RØalitØEnrichie par SynthŁse

Pierre JancŁne, Christophe Meilhac, Fabrice Neyret,
Xavier Provot, Jean-Philippe Tarel,
Jean-Marc VØen, Anne Verroust

INRIA
Domaine de Voluceau, ROCQUENCOURT
78153 Le Chesnay cedex FRANCE
Contact e-mail: Christophe.Meilhac@inria.fr
http://www-rocq.inria.fr/syntim/analyse/video-fra.html

RØsumØ

Nous proposons une mØode pour automatiser l'insertion cohØente d'objets 3D de synthŁse dans des sØquences d'images rØlles. Celle-ci fait appel conjointement des techniques d'analyse et de synthŁse d'images pour calculer les images modifiØs.

Mots-clØ
analyse d'images, synthŁse d'images,rØalitØirtuelle .

Abstract

In this paper, a technical solution is presented to automate the mixing of real and synthetic objects in a same animated video sequence. We aim at achieving a close binding between 3D-based analysis and synthesis techniques to compute the interaction between a real scene captured in a sequence of calibrated images, and a computer-generated environment.

Keywords:
Image analysis, Image synthesis, virtual reality.

1 Introduction

1.1 Etat de l'Art

La rØalitØaugmentØprend de l'ampleur dans diffØents domaines comme les processus de fabrication pour l'industrie [1, 2], l'architecture d'intØieur [3, 4] et le domaine mØical [5]. Dans la plupart de ces applications, le problŁme principal est d'assurer la prØision de la superposition des images calculØs sur les images rØlles.

En effet quand on insŁre des objets 3D de synthŁse dans une sØquence d'images pour des applications de vidØ, il faut s'assurer de la cohØence de l'illumination, des ombres et des occultations des objets virtuels par les objets rØls. Il existe plusieurs approches en rØponse ce problŁme :

- Les applications actuelles d'effets spØiaux pour la production vidØsuperposent simplement l'image de synthŁse devant l'image rØlle (incrustation ou *alpha keying*). Si l'image de synthŁse doit Œe partiellement occultØ un masque 2D doit alors Œe fourni, ce qui interdit l'animation ou contraint interpoler l'Øolution du masque au cours du temps.
- La *Synthetic TV* [6], garantit la cohØence gØmØtrique entre images rØlles et objets de synthŁse en synchronisant les mouvements de la 'camØa' de synthŁse, correctement modifiØ avec les mouvements de la camØa rØlle.
- La *Computer Augmented Reality* [7, 8], s'attache au calcul de l'illumination globale dans une scŁne rØlle enrichie d'objets de synthŁse.

Ces mØodes nØessitent une forte interaction avec l'utilisateur lors de l'opØation de mixage. Notre but est d'automatiser partiellement ce traitement.

1.2 Notre Approche

La RØalitØEnrichie par SynthŁse (RES) a pour but l'insertion d'objets 3D de synthŁse dans une sØquence 2D d'images rØlles, en assurant la cohØence 3D pour les occultations, les ombres, les collisions et les contacts. L'idØconsiste reconnatre et localiser des objets 3D dans les images 2D, afin de construire des sortes de masques 3D.

Notre approche associe des techniques d'analyse et de synthŁse d'images via la constitution de modŁles 3D synthØisables compatibles, autorisant un mixage 3D-cohØent des objets de synthŁse dans les images 2D.

Les applications de cette technique peuvent se trouver en visualisation (par exemple en imagerie mØicale), et pour la production vidØ(effets spØiaux pour les films, publicitØude d'impact...).

Un modŁle synthØisable dØit une scŁne composØ d'objets, de sources de lumiŁres et d'une camØa. Les objets ont une forme, une position, une orientation, des propriØde surface (ex : couleurs ambiante, diffuse et spØulaire) et Øventuellement une Øolution temporelle (mouvement rigide ou dØormation). Un modeleur est un outil interactif permettant un utilisateur de dØire un modŁle synthØisable 3D.

Les traitements effectuØpour la RES sont effectuØen diffØ

- Les objets rØls choisis pour l'interaction sont prØlablement modØisØ l'aide du modeleur, qui dispose de fonctionnalitØspØifiques pour 'dØalquer' interactivement les objets rØls partir de vues sous diffØents angles. Cet outil permet de constituer une base de modŁles d'objets susceptibles d'Œe prØents dans la scŁne enrichir.

- La séquence réelle est tout d'abord numérisée et analysée: les modèles des objets 'reconnus' dans la scène réelle sont recalés (cf. partie 2), de manière reconstituer une scène synthétique les remettant aux places et orientations identiques celles de la scène originale. Les caractéristiques de la caméra sont également estimées.
- L'utilisateur spécifie sa scène animée de synthèse l'aide de notre modeleur ACTION3D (cf. partie 3), en partant de la scène recalée construite par l'analyse.
- Pour chaque image de la séquence, le processeur de production d'images lit les données géométriques d'ACTION3D qui contiennent les objets de synthèse et les objets reconnus, appelle les modules de simulation dynamique qui complètent la scène, et lance le logiciel de rendu RAYSHADE afin de construire quatre images partielles, qui seront mixées avec l'image originale au moyen d'un module de composition (cf. partie 4).

Ceci conduit aux séquences d'images présentées dans la partie 5, montrant l'insertion d'objets 3D de synthèse dans des images ou séquences réelles avec maintien de la cohérence 3D pour les occultations, les ombres et les collisions.

2 Analyse d'image

2.1 Introduction

L'une des spécificités de l'application Réalité Enrichie par la Synthèse (RES), par rapport aux applications plus classiques d'analyse comme celles dédiées la robotique, est la nécessité d'obtenir une description de certains objets observés plus complète que pour les applications d'analyse classiques, tant dans son aspect géométrique que photométrique. En particulier, la connaissance géométrique nécessaire pour pouvoir ajouter des objets de synthèse en interaction 3D réaliste avec la scène dépasse de loin les données que peuvent extraire les méthodes de reconstruction connues en vision par ordinateur l'heure actuelle. Nous avons donc choisi une méthode d'analyse par modèle, qui suppose d'introduire des connaissances géométriques données a priori sous la forme d'un modèle des objets 3D présents dans la scène. Ceci permet de restituer les parties importantes de la scène de faon robuste, partir d'informations dans les images.

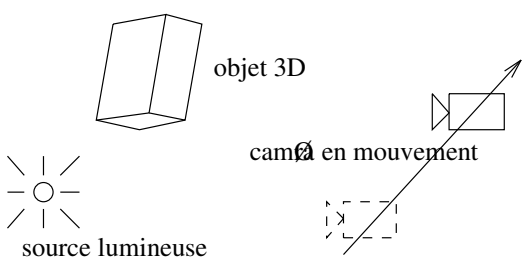

FIG. 1 - Qu'est ce qu'une scène?

Dans notre cadre, une description du monde contient un observateur, des sources lumineuses et des objets (figure 1). Le but que se fixe l'analyse est de fournir les

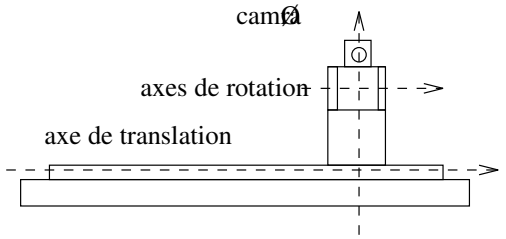

FIG. 2 - Le portique du laboratoire.

informations nécessaires (géométrie et photométrie) sur ces trois types d'entités présentes dans la scène observée. La description de l'observateur est réalisée en ligne par calibration (section 2.3). Les modèles des objets sont positionnés par recalage sur des données 3D extraites des images (section 2.4) . Enfin, on identifie les sources de lumière sous la forme de conditions d'éclairage équivalentes (section 2.5).

La description, ou modèle de la scène, est constitué:

1. d'un aspect géométrique : localisation des objets présents dans l'image dont on a un modèle, modélisation et mouvement de la caméra,
2. d'un aspect photométrique : localisation des sources de lumière équivalentes et estimation de la réflectance des facettes.

2.2 Contraintes sur la scène

Nous récapitulons dans cette partie les contraintes dues au matériel (cf. fig 2) et les hypothèses qui simplifient le traitement. Ces hypothèses limitent la généralité du système, la relaxation de ces contraintes constitue une part de notre travail actuel. Cependant, l'objectif de cet article est avant tout de démontrer la faisabilité de la RES, c'est dire la possibilité d'ajouter des objets de synthèse dans une scène réelle en interaction 3D réaliste, de manière dépasser les limites des systèmes d'incrustation actuels.

2.2.1 Contraintes d'acquisition

Le laboratoire dispose d'une caméra couleur Sony XC007 avec un objectif zoom Canon J15x9.5B. La caméra est placée sur un portique qui permet le pilotage du zoom, le réglage du diaphragme de l'objectif et son déplacement selon un axe de translation et deux axes de rotation (figure 2).

Un mouvement de caméra calibré. Puisque le système d'acquisition est pilotable avec une bonne précision, nous supposons que la position de la caméra en rotation et translation ainsi que son zoom et sa mise au point sont connues. Ces informations sont accessibles dès que le portique est calibré cinématiquement [9]. Dans l'état actuel de nos développements, seuls les déplacements en translation de la caméra sont calibrés.

Des scènes statiques. Les traitements sur les images décrits par la suite (partie 2.4 et 2.5) doivent être assez précis pour permettre une incrustation de bonne qualité

sinon on risque d'observer des effets de bavures entre les images des objets réels et de synthèse. Nous supposons donc que l'image observée est propre, et en particulier qu'elle ne contient pas trop de bruit, ce qui nécessite une moyenne sur plusieurs images d'un même point de vue. Puisque nous ne nous intéressons qu' des scènes statiques, il suffit de faire l'acquisition la cadence de 2 3 images par seconde.

Pas d'analyse en temps réel. Comme indiqué en introduction, la RES se déroule dans un contexte 'différé' et non 'temps réel', ce qui permet de traiter des séquences préexistantes calibrées, et de procéder des opérations d'analyse complètes donc coûteuses. Dans l'état actuel de notre système, il faut compter une durée d'une journée pour analyser avec succès une paire stéréoscopique extraite d'une séquence de deux secondes.

2.2.2 Hypothèses simplificatrices

Comme il n'est pas possible d'obtenir toute l'information géométrique utile la synthèse d'images sur un objet, moins de disposer de tous les points de vue, seuls quelques indices 3D robustes sont extraits des images pour permettre le recalage avec les modèles complets des objets de la scène.

Modèle des objets. On suppose lors du processus d'analyse des images qu'un modèle des objets les plus importants dans la scène est disponible.

Le modèle géométrique d'un objet est défini par ses surfaces les plus caractéristiques. La description utilisée est une représentation par bords 3D, ou les surfaces sont définies comme des polyèdres, eux-mêmes représentés par un ensemble de polygones orientés, ou facettes. Les parties courbes d'un objet sont approximables par ce type de représentation. Ainsi, certaines parties des objets manipulés peuvent être courbes, mais elles ne seraient pas utilisées pour l'analyse stéréoscopique. Cette représentation est complète du point de vue géométrique, puisque le point de vue et les conditions d'éclairage n'interviennent pas dans la caractérisation de l'objet.

Des facettes planes. Nous supposerons que les objets observés dont on désire obtenir une localisation possèdent dans l'image suffisamment de surfaces planes visibles photométrie uniforme. Un minimum de trois facettes visibles et non occultées est nécessaire pour chaque objet afin d'obtenir un recalage robuste et précis de son modèle géométrique. Cela induit les hypothèses suivantes :

1. les conditions d'illumination doivent être suffisamment bonnes : éclairage lointain, diffus et important ;
2. les matériaux doivent être de nature peu spéculaire de manière réduire les reflets de lumière.

Ces hypothèses permettent de réaliser une analyse en région et d'obtenir de manière directe, par reconstruction, des facettes 3D (voir section 2.4). De plus, la face est la primitive la géométrie la plus simple qui peut contenir une information photométrique de nature texturelle.

On suppose donc que la scène observée est localement plane : cette hypothèse est toujours vérifiée si l'on observe des surfaces suffisamment petites, mais raisonnable uniquement pour des scènes de type 'intérieur', comme des scènes de bureau ou des scènes architecturales.

2.3 Analyse du point de vue

Dans cette partie, nous explicitons la formation d'une image, et comment on acquiert les connaissances qui concernent l'observateur (position, mouvement, caractéristiques propres).

L'observateur est décrit par un ensemble de paramètres. Ceux-ci sont appelés *point de vue*. Celui-ci est modifiable par les paramètres de contrle du portique : commande de translation, rotation, zoom. La calibration, c'est--dire la connaissance de la transformation que subit un point de vue par les paramètres de contrle, est réalisée par processus hors-ligne grce une mire de calibration (objet de géométrie parfaitement connue). Ainsi dès l'acquisition, le point de vue de l'observateur est connu.

2.3.1 Définition d'un point de vue

Le point de vue est déterminé par :

- une origine, appelée centre de la caméra,
- une direction de visée dont l'axe passe par l'origine,
- un rectangle image, le plan image est orthogonal la direction de visée,
- un twist, qui fixe l'angle de l'un des côtés de l'image avec l'horizontale,
- une focale et
- une distance de mise au point, les objets qui ne sont pas cette distance sont flous.

Un certain nombre d'actions sur la caméra peuvent modifier les paramètres du point de vue (le focus ou mise au point, le zoom, le diaphragme et les déplacement du robot en rotation et translation).

Nous utiliserons une version simplifiée de la modélisation d'une caméra, le modèle sténopé (cf. figure 3). L'image d'un point dans l'espace euclidien 3D est l'intersection du plan image et de la droite passant par le point objet et le centre. L'objectif de la calibration est d'extraire les paramètres de l'observateur et d'être en mesure de suivre les transformations de ces paramètres sous l'effet des commandes du portique.

2.3.2 Calibration de l'objectif de la caméra

La calibration d'une caméra est une méthodologie permettant d'extraire les paramètres du modèle. Elle est réalisée en général grce une mire de calibration (objet de géométrie parfaitement connue). On repère la position d'un certain nombre de points caractéristiques sur la mire par rapport un repère lié celle-ci, et on extrait dans l'image de calibration le point image correspondant chacun d'entre eux [10].

Une caméra possède deux commandes pour améliorer la qualité de l'image, qui ont des conséquences sur la projection perspective : le zoom et le focus. Chacune

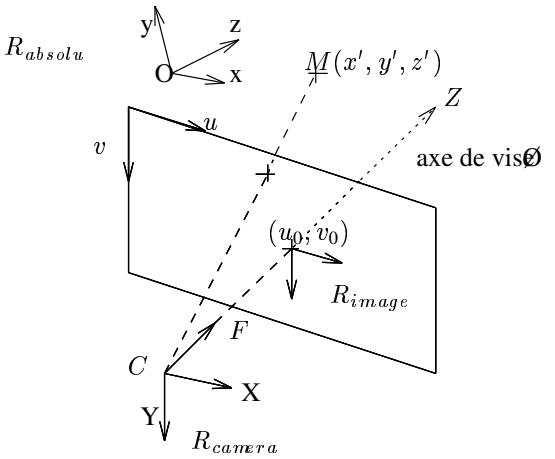

FIG. 3 - Modèle sténopé d'une caméra.

d'entre elles modifie le point de vue : le centre se translate sur l'axe de visée et la focale augmente ou diminue.

2.3.3 Calibration du portique

Pour étudier le mouvement de la caméra, on peut distinguer trois types de repères. Le premier est lié au monde donc immobile lors d'un déplacement de caméra, le second est le repère caméra dans lequel la projection centrale sur le plan image s'exprime naturellement, le dernier est le repère pince dans lequel les commandes de déplacement sont connues.

La calibration caméra-pince est une méthodologie qui doit fournir la matrice de passage entre le repère pince et le repère caméra. L'article [9] présente une méthodologie de calibration.

2.3.4 Récapitulatif

Nous venons de décrire notre démarche pour obtenir un système entièrement calibré dès l'acquisition. Nous avons décomposé l'analyse du point de vue en deux parties : d'une part l'extraction des paramètres liés à la caméra elle-même (les commandes robotiques susceptibles de les modifier sont le zoom et le focus) et d'autre part l'extraction des paramètres dérivant le mouvement de la caméra en translation et rotation. Dans les deux cas notre calibration repose sur des données images et sur l'observation d'une image de mire.

On peut améliorer notre système de calibration en relchant un certain nombre de contraintes tel que la restitution du mouvement par des calculs sur l'image, ou encore se passer totalement de mire de calibration, les paramètres de modélisation étant entièrement (ou partiellement) calculés à partir des images, a posteriori [11].

2.4 Analyse de la position des objets

Une fois les caractéristiques de l'observateur obtenues, il est possible d'extraire des informations 3D sur les objets de la scène. Comme nous connaissons par hypothèse un modèle des principaux objets qui constituent la scène, le processus d'analyse des paires d'images consiste à extraire suffisamment de primitives 3D pour localiser robustement ces objets dans la scène. Ainsi, dans notre processus chaque modèle des objets importants est recalé précisément par rapport aux facettes 3D reconstruites. Seuls les objets qui peuvent être représentés par un ensemble de contours rigides et qui possèdent suffisamment de surfaces homogènes planes sont actuellement recalables par notre système.

2.4.1 Extraction de caractéristiques 3-D

Dans le processus d'analyse de la séquence, nous choisissons des paires d'images ou l'objet qui nous intéresse est observé sous deux points de vue assez différents pour permettre une reconstruction de qualité. L'objet doit être bien visible dans l'image pour pouvoir effectuer le traitement dans de bonnes conditions. La position des objets dans l'espace 3D n'est alors calculée que sur ces deux images extraites de la séquence.

Notre méthode repose donc sur une analyse stéréoscopique, à partir d'une paire d'images gauche-droite, dont les étapes sont (figure 4) :

Segmentation : chaque image est découpée en régions, indépendamment de l'autre image,

Appariement 2D : Les régions gauches identifiées précédemment sont si possible mises en correspondance avec les régions de l'image droite,

Reconstruction 3D : les régions, de chaque appariement issues d'une même surface plane, sont localisées dans l'espace 3D euclidien.

Ces étapes nous fournissent un ensemble de primitives 3D qui nous permettra par la suite de faire du recalage. Le but n'est donc pas d'obtenir une reconstruction exhaustive mais plutôt d'avoir suffisamment d'éléments reconstruits pour que le recalage 3D soit possible avec une bonne précision.
Nous allons décrire en détail les différentes étapes.

a) **Correction des images** Grâce aux informations fournies par la calibration, les images du couple stéréoscopique sont corrigées de leurs distorsions pour s'approcher d'un modèle de projection sténopé parfait [10]. Les distorsions photométriques doivent être aussi corrigées pour ne pas biaiser les algorithmes qui s'appuient sur les intensités dans l'image [12].

b) **Segmentation en régions : primitive 2-D** Les paires stéréoscopiques calibrées sont segmentées en régions, afin d'extraire les principales zones homogènes au sens d'un certain critère, qui seront reconstruites sous forme de faces. Ces facettes servent d'ancrage au recalage des modèles des objets les plus importants de la scène. Le principe de la segmentation est de regrouper progressivement les pixels de l'image avec leurs voisins en fonction de critères d'homogénéité. Ceci peut être réalisé par croissance de régions [13, 14], ou par division récursive de l'image pilotée par la mise en correspondance [15, 16], par classification floue [17] ou par une minimisation d'énergie [18].

L'idéal serait un processus de segmentation aussi peu paramétrique possible (par exemple, ne dépendant

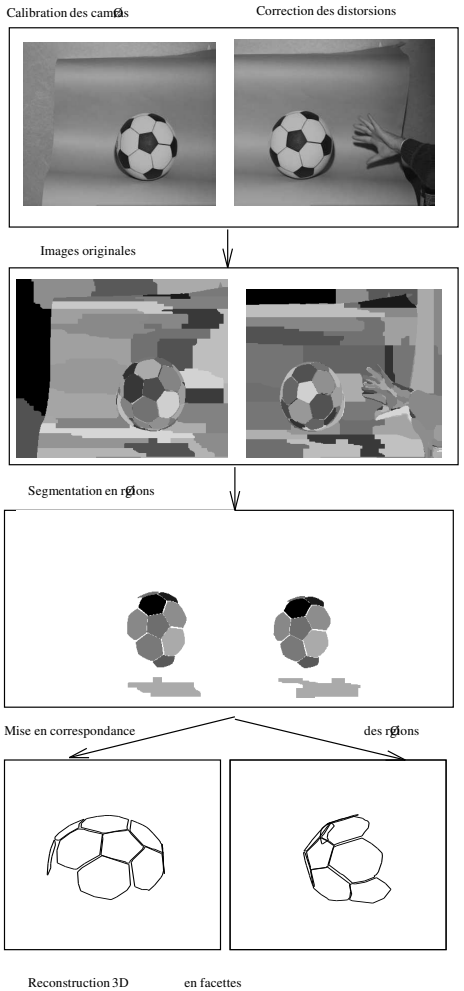

FIG. 4 - Le processus d'analyse stéréoscopique.

que d'un paramètre d'échelle), afin d'automatiser la procédure. De ce point de vue, la dernière méthode conduit des résultats facilement contrlables.

c) **Mise en correspondance 2D** Comme pour des primitives plus classiques, chaque région, une fois segmentée, doit être appariée avec son correspondant dans l'autre image de la paire. Un des avantages de l'approche par région tient à la quantité d'information disponible pour opérer les bons appariements : des critères géométriques (forme, contrainte épipolaire), mais aussi photométriques (l'aspect de la projection d'une région change peu entre deux points de vue stéréo) peuvent être utilisés.

Pratiquement, la méthode mise au point cherche, pour chaque région de l'image gauche, un ensemble de régions candidates dans l'image droite, suivant un critère lche de ressemblance photométrique (variance et moyenne similaire). On trie ensuite les candidats par un ensemble de contraintes : taille semblable, contrainte épipolaire sur le centre de gravité [13]. Cette technique conduit assez facilement aux bons appariements pour les principales régions de l'image, même en présence de quelques erreurs de segmentation locales.

L encore, comme le choix des critères et des seuils attachés peut poser problème, nous avons défini une stratégie systématique qui conduit à de bons résultats pour une gamme de scènes d'intérieur. Cependant, il serait possible de raffiner le processus selon le contenu désiré dans les images.

d) **Reconstruction 3D** Nous disposons de régions appariées, pour lesquelles une triangulation directe est impossible. Une mise en correspondance explicite des points extraits des régions aurait pu résoudre cette difficulté. Des points particuliers des contours, comme par exemple ceux de forte courbure, conviennent bien. Mais cette approche souffre souvent des erreurs de segmentation intervenues lors des étapes de traitement bas-niveau.

C'est pourquoi, nous avons développé une méthode de reconstruction de faces 3D, qui est globale sur les régions appariées. Les facettes sont reconstruites comme des morceaux de surfaces planes par le biais de leur équation. Quant aux frontières des régions 2D, elles servent uniquement à découper le contour 3D dans le plan calculé [19, 20].

2.4.2 Recalage 3D

A cette étape du processus d'analyse, nous avons notre disposition les modèles 3D d'objets susceptibles d'être présents dans la scène. Grâce à l'étape de reconstruction 3D décrite précédemment, nous disposons de plus d'un ensemble de facettes planes présentes et bien localisées sur les objets. Retrouver la position de ces objets se réduit donc par la mise en correspondance des facettes incluses dans l'un et l'autre de ces deux ensembles de faces. Deux types de méthodes ont été développés et sont utilisés de manière complémentaire :

- Le premier algorithme est dérivé de la méthode de recalage ICP [21]. Il s'agit d'un algorithme itératif qui calcule à chaque pas une mise en correspondance de chaque point de donnée avec le point le plus proche du modèle à positionner. Cette correspondance permet le calcul d'un déplacement optimal entre le repère de reconstruction et le repère propre a l'objet au sens des moindres carrés. L'algorithme itère les deux étapes appariement/recalage jusqu' ce que l'erreur résiduelle soit stable. A partir d'une position initiale, une telle procédure converge systématiquement vers le plus proche minimum local.

  Les mises en correspondance sont analysées statistiquement à chaque étape de l'algorithme pour ne retenir que les plus robustes. Ceci a l'avantage de permettre le recalage d'un modèle d'objet isolé dans une scène complexe même si la reconstruction est partielle, puisque les fausses mises en correspondance sont alors en grande partie ignorées [19, 22].

- Les méthodes de classification floue, de par leur robustesse, fournissent un outil intéressant pour la reconnaissance 3D. Une méthode basée sur l'algorithme de classification avec du bruit a été développée qui permet de réaliser un recalage grossier rapidement sur les faces. Le recalage est ensuite affiné par l'algorithme décrit au point précédent. Cette méthode a l'avantage de gérer en parallèle plusieurs hypothèses de recalage intéressantes en imposant une contrainte

globale de bon recouvrement de l'ensemble des dØ placements possibles [23].

2.5 Analyse des sources lumineuses

L'illumination de la scŁne est modØlisØ par un flux lumineux parallŁle qui provient d'une source situØ l'infini. Elle est ainsi uniquement dØterminØ par sa direction d'illumination et son intensitØ lumineuse. Le calcul de l'Øclairement sur chaque facette est donc particuliŁrement simple, ce qui permet, en minimisant l'Øcart entre la distribution d'intensitØ prØvue et les intensitØ effectivement observØes sur l'intØrieur des rØgions, d'obtenir une estimation de la source Øquivalente et de rØflectance des surfaces. Le mŒme type d'approche a ØtØ appliquØ des modŁles de sources plus complexes comme une source ponctuelle uniforme Øquivalente [24]. Evidemment, l'obtention d'une source lumineuse Øquivalente ne permet pas de faire des estimations ayant une grande valeur du point de vue strictement physique, mais elle fourni des rØsultats suffisants dans le cadre de la RES.

2.6 RØcapitulatif

Nous avons prØsentØ un ensemble d'outils qui permettent d'analyser une sØquence d'images rØelles dans le but de fournir un modŁle gØomØtrique et photomØtrique de la scŁne suffisamment riche pour Œtre utilisØ dans le cadre de la RES. Nous avons dØcrit les possibilitØs du systŁme actuel et les contraintes que cela implique sur la nature des objets observØs et du systŁme d'acquisition des images. En particulier, les objets doivent possØder des facettes planes uniformes et ne pas se dØplacer dans la sØquence. De plus, la camØra et son mouvement doivent Œtre calibrØs. Enfin, les sources lumineuses doivent Œtre Øloignées des objets Øclairés. C'est dans ce contexte, que les processus d'analyse du point de vue, d'extraction de facettes 3D, de positionnement des modŁles d'objets, et d'estimation d'Øclairage Øquivalent, prØcØdemment dØcrits, peuvent fonctionner.

3 SynthŁse d'images

3.1 Introduction

Comme indiquØ prØcØdemment, le modŁle synthØtisable d'une scŁne comprend des objets, des sources de lumiŁre et une camØra (point de vue).

La problØmatique de l'analyse d'images est de reconstituer un modŁle aussi complet que possible partir d'images 2D. Les problØmatiques de la synthŁse d'images sont l'inverse de celles de l'analyse d'images. Plus prØcisément, la synthŁse d'images s'intØresse au problŁme direct de la simulation des divers phØnomŁnes qui conduisent de la description d'une scŁne la production d'une image [25] :

modØlisation : Une premiŁre problØmatique concerne la description gØomØtrique des objets, puis la structuration d'ensembles d'objets. Les modalitØs de description sont trŁs variØes : la modØlisation peut Œtre interactive, constructive, analytique, sous contraintes. Il faut Ødier non seulement les diverses reprØsentations, mais aussi les procØdures conduisant ces modØlisations : dØformations globales, 'sculpture' par modifications successives, grammaires.

De mŒme, les types de structures liant les objets sont multiples : hiØrarchies, articulations, squelettes. Il existe Ølement nombre de modŁles dØdiØs des types d'objets particuliers (systŁmes de particules, liquides, fumØes et nuages, objets complexes, fractales). Le logiciel permettant de constituer cette description est un modeleur gØomØtrique.

animation : Une seconde problØmatique est liØe l'Øvolution temporelle de ces objets [26], qui comprend la cinØmatique, la simulation physique, l'animation des dØformations, le contrle du mouvement, la simulation comportementale. Le logiciel permettant de spØcifier ces animations est un modeleur d'animation.

rendu : Une troisiŁme problØmatique s'attache la simulation des interactions lumineuses et au calcul de l'image proprement dite, et se spare en deux branches, l'illumination globale et l'illumination locale [1]. L'illumination globale prend en charge la rØsolution de l'Øquilibre dans les Øchanges d'Ønergie lumineuse, tandis que l'illumination locale s'occupe de l'interaction de la lumiŁre et de la matiŁre (modØlisation de la rØflectance, des textures...). Le logiciel effectuant le calcul d'une image est un programme de rendu (ou render). Il est gØnØralement contrlØ par un processeur de production d'image, qui se charge de reconstituer l'Øtat de la scŁne chaque pas de temps puis de post-traiter et de stocker les images aprŁs calcul.

3.2 SynthŁse d'images pour la RES

3.2.1 construction de modŁles pour l'analyse

Notre modeleur ACTION3D est tout d'abord utilisØ pour construire interactivement les modŁles gØomØtriques des objets rØels avec lesquels on souhaite interagir. On utilise plusieurs images de l'objet 'dØcalquer' que l'on place dans deux vues sur lesquelles on reprojette le modŁle en cours de construction. Un systŁme permet de limiter le placement des nouveaux points successifs dans les plans des faces dØjacquises [27].

3.2.2 construction de modŁles pour la synthŁse

Le modeleur est par la suite utilisØ classiquement pour modØliser et animer les objets de synthŁse rigides ou dØformØs gØomØtriquement. Les objets rØels par les lois de la dynamique sont construits par des modules dØdiØs directement lors de l'exØcution du rendu.
Le modeleur ACTION3D permet de :

- construire interactivement des objets polyØdriques et des surfaces de forme libre,
- dØformer et animer ces objets par des mØthodes de dØformation de l'espace 3D et par des mØthodes de mØtamorphose 3D,

1. Sans compter la modØlisation des sources de lumiŁres et des camØras.

- construire interactivement les textures, et spØcifier leur plaquage sur les surfaces.

3.2.3 modŁles dynamiques pour la synthŁse

Dans l'exemple prØsentØen partie 5, on utilise un module de simulation dynamique de tissus, constituØ d'un rØseau masse-ressort qui rØagit un champ de force et aux collisions avec d'autres objets [28]. A chaque pas de temps, la nouvelle position des masses est calculØe en fonction des forces internes (modØlisØes par les ressorts et la dissipation) et des forces externes (vent, collisions...). Il n'y a aucune diffØrence de traitement selon que le tissu rencontre un objet virtuel ou le modŁle d'un objet reconnu.

3.2.4 calcul du rendu

Comme indiquØplus haut et plus amplement dØcrit dans la partie 4, quatre images de synthŁses partielles sont calculØes, puis mixØes avec l'image rØelle.

Pour le calcul du rendu, nous utilisons le logiciel de rendu du domaine public RAYSHADE, qui opŁre par lancer de rayons [25] : l'Ønergie reue par chaque pixel du plan image est Øvaluøen lanant quelques rayons via ce pixel travers la scŁne. Ces rayons 'remontent' les trajectoires des photons en partant de l'observateur. L'illumination locale est Øvaluøau point d'impact de chaque rayon sur un objet de la scŁne, en fonction des propriØtØde surface de l'objet et de la lumiŁre parvenant en ce point, selon un modŁle de rØflectance qui indique quelle proportion de l'Ønergie est Ømise vers l'observateur. L'Ønergie incidente est elle-mŒme Øvaluøen lanant un rayon vers les sources de lumiŁre, afin de tester les occultations (gØnØrant les ombres). Un rayon est Ølement lancØdans la direction de rØflection donnØpar les lois de Descartes, afin de collecter l'Ønergie d'un Øventuel reflet.

4 Mixage (*compositing*)

En incrustant (figure 5) l'image de synthŁse sur l'image rØelle, l'occultation d'objets rØels par des objets virtuels Øtait dØjeffective dŁs les premiŁres approches pratiquØes en vidØo. Un masque associØl'image de synthŁse indique les pixels occupØs(avec une rØsolution subpixel), contrlant le mixage des images. Celui-ci peut Œtre obtenu en effectuant un calcul de rendu aprŁs avoir peint les objets en blanc uniforme (i.e. isotrope et non lambertien).

En fusionnant les modŁles reconnus avec les modŁles de synthŁse, l'occultation des objets virtuels par les objets rØels est automatique avec notre reprØsentation. Il faut cependant en tenir compte pour la gØnØration du masque : les objets reconnus ne doivent pas apparatre dans l'image de synthŁse ni dans le masque, dans la mesure ou ils sont dØjreprØsentØdans l'image rØelle, mais ils peuvent nØanmoins cacher en partie des objets de synthŁse. Ceci est obtenu en peignant les objets reconnus en noir et les objets virtuels en blanc uniforme. Les objets reconnus sont alors la fois invisibles et occultants, et peuvent Øventuellement Œtre rØflØchissants.

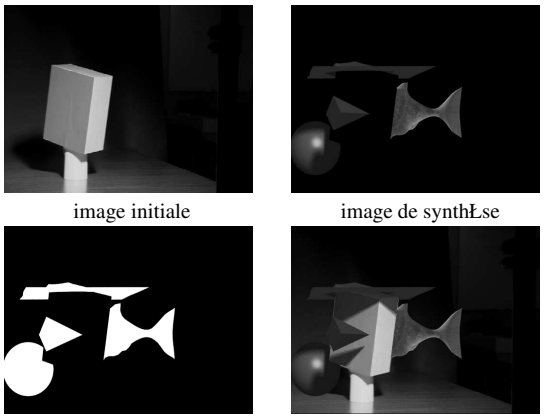

FIG. 5 - Noter la trace des objets rØels reconnus (la bote), invisibles mais ombrants et occultants.

Les ombres (figure 6) des objets rØels sur les objets virtuels apparaissent automatiquement avec notre reprØsentation unifiØ Les ombres des objets virtuels sur les objets rØels sont obtenues en calculant un coefficient d'attØnuation, multiplier l'image originale. Celui-ci est Øvaluøen effectuant un rendu des objets reconnus en blanc lambertien, avec et sans les autres objets, afin d'Øvaluer l'effet d'ombrage dû aux objets virtuels. Le facteur d'attØnuation est le ratio de l'image avec les ombres et de l'image de rØfØrence contenant seulement les objets reconnus. (Cette valeur obtenue pour tout point de l'image est Ølement calculØavec une rØsolution subpixel.)

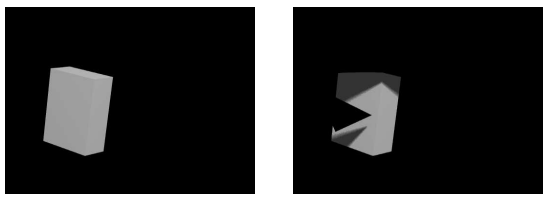

FIG. 6 - objets reconnus, seuls ou avec les objets de synthŁse ombrants. L'assombrissement est donnØpar le ratio des images.

Le mixage (figure 5) est alors dØfini par :
image enrichie = image rØelle × attØnuation × (1-masque) + image de synthŁse. [2]

5 Exemples

Les figures 7 et 8 sont extraites de sØquences d'images enrichies.

On peut remarquer dans la figure 7 les diverses formes d'interaction 3D entre objets virtuels et objets rØels (ici la bote) : appui de la nappe sur la bote, ombres de la nappe sur la bote et de la bote sur la balle verte, occultations entre les objets.

La figure 8 prØsente moins de cas d'interaction, mais illustre la possibilitØde rendre rØflØchissant des objets rØels. De plus, l'animation de la sØquence permet de

2. on remarquera que l'image de synthŁse contient : objets de synthŁse × masque + objets reconnus en noir rØflØchissants.

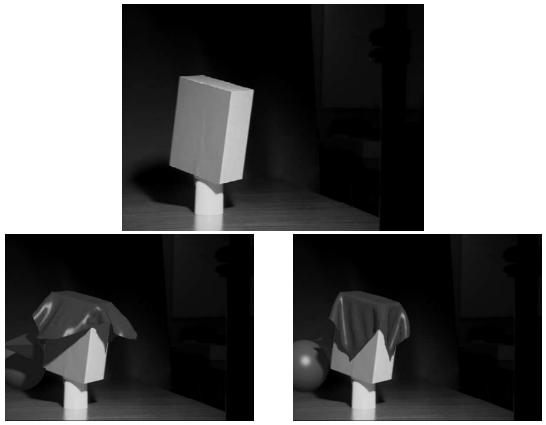

FIG. 7 - Image originale et deux images de la séquence enrichie par un drap de synthèse rouge mobile, un tetraèdre bleu et une sphère verte.

juger de la qualité du positionnement du cube rose, dont le déplacement est parfaitement synchrone avec celui de l'objet réel (la mire) sur lequel il est 'posé' (l'animation est visible à l'adresse WWW mentionnée en entête de l'article).

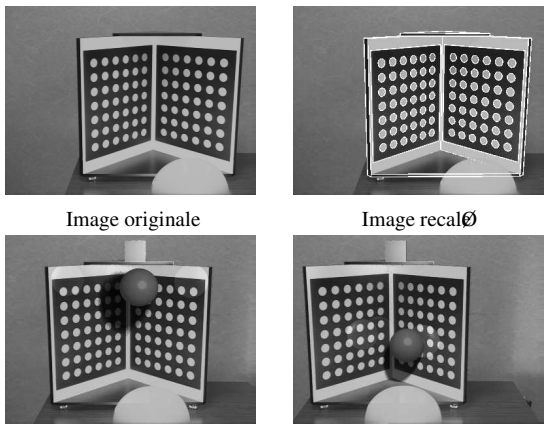

Image originale Image recalée

FIG. 8 - Images de la séquence enrichie par un une sphère verte mobile et un cube rose statique.

6 Conclusion

Nous avons démontré la faisabilité de l'automatisation de l'insertion d'objets 3D virtuels dans des séquences vidéo, en s'appuyant sur une coopération de techniques d'analyse et de synthèse d'images.

De plus, cette réalisation permet, du point de vue de la recherche en analyse d'images, de valider les algorithmes mis en uvre, et de poser de nouveaux problèmes.

Diverses directions de recherche sont ouvertes pour relcher ces contraintes et aboutir à un système plus performant et plus facile d'utilisation :

- Le calcul en différé des paramètres du mouvement d'un objet relativement à la caméra permettrait de traiter des images non acquises par un système entièrement calibré ou qui contient des objets en mouvement.
- Le fait de relcher les contraintes de planarité en disposant d'autres types de connaissances sur la scène 3D est un autre axe de recherches futures important.

En effet, nous avons opté pour une approche région car la RES nécessite l'obtention d'un modèle, à la fois géométrique et photométrique, relativement complet. Mais, le choix de facettes n'est pas exclusif, et l'association de régions avec des approches complémentaires à base par exemple de points, segments, courbes et cartes de profondeur, permettrait d'améliorer encore les résultats du processus d'analyse. En particulier, on gagnerait en efficacité de recalage en disposant de plusieurs types de primitives 3D, et la classe des objets que l'on pourrait traiter serait grandement étendue.

Du point de vue des applications de synthèse d'images, le mélange de séquences de synthèse et de séquences réelles permet de mettre en valeur le réalisme de la modélisation de l'animation et du rendu.

D'autre part, la faisabilité d'un tel mélange montre qu'il peut s'avérer très intéressant pour les applications pratiques d'exploiter au maximum des scènes et des éléments réels, pour n'utiliser des objets virtuels que là ou ils sont vraiment indispensables (ce qu'avait déjà commencé à illustrer l'équipe d'A. Fellous [6]).

Références

[1] Feiner (S.), MacIntyre (B.) et Seligmann (D.). – Knowledge based augmented reality. *Communications of the ACM*, vol. 36, juillet 1993.

[2] Caudel (T.) et Mizell (D.W.). – Augmented reality: An application of heads-up display technology to manual manufacturing processes. *In : Hawaii International Conference on System Sciences*.

[3] Ahlers (K. H.), Kramer (A.), Breen (D. E.), Chevalier (P. Y.), Crampton (C.), Rose (E.), Tuceryan (M.), Whitaker (R. T.) et Greer (D.). – Distributed augmented reality for collaborative design applications. *In : Eurographics'95*, pp. 3–14.

[4] Whitaker (R. T.), Crampton (C.), Breen (D. E.), Tuceryan (M.) et Rose (E.). – Object calibration for augmented reality. *In : Eurographics'95*, pp. 15–27.

[5] Bajura (M.), Fuchs (H.) et Ohbuchi (R.). – Merging virtual objects with the real world: seeing ultrasound imagery within the patient. *Computer Graphics (SIGGRAPH '92 Proceedings)*, juillet 1992, pp. 203–210.

[6] Fellous (A.). – The STV-synthetic TV system. *In : Journees INRIA Analyse-Synthèse d'images*, pp. 28–35. – ENST, Paris, janvier 1994.

[7] Fournier (A.), Gunawan (A. S.) et Romanzin (C.). – Common illumination between real and computer generated scenes. *In : Proceedings of Graphics Interface '93*. Canadian Information Processing Society, pp. 254–262. – Toronto, Ontario, Canada, mai 1993.

[8] Fournier (A.). – Illumination problems in computer augmented reality. *In : Journees INRIA*

Analyse-SynthŁse d'images, pp. 1–21. – ENST, Paris, janvier 1994.

[9] Horaud (Radu) et Dornaika (Fadi). – Hand-eye calibration. *Int. Journal of Robotics Research*, vol. 14, Spring 1995. – A paratre.

[10] Tarel (J.P.) et Gagalowicz (A.). – Calibration de camØa base d'ellipses. *Traitement du Signal*, vol. 12, n 2, 1995, pp. 177–187. – http://www-rocq.inria.fr/syntim/textes/calib-eng.html.

[11] Boufama (B.), Mohr (R.) et Veillon (F.). – Euclidean constraints for uncalibrated reconstruction. *In : IEEE International Conference on Computer Vision*, pp. 466–470. – Berlin, March 1993.

[12] Tarel (J.P.). – Une mØthode de calibration radiomØtrique de camØa focale variable. *In : 10Łme congrŁs AFCET, Reconnaissance des Formes et Intelligence Artificielle*. – Rennes, France, 1996.

[13] Vinet (L.). – *Segmentation et mise en correspondance de rØgions de paires d'images stØrØoscopiques*. – ThŁse de PhD, UniversitØ Paris-IX Dauphine, 1991.

[14] Sander (Peter T.), Vinet (Laurent), Cohen (Laurent) et Gagalowicz (AndrØ). – Hierarchical regions based stereo matching. *In : Proceedings of the Sixth Scandinavian Conference on Image Analysis*, pp. 71–78. – Oulu, Finland, June 1989.

[15] Randriamasy (S.). – *Segmentation descendante coopØrative en rØgions de paires d'images stØrØoscopiques*. – ThŁse de PhD, UniversitØ Paris-IX Dauphine, 1992.

[16] Randriamasy (S.) et Gagalowicz (A.). – Region based stereo matching oriented image processing. *In : Proceedings of Computer Vision and Pattern Recognition*. – Maui, Hawaii, June 1991.

[17] Boujemaa (Nozha), Stamon (Georges) et Gagalowicz (AndrØ). – ModØlisation floue pour la segmentation d'images. *In : 9Łme congrŁs AFCET, Reconnaissance des Formes et Intelligence Artificielle*.

[18] Ackah-Miezan (A.) et Gagalowicz (A.). – Discrete models for energy minimizing segmentation. *In : Proceedings of the 3rd International Conference on Computer Vision*, pp. 200–207. – 11-13 May, Berlin, 1993. http://www-rocq.inria.fr/syntim/textes/iccv93-eng.html.

[19] VØlen (J.M.). – *Techniques de reconstruction globale par analyse de paires d'images stØrØoscopiques*. – ThŁse de PhD, UniversitØ Paris-VII, 1995.

[20] Tarel (Jean-Philippe) et VØlen (Jean-Marc). – A generic approach for planar patches stereo reconstruction. *In : Proceedings of the Scandinavian Conference on Image Analysis*, pp. 1061–1070. – Uppsala, Sweden, 1995. http://www-rocq.inria.fr/syntim/textes/scia95-eng.html.

[21] Koivunen (V.), VØlen (J-M.) et Bajcsy (R.). – Procedural cad-models from range data. *In : Proceedings of Towards World Class Manufacturing '93*.

[22] Zhang (Z.). – Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, vol. 13, n 2, 1994, pp. 119–152. – also INRIA Tech. Report #1658.

[23] Tarel (J.P.) et Boujemaa (N.). – Une approche floue du recalage 3d : gØnØricitØ et robustesse. *In : 10Łme congrŁs AFCET, Reconnaissance des Formes et Intelligence Artificielle*. – Rennes, France, 1996.

[24] Serfaty (V.), Ackah-Miezan (A.), Lutton (E.) et Gagalowicz (A.). – Photometric analysis as an aid to 3D reconstruction of indoor scenes. *In : Image Modeling, L.A. Ray, J.R. Sullivan*, pp. 196–207. – Editors, proc. SPIE 1904.

[25] Foley (J. D.), van Dam (A.), Feiner (S. K.) et Hughes (J. F.). – *Computer Graphics: Principles and Practices (2nd Edition)*. – Addison Wesley, 1990.

[26] Watt (A.) et Watt (M.). – *Advanced Rendering and Animation Techniques: Theory and Practice*. – Addison-Wesley, 1991.

[27] Meas-Yedid (V.), Tarel (J.P.) et Gagalowicz (A.). – Calibration mØtrique faible et construction interactive de modŁles 3D de scŁnes. *In : CongrŁs Reconnaissance des Formes et Intelligence Artificielle, (RFIA'94)*. AFCET. – Paris, France, 1994.

[28] Provot (X.). – Deformation constraints in a mass-spring model to describe rigid cloth behavior. *In : Proceedings of Graphic Interface'95*, pp. 147–154. – QuØbec, Canada, mai 1995.