 Février 2004, Alger, Algérie.

[70] Nini Brahim and Chaouki Batouche. Real Time Virtualized Real Object Manipulation in an Augmented Reality Environment. 1st International Symposium on Brain, Vision and Artificial Intelligence, 19-21 October. M. De Gregorio et al. (Eds.): BVAI 2005, LNCS 3704, pp. 477 – 486, 2005. Naples. Italy.

[71] Nini Brahim and Chaouki Batouche. Simulation of the Handling of Real Objects with a Complete Control of Rotation. Information Technology Journal 6(5): 672-680, 2007. ISSN 1812-5638. Asian Network for Scientific Information.

[72] Nini Brahim and Chaouki Batouche. Utilisation d'une Séquence pour l'Augmentation en Réalité Augmentée. JIG'05, Biskra. Algérie.

[73] Nini Brahim and Chaouki Batouche. Virtual Object Manipulation in Collaborative Augmented Reality Environment. IEEE-ICIT 2004, December 8-10 (2004). Tunis.

[74] Nini Brahim and Chaouki Batouche. Virtualized Real Object Integration and Manipulation in an Augmented Scene. CAIP 2005. The 11th International Conference on Computer Analysis of Images and Patterns, 5-9 Septenber, 2005. A. Gagalowicz and W. Philips (Eds.): CAIP 2005, LNCS 3691, pp. 248 – 255,

[75] Ouhaddi H. et P. Horain, "Conception et ajustement d'un modèle 3D articulé de la main", *Actes des 6èmes Journées de Travail du GT Réalité Virtuelle*, Issy-les-Moulineaux, 12-13 mars, 1998, pp. 83-90. *http://www-sim.int evry/Publications.*

[76] Ouhaddi H. , P. Horain, K. Mikolajczyk, « Modélisation et suivi de la main», CORESA juin1998, http://www-sim.int-evry.fr.

[77] Pérez P., Hue C., Vermaak J. and Gangnet M., "Color based probabilistic tracking", ECCV'02, 2002, pages 661–675.

[78] Poupyrev Ivan, Desney Tan, Mark Billinghurst, Hirokazu Kato, Holger Regenbrecht, Nobuji Tetsutani. "Developing a Generic Augmented-Reality Interface". IEEE Computer, March 2002. pp 44-50.

[79] Shahzad Malik, Gerhard Roth, Chris McDonald. "Robust 2D Tracking for Real-time Augmented Reality". In Proceedings of Vision Interface (VI) 2002, Calgary, Alberta, Canada.

[80] Simon G., "Détermination du point de vue à partir d'une observation d'un objet 3D dont le modèle est connu", Rapport de DEA, Université Henri Poincaré Nancy I, September 1995.

[81] Simon Gilles and Marie-Odile Berger. Reconstructing while Registering: a novel approach for markerless augmented reality Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'02), 2002.

[82] Simon Gilles, "Vers un Système de Réalité Augmentée Autonome" , Thèse de doctorat de l'Université Henri Poincaré , Décembre 1999.

[83] Simon Gilles, Andrew Fitzgibbon, Andrew Zisserman. "Markerless Tracking using Planar Structures in the Scene". Proceedings. IEEE and ACM International Symposium on Augmented Reality ISAR 2000, 2000. pp 120-128.

[84] Stricker Didier, "Tracking with Reference Images: A Real-Time and Markerless Tracking Solution for Outdoor Augmented Reality Applications, ACM Inc, 2002.

[85] Terzopulos D., Szeliski R., "Tracking with Kalman snakes", In active vision, MIT press, Cambridge, 1992.

[86] Thevenin, D. Bérard, F. and Coutaz, J. (1999) Capture d'Inscriptions pour la Réalité Augmentée, actes de la 11ème conférence francophone sur l'Interaction Homme-Machine, Montpellier, France.

[87] Trucco Emanuele, Alessandro Verri. Introductory Techniques for 3D Computer Vision. Prentice-Hall, 1998.

[88] Vincze M., "Robust Tracking of Ellipses at Frame Rate", Pattern Recognition, 34(2):487-498, février 2001.

[89] Ware C. et Balakrishnan, R. “*Reaching for Objects in VR Displays: Lag and Frame Rate*”. ACM Transactions on Computer-Human Interaction (TOCHI), Vol. 1, No. 4, pages 331-356, Décembre 1994.

[90] Wellner, “*Interacting with paper on the DigitalDesk*”. Communication of the ACM n. 7, p. 87-96, Juillet 1993.

[91] Wellner, P. “*The DigitalDesk Calculator: Tangible Manipulation on a Desk Top Display”. ACM conference on User Interface Systems and Toolkits (UIST), ACM publ, p. 27-33, 1991*.

[92] Worral A.D., Ferryman J.M. and Sullivan G.D., "Pose and structure recovery using active models", Proc. 6th British Machine Vision, pp 137-146.

[93] Worral A.D., Sullivan G.D. and Baker K.D., "Pose refinement of active models using forces in 3D", Third European Conference on Computer Vision, Stockholm, Sweden, 1994.

[94] Zisserman, A. “*Projective transformation between two images of a planar scene*”, dans “Computer Vision Online: Vision Geometry and Mathematics”, édité par Fisher, R. B. à l’Université d’Edinburgh.