Réalité Enrichie par Synthèse

Pierre Jancène, Christophe Meilhac, Fabrice Neyret,
Xavier Provot, Jean-Philippe Tarel,
Jean-Marc Vézien, 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

Abstract

Nous proposons une méthode pour automatiser l'insertion cohérente d'objets 3D de synthèse dans des séquences d'images réelles. Celle-ci fait appel conjointement à des techniques d'analyse et de synthèse d'images pour calculer les images modifiées.
Mots-clés:
analyse d'images, synthèse d'images,réalité virtuelle .

1 Introduction

1.1 Etat de l'Art

La réalité augmentée prend de l'ampleur dans différents domaines comme les processus de fabrication pour l'industrie [1, 2], l'architecture d'intérieur [3, 4] et le domaine médical [5]. Dans la plupart de ces applications, le problème principal est d'assurer la précision de la superposition des images calculées sur les images réelles.

En effet quand on insère des objets 3D de synthèse dans une séquence d'images pour des applications de vidéo, il faut s'assurer de la cohérence de l'illumination, des ombres et des occultations des objets virtuels par les objets réels. Il existe plusieurs approches en réponse à ce problème :

- Les applications actuelles d'effets spéciaux pour la production vidéo superposent simplement l'image de synthèse devant l'image réelle (incrustation ou *alpha keying*). Si l'image de synthèse doit être partiellement occultée, un masque 2D doit alors être fourni, ce qui interdit l'animation ou contraint à interpoler l'évolution du masque au cours du temps.

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.

- La *Synthetic TV* [6], garantit la cohérence géométrique entre images réelles et objets de synthèse en synchronisant les mouvements de la 'caméra' de synthèse, correctement modélisée, avec les mouvements de la caméra réelle.
- La *Computer Augmented Reality* [7, 8], s'attache au calcul de l'illumination globale dans une scène réelle enrichie d'objets de synthèse.

Ces méthodes nécessitent une forte interaction avec l'utilisateur lors de l'opération de mixage. Notre but est d'automatiser partiellement ce traitement.

1.2 Notre Approche

La Réalité enrichie par Synthèse (RES) a pour but l'insertion d'objets 3D de synthèse dans une séquence 2D d'images réelles, en assurant la cohérence 3D pour les occultations, les ombres, les collisions et les contacts. L'idée consiste à reconnaître 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 synthèse d'images via la constitution de modèles 3D synthétisables compatibles, autorisant un mixage 3D-cohérent des objets de synthèse dans les images 2D.

Les applications de cette technique peuvent se trouver en visualisation (par exemple en imagerie médicale), et pour la production vidéo (effets spéciaux pour les films, publicité, étude d'impact...).

Un modèle synthétisable décrit une scène composée d'objets, de sources de lumières et d'une caméra. Les objets ont une forme, une position, une orientation, des propriétés de surface (ex : couleurs ambiante, diffuse et spéculaire) et éventuellement une évolution temporelle (mouvement rigide ou déformation). Un modeleur est un outil interactif permettant à un utilisateur de décrire un modèle synthétisable 3D.
Les traitements effectués pour la RES sont effectués en différé :

- Les objets réels choisis pour l'interaction sont préalablement modélisés à l'aide du modeleur, qui dispose

de fonctionnalités spécifiques pour 'décalquer' interactivement les objets réels à partir de vues sous différents angles. Cet outil permet de constituer une base de modèles d'objets susceptibles d'être présents dans la scène à 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 à reconstituer une scène synthétique les remettant aux places et orientations identiques à celles de la scène originale. Les caractéristiques de la caméra sont également estimées.
- L'utilisateur spécifie sa scène animée de synthèse à 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 à 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 à 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, à partir d'informations dans les images.

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 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 lumières sous la forme de conditions d'éclairage équivalentes (section 2.5).

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

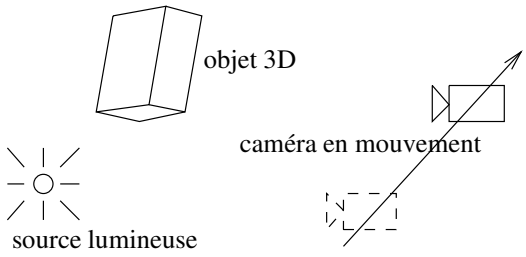

Figure 1: Qu'est ce qu'une scène?

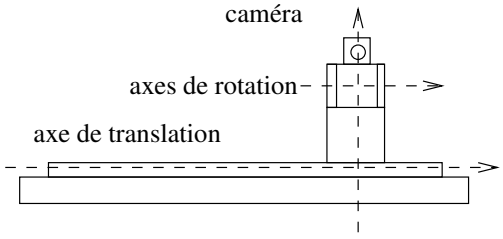

Figure 2: Le portique du laboratoire.

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 à dire la possibilité d'ajouter des objets de synthèse dans une scène réelle en interaction 3D réaliste, de manière à 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'à des scènes statiques, il suffit de faire l'acquisition à la cadence de 2 à 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 pré-existantes calibrées, et de procéder à des opérations d'analyse complètes donc coteuses. 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 à la synthèse d'images sur un objet, à 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, où 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 à 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 à 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 à 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-à-dire la connaissance de la transformation que subit un point de vue par les paramètres de contrle, est réalisé par processus hors-ligne grâce à 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 à la direction de visée,
- un twist, qui fixe l'angle de l'un des ctés de l'image avec l'horizontale,
- une focale et
- une distance de mise au point, les objets qui ne sont pas à 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 grâce à 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 à un repère lié à celle-ci, et on extrait dans l'image de calibration le point image correspondant à 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 d'entre elles

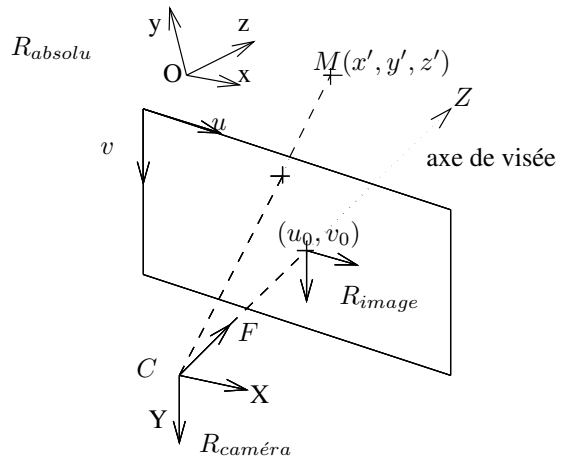

Figure 3: Modèle sténopé d'une caméra.

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 décrivant 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 relâchant 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) calculé 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 où 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 plutt 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 paramétré que possible (par exemple, ne dépendant que d'un paramètre d'échelle), afin d'automatiser la procédure. De

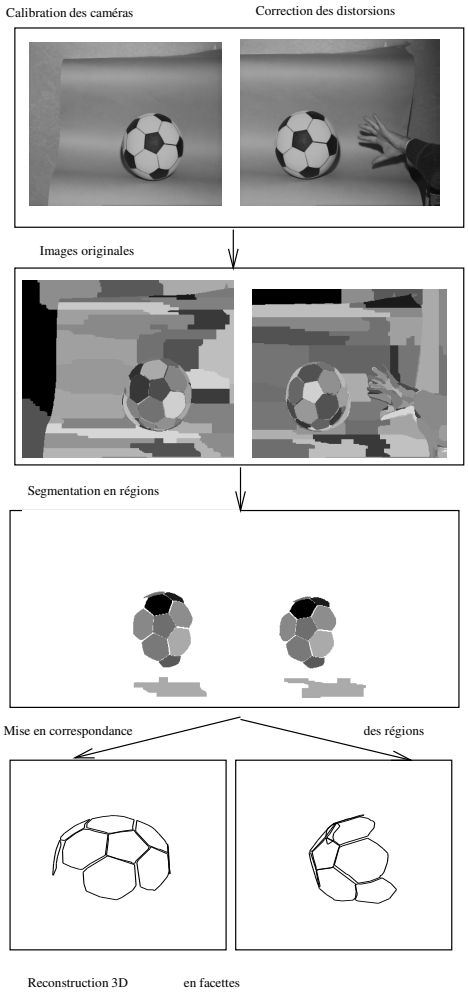

Figure 4: Le processus d'analyse stéréoscopique.

ce point de vue, la dernière méthode conduit à 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 lâche 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.

Là 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 détecté 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 à 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 réalise 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 développées et sont utilisées 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'à 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 déplacements possibles [23].

2.5 Analyse des sources lumineuses

L'illumination de la scène est modélisée par un flux lumineux parallèle qui provient d'une source située à l'infini. Elle est ainsi uniquement déterminée par sa direction d'illumination et son intensité lumineuse. Le calcul de l'éclairement sur chaque facette est donc particulièrement simple, ce qui permet, en minimisant l'écart entre la distribution d'intensité prévue et les intensités effectivement observées sur l'intérieur des régions, d'obtenir une estimation de la source équivalente et de réflectance des surfaces. Le même type d'approche a été appliqué à des modèles de sources plus complexes comme une source ponctuelle uniforme équivalente [24]. Evidemment, l'obtention d'une source lumineuse équivalente ne permet pas de faire des estimations ayant une grande valeur du point de vue strictement physique, mais elle fourni des résultats suffisants dans le cadre de la RES.

2.6 Récapitulatif

Nous avons présenté un ensemble d'outils qui permettent d'analyser une séquence d'images réelles dans le but de fournir un modèle géométrique et photométrique de la scène suffisamment riche pour être utilisé dans le cadre de la RES. Nous avons décrit les possibilités du système actuel et les contraintes que cela implique sur la nature des objets observés et du système d'acquisition des images. En particulier, les objets doivent posséder des facettes planes uniformes et ne pas se déplacer dans la séquence. De plus, la caméra et son mouvement doivent être calibrés. Enfin, les sources lumineuses doivent être éloignée des objets éclairés. C'est dans ce contexte, que les processus d'analyse du point de vue, d'extraction de facettes 3D, de positionnement des modèles d'objets, et d'estimation d'éclairage équivalent, précédemment décrits, peuvent fonctionner.

3 Synthèse d'images

3.1 Introduction

Comme indiqué précédemment, le modèle synthétisable d'une scène comprend des objets, des sources de lumière et une caméra (point de vue).

La problématique de l'analyse d'images est de reconstituer un modèle aussi complet que possible à partir d'images 2D. Les problématiques de la synthèse d'images sont l'inverse de celles de l'analyse d'images. Plus précisément, la synthèse d'images s'intéresse au problème direct de la simulation des divers phénomènes qui conduisent de la description d'une scène à la production d'une image [25] :

modélisation : Une première problématique concerne la description géométrique des objets, puis la structuration d'ensembles d'objets. Les modalités de description sont très variées : la modélisation peut être interactive, constructive, analytique, sous contraintes. Il faut étudier non seulement les diverses représentations, mais aussi les procédures conduisant à ces modélisations : déformations globales, 'sculpture' par modifications successives, grammaires.

De même, les types de structures liant les objets sont multiples : hiérarchies, articulations, squelettes. Il existe également nombre de modèles dédiés à des types d'objets particuliers (systèmes de particules, liquides, fumées et nuages, objets complexes, fractales). Le logiciel permettant de constituer cette description est un modeleur géométrique.

animation : Une seconde problématique est liée à l'évolution temporelle de ces objets [26], qui comprend la cinématique, la simulation physique, l'animation des déformations, le contrle du mouvement, la simulation comportementale. Le logiciel permettant de spécifier ces animations est un modeleur d'animation.

rendu : Une troisième problématique s'attache à la simulation des interactions lumineuses et au calcul de l'image proprement dite, et se sépare en deux branches, l'illumination globale et l'illumination locale[1]. L'illumination globale prend en charge la résolution de l'équilibre dans les échanges d'énergie lumineuse, tandis que l'illumination locale s'occupe de l'interaction de la lumière et de la matière (modélisation de la réflectance, des textures...). Le logiciel effectuant le calcul d'une image est un programme de rendu (ou render). Il est généralement contrlé par un processeur de production d'image, qui se charge de reconstituer l'état de la scène à chaque pas de temps puis de post-traiter et de stocker les images après calcul.

3.2 Synthèse d'images pour la RES

3.2.1 construction de modèles pour l'analyse

Notre modeleur ACTION3D est tout d'abord utilisé pour construire interactivement les modèles géométriques des objets réels avec lesquels on souhaite interagir. On utilise plusieurs images de l'objet à 'décalquer' que l'on place dans deux vues sur lesquelles on reprojette le modèle en cours de construction. Un système permet de limiter le placement des nouveaux points successifs dans les plans des faces déjà acquises [27].

3.2.2 construction de modèles pour la synthèse

Le modeleur est par la suite utilisé classiquement pour modéliser et animer les objets de synthèse rigides ou déformés géométriquement. Les objets régis par les lois de la dynamique sont construits par des modules dédiés directement lors de l'exécution du rendu.

Le modeleur ACTION3D permet de :

- construire interactivement des objets polyédriques et des surfaces de forme libre,
- déformer et animer ces objets par des méthodes de déformation de l'espace 3D et par des méthodes de métamorphose 3D,
- construire interactivement les textures, et spécifier leur plaquage sur les surfaces.

[1] Sans compter la modélisation des sources de lumières et des caméras.

3.2.3 modèles dynamiques pour la synthèse

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

3.2.4 calcul du rendu

Comme indiqué plus haut et plus amplement décrit dans la partie 4, quatre images de synthèses partielles sont calculées, puis mixées avec l'image réelle.

Pour le calcul du rendu, nous utilisons le logiciel de rendu du domaine public RAYSHADE, qui opère par lancer de rayons [25] : l'énergie reçue par chaque pixel du plan image est évaluée en lançant quelques rayons via ce pixel à travers la scène. Ces rayons 'remontent' les trajectoires des photons en partant de l'observateur. L'illumination locale est évaluée au point d'impact de chaque rayon sur un objet de la scène, en fonction des propriétés de surface de l'objet et de la lumière parvenant en ce point, selon un modèle de réflectance qui indique quelle proportion de l'énergie est émise vers l'observateur. L'énergie incidente est elle-même évaluée en lançant un rayon vers les sources de lumière, afin de tester les occultations (générant les ombres). Un rayon est également lancé dans la direction de réflection donnée par les lois de Descartes, afin de collecter l'énergie d'un éventuel reflet.

4 Mixage (*compositing*)

En incrustant (figure 5) l'image de synthèse sur l'image réelle, l'occultation d'objets réels par des objets virtuels était déjà effective dès les premières approches pratiquées en vidéo. Un masque associé à l'image de synthèse indique les pixels occupés (avec une résolution subpixel), contrlant le mixage des images. Celui-ci peut être obtenu en effectuant un calcul de rendu après avoir peint les objets en blanc uniforme (i.e. isotrope et non lambertien).

En fusionnant les modèles reconnus avec les modèles de synthèse, l'occultation des objets virtuels par les objets réels est automatique avec notre représentation. Il faut cependant en tenir compte pour la génération du masque : les objets reconnus ne doivent pas apparaître dans l'image de synthèse ni dans le masque, dans la mesure où ils sont déjà représentés dans l'image réelle, mais ils peuvent néanmoins cacher en partie des objets de synthèse. Ceci est obtenu en peignant les objets reconnus en noir et les objets virtuels en blanc uniforme. Les objets reconnus sont alors à la fois invisibles et occultants, et peuvent éventuellement être réfléchissants.

Les ombres (figure 6) des objets réels sur les objets virtuels apparaissent automatiquement avec notre représentation unifiée. Les ombres des objets virtuels sur les objets réels sont obtenues en calculant un coefficient d'atténuation, à multiplier à l'image originale. Celui-ci est évalué en effectuant un rendu des objets reconnus en blanc lambertien, avec et sans les autres objets, afin d'évaluer l'effet d'ombrage

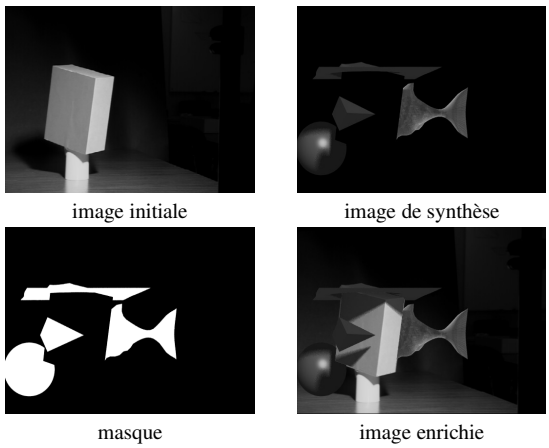

Figure 5: Noter la trace des objets réels reconnus (la boîte), invisibles mais ombrants et occultants.

d aux objets virtuels. Le facteur d'atténuation est le ratio de l'image avec les ombres et de l'image de référence contenant seulement les objets reconnus. (Cette valeur obtenue pour tout point de l'image est également calculée avec une résolution subpixel.)

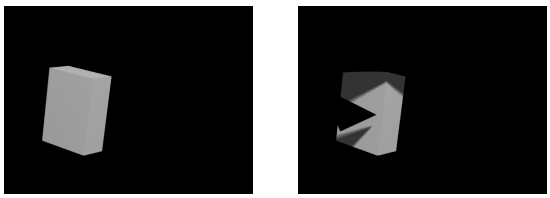

Figure 6: objets reconnus, seuls ou avec les objets de synthèse ombrants. L'assombrissement est donné par le ratio des images.

Le mixage (figure 5) est alors défini par :

image enrichie = image réelle $\times$ atténuation $\times$ (1-masque) + image de synthèse. [2]

5 Exemples

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

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

La figure 8 présente moins de cas d'interaction, mais illustre la possibilité de rendre réfléchissant des objets réels. De plus, l'animation de la séquence permet de 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).

[2] on remarquera que l'image de synthèse contient : objets de synthèse $\times$ masque + objets reconnus en noir réfléchissants.

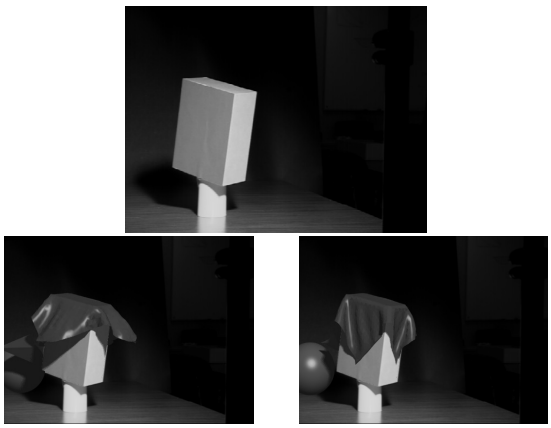

Figure 7: Image originale et deux images de la séquence enrichie par un drap de synthèse rouge mobile, un tetraèdre bleu et une sphère verte.

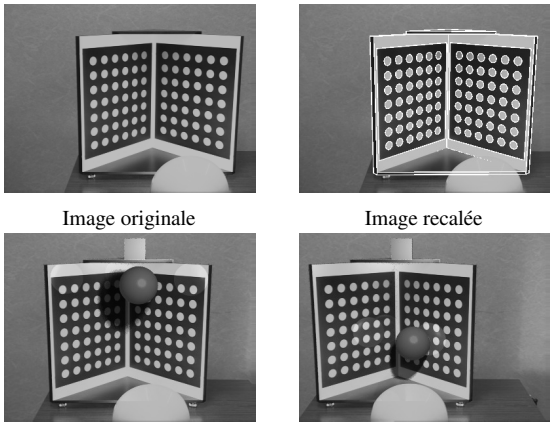

Image originale Image recalée

Figure 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 œuvre, et de poser de nouveaux problèmes.

Diverses directions de recherche sont ouvertes pour relâcher 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 relâcher 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 où ils sont vraiment indispensables (ce qu'avait déjà commencé à illustrer l'équipe d'A. Fellous [6]).

References

[1] Steven Feiner, Blair Macintyre, and Dorée Seligmann. Knowledge-based augmented reality. *Commun. ACM*, 36(7):53–62, July 1993.

[2] T.P. Caudell and D.W. Mizell. Augmented reality: an application of heads-up display technology to manual manufacturing processes. In *System Sciences, 1992. Proceedings of the Twenty-Fifth Hawaii International Conference on*, volume ii, pages 659 –669 vol.2, jan 1992.

[3] Klaus H. Ahlers, André Kramer, David E. Breen, Pierre-yves Chevalier, Douglas Greer, Chris Crampton, Eric Rose, Mihran Tuceryan, and Ross T. Whitaker. Distributed augmented reality for collaborative design applications. In *Proc. of Eurographics'95 Conference, (Maastricht, NL*, pages 3–14, 1995.

[4] Ross T. Whitaker, Chris Crampton, David E. Breen, Mihran Tuceryan, and Eric Rose. Object calibration for augmented reality. In *Proc. of Eurographics'95 Conference, (Maastricht, NL*, pages 15–27, 1995.

[5] Michael Bajura, Henry Fuchs, and Ryutarou Ohbuchi. Merging virtual objects with the real world: seeing ultrasound imagery within the patient. *SIGGRAPH Comput. Graph.*, 26(2):203–210, July 1992.

[6] Armand Fellous. Stv-synthetic tv: from laboratory prototype to production tools. In Nadia Magnenat Thalmann and Daniel Thalmann, editors, *Virtual worlds and multimedia*, pages 127–133. John Wiley & Sons, Inc., New York, NY, USA, 1993.

[7] Alain Fournier, Atjeng S. Gunawan, and Chris Romanzin. Common illumination between real and computer generated scenes. Technical report, University of British Columbia, Vancouver, BC, Canada, Canada, 1992.

[8] Alain Fournier. Illumination problems in computer augmented reality, 1994.

[9] R. Horaud and F. Dornaika. Hand eye calibration. *International Journal of Robotics Research*, 14(3):195–210, June 1995.

[10] Jean-Philippe Tarel and André Gagalowicz. Calibration de caméra à base d'ellipses. *Traitement du Signal*, 12(2):177–187, 1995.

[11] B. Boufama, R. Mohr, and F. Veillon. Euclidean constraints for uncalibrated reconstruction. In *Fourth International Conference on Computer Vision (Berlin, Germany, May 11–14, 1993)*, pages 466–470, 1993.

[12] Jean-Philippe Tarel. Une méthode de calibration radiométrique de caméra à focale variable. In *10ème congrès AFCET, Reconnaissance des Formes et Intelligence Artificielle*, Rennes, France, 1996.

[13] L. Vinet. *Segmentation et mise en correspondance de régions de paires d'images stéréoscopiques*. PhD thesis, Université Paris-IX Dauphine, 1991.

[14] L. Cohen, L. Vinet, P. T. Sander, and A. Gagalowicz. Hierarchical region based stereo matching. In *CVPR'89 (IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, June 4–8, 1989)*, pages 416–421, Washington, DC., June 1989. Computer Society Press.

[15] S. Randriamasy. *Segmentation descendante coopérative en régions de paires d'images stéréoscopiques*. PhD thesis, Université Paris-IX Dauphine, 1992.

[16] S. Randriamasy and A. Gagalowicz. Region based stereo matching oriented image processing. In *CVPR'91 (IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Lahaina, Maui, HI, June 3-6, 1991)*, Washington, DC., June 1991. Computer Society Press.

[17] N. Boujemaa, G. Stamon, and A. Gagalowicz. Modélisation floue pour la segmentation d'images. In *9ème congrès AFCET, Reconnaissance des Formes et Intelligence Artificielle*, 1994.

[18] A. Ackah-Miezan and A. Gagalowicz. Discrete models for energy minimizing segmentation. In *Fourth International Conference on Computer Vision (Berlin, Germany, May 11–14, 1993)*, pages 200–207, 1993.

[19] J.-M. Vézien. *Techniques de reconstruction globale par analyse de paires d'images stéréoscopiques*. PhD thesis, Université Paris-VII, 1995.

[20] Jean-Philippe Tarel and Jean-Marc Vézien. A generic approach for planar patches stereo reconstruction. In *Proceedings of the Scandinavian Conference on Image Analysis*, volume 2, pages 1061–1070, Uppsala, Sweden, 1995. Swedish Society for Automated Image Analysis.

[21] V. Koivunen and J.-M. Vézien. Machine vision tools for CAGD. *International Journal of Pattern Recognition and Artificial Intelligence*, 10(2):165–182, 1996.

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

[23] Jean-Philippe Tarel and Nozha Boujemaa. Une approche floue du recalage 3D : généricité et robustesse. In *10ème congrès AFCET, Reconnaissance des Formes et Intelligence Artificielle*, Rennes, France, 1996.

[24] V. Serfaty, A. Ackah-Miezan, E. Lutton, and A. Gagalowicz. Photometric analysis as an aid to 3D reconstruction of indoor scenes. In *Image Modeling, L.A. Ray, J.R. Sullivan*, pages 196–207, 1993. Editors, proc. SPIE 1904.

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

[26] Alan Watt and Mark Watt. *Advanced animation and rendering techniques*. ACM, New York, NY, USA, 1991.

[27] Vannary Meas-Yedid, Jean-Philippe Tarel, and André Gagalowicz. Calibration métrique faible et construction interactive de modèles 3D de scènes. In *Congrès Reconnaissance des Formes et Intelligence Artificielle*, Paris, France, 1994. AFCET.

[28] Xavier Provot. Deformation constraints in a mass-spring model to describe rigid cloth behavior. In *In Graphics Interface*, pages 147–154, 1995.