REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Hadj Lakhder- BATNA
Faculté des Sciences de l'Ingénieur
Département de l'Informatique

Mémoire présenté en vue d'obtenir le titre du

MAGISTER EN INFORMATIQUE

Option : ***Informatique Industrielle***

Synthèse de Nouvelles Vues pour les Applications en Réalité Augmentée

Présenté par :

Dib Abderrahim

Dirigé par :

Pr. Batouche M^{ed} Chowki

Composition du jury :

Président :

Dr. Zidani Abdelmadjid MC, Université de BATNA

Rapporteur :

Pr. Batouche M^{ed} Chowki Pr, Université Mentouri de Constantine

Examinateurs :

Dr. Meshoul Souham MC, Université Mentouri de Constantine

Dr. Kholladi M^{ed} Khireddine MC, Université Mentouri de Constantine

Mémoire préparé au sein du laboratoire LIRE , Université Mentouri de Constantine

Juin 2006


بسم الله الرحمن الرحيم

Remerciements

Je tiens à remercier sincèrement Monsieur Batouche M^{ed} Chowki, Professeur à l'Université Mentouri de Constantine, d'avoir accepté de m'encadrer et de m'avoir accueilli au sein de son équipe de recherche au laboratoire LIRE. J'ai beaucoup bénéficie de ses précieux conseils, de ses suggestions pertinentes pour la réalisation de ce travail.

Je tiens à exprimer ma reconnaissance à Monsieur Mourad Bouzenada, Maître assistant à l'université Mentouri de Constantine, pour m'avoir dirigé dans mon travail. Son aide tant scientifique que personnelle, ainsi que ses conseils ont soutenu mon travail de recherche.

Je tiens également à remercier les membres du jury qui m'ont fait l'honneur de bien vouloir évoluer mon travail, et plus précisément :

Monsieur Zidani Abdelmadjid, Maître de Conférence à l'Université Hadj Lakhder de BATNA, pour l'honneur qu'il me fait, en acceptant la présidence de ce jury.

Madame Meshoul Souham, Maître de Conférence à l'université Mentouri de Constantine et Monsieur Kholladi Mohamed Khireddine, Maître de Conférence à l'université Mentouri de Constantine, auxquels je suis très reconnaissant d'avoir accepté d'être examinateurs de ce travail.

Et finalement, je n'oublis pas à remercier tous ceux qui ont participé directement ou indirectement à l'aboutissement de ce travail. Je remercie en particulier :

Tous les membres de ma famille.

Mes chers amis : B. Hocine, G. Hichem, K. Frouk, A. Yacer, K. Fateh, Z. Lamine.
T. Halim,T. Hacene, B. Djamel, Z. Karim, B.C. Madjed, L. Kamel, B. Menseur.

Et à toute personne qui m'est cher.

Dib Abderrahim

T a b l e d e s m a t i è r e s

Remerciement …….....…….....……. ***I***

Table des matières et des figures……..……....…….....…. ***II***

Résumé…….……....…….....……. ***XIII***

Chapitre I :

Introduction

1. Introduction……..…….....…….....**02**

2. Synthèse d'images classique (à base de modèles géométriques) ….....… 02

2.1 Phase de modélisation……..…….....…….....…….....02

2.2 Phase de rendu……...……...…….....…….....…... 03

3. Synthèse de nouvelles vues à partir d'images réelles…...…….....…….....05

4. Applications ……...…...….....08

4.1 Entraînement……....…….....…….....08

4.2 Simulation…….……....…...…...…....09

4.3 Commerce, conservation, loisirs……....…….....… 09

4.4 Compression de données……..…....……...….....…...……..…. 09

5. Organisation du mémoire…….....10

Chapitre II :

Synthèse de nouvelles vues à partir d'images réelles

(IBMR) : Principes et méthodes

1. Introduction …….....……..…….....…….....13

1.1 Les méthodes actives ……..…….....…….... 13

1.2 Les méthodes passives ……....…….....…….... 17

1.2.1 La stéréovision…….....…...……..… 17

1.2.2 Système de stéréovision…….....…….... 18

2. La Géométrie de la vision…….…….....….....…… **19**

2.1 Systèmes de coordonnées…… 20
2.2 Coordonnées homogènes et transformations géométrique…… 20
2.2.1 Coordonnées homogènes…… 21
2.2.2 Les transformations élémentaires que subissent les objets dans l'espace 3D 21
2.2.3 Les Homographies…… 22
2.3 Espaces Euclidien, métrique, affine et projectif…… 23
2.4 Modèle de la camera et la formation de l'image…… 26
2.4.1 Modèle de la caméra…… 26
2.4.2 Limitations du modèle de trou d'épingle…… 27
2.5 Formation de l'image…… 27
2.6 Géométrie épipolaire et relation entre deux images…… 30
2.6.1 La contrainte épipolaire…… 30
2.6.2 La matrice fondamentale…… 30
2.6.3 La matrice essentielle…… 32
2.7 Calibrage …… 33
2.7.1 Calibrage d'une caméra…… 33
2.7.1.1 Calibrage classique…… 33
2.7.2.2 Calibrage automatique ou auto-calibrage…… 34
3. La mise en correspondance…… 34
3.1 Correspondance stéréo…… 35
3.2 Primitives stéréoscopiques…… 36
3.3 Contraintes de la mise en correspondance…… 36
3.3.1 Contraintes géométriques…… 36
3.3.1.1 Contrainte épipolaire…… 36
3.3.1.2 Contrainte de limites de disparité…… 37
3.3.1.3 Contrainte d'ordre…… 38
3.3.1.4 Contrainte d'unicité …… 38
3.3.2 Contraintes figurales…… 39
3.3.2.1 Disparité locale constante…… 39
3.3.2.2 Continuité figurale …… 39
3.4 Les méthodes de la mise en correspondance…… 40
3.4.1 Méthodes locales d'appariement…… 41
3.4.1.1 Mesures de similarité…… 41

3.4.1.2 Méthodes de gradient…. 43
3.4.2 Forces et lacunes des méthodes locales…. 43
3.4.3 Méthodes globales d'appariement…. 44
3.4.3.1 Programmation Dynamique…. 45
3.4.3.2 Max-flow / min-cut…. 46
4. La reconstruction…. 48
5. Les méthodes de rendu et modélisation à base d'images IBMR…. 50
5.1 Techniques de rendu purement à base d'images…. 51
5.1.1 Imposteurs…. 51
5.1.2 Forme à partir de X (Shape from X) …. 52
5.1.2.1 Forme à partir de contour (Shape from contour) …. 52
5.1.2.2 Forme à partir de l'ombrage (shape from shading) …. 53
5.1.2.3 Forme à partir de silhouette (shape from silhouette) …. 54
5.1.2.4 Forme à partir de texture (shape from texture) …. 55
5.1.3 Morphing/Interpolation…. 56
5.1.3.1 Le morphing…. 56
5.1.3.2 Interpolation de point de vue (View Interpolation) …. 57
5.1.3.3 Déformation de point de vue (View Morphing) …. 58
5.1.4 Interpolation de rayons lumineux et fonction plénoptique…. 58
5.1.4.1 Fonction plénoptique…. 59
5.1.4.2 Modélisation plénoptique (Plenoptic Modeling) …. 59
5.1.4.3 Light Fields et Lumigraphes…. 61
5.1.4.4 Mosaïques concentriques (Concentric mosaic) …. 63
5.1.4.5 Panoramas (image mosaicing) …. 64
5.1.5 Modélisation volumétrique de scène (volumetric scene medeling) …. 64
5.1.5.1 Coloration de voxel (Voxel coloring) …. 65
5.1.5.2 Space carving…. 66
5.2 Techniques basées images / basées géométrie …. 67
5.2.1 Transfert d'images classique…. 67
5.2.2 Déformation 3D d'images (3D image warping) …. 68
5.2.3 Images à plans de profondeurs (LDI : Layered Depth Images) …. 69
5.3 Techniques hybrides…. 71
5.3.1 Placage de texture…. 71

5.3.2 Placage de texture en relief…….. 71
5.3.3 *Façade* : une approche hybride entre image et géométrie……..…..…… 72
6. Conclusion…….....…. 74

Chapitre III :

Les techniques d'IBMR Pour Les applications en Réalité Augmentée

1. Introduction…….....…….....…..……...…. 77
2. La réalité augmentée…….....…..…….....…….…. 77
2.1 Intérêts…….....…….....…. 78
2.2 Réalité augmentée contre réalité virtuelle…….....……. 78
2.3 Technologies d'affichage……..…....…….....…. 79
2.3.1 Affichage à base de moniteurs……..…..……...…… 79
2.3.2 HMD optiques……..…….....…… 79
2.3.3 HMD Vidéos……..…...…… 80

3. Rendu et modélisation à base d'images et les applications en réalité augmentée… 81
3.1 Applications……....…...…… 81
3.1.1 Médecine……....….……..…. 81
3.1.2 Design intérieur……..…….…… 82
3.1.3 Effets spéciaux…….....…… 84
3.1.4 Musé virtuel…….....…….....…. 85
4. Conclusion…….....…….....…….....…..…….... 87

Chapitre IV :

Une nouvelle approche de reconstruction d'objets 3D par la combinaison : enveloppe visuelle / Stéréovision

1. Introduction…….....….....…… 89
2. L'enveloppe visuelle…….....…….....…. 90
3. La stéréovision……..….....….…….....…. 91

4. L'approche proposée…… 92
4.1 Système de prise de vues …… 92
4.2 Etapes du processus de reconstruction…… 93
4.2.1 Calcul de l'enveloppe visuelle…… 93
4.2.2 Restriction du champ de recherche d'un correspondant d'un point…. 93
4.2.3 Mise en correspondance et reconstruction…… 94
4.2.4 Placage de texture…… 95
5. Résultats…… 96
6. Conclusions…… 98

Conclusions et Perspectives…… 100

Références bibliographiques…… 104

L i s t e d e s f i g u r e s

Chapitre I

Figure I.1 **:** Modélisation des objets 3D…… 03

Figure I.2 **:** Effets du nombre de polygones sur la précision du modèle 3D…… 03

Figure I.3 *: Rendu des objets 3D*…… 04

Figure I.4 *: Principe du rendu à base d'images*…… 06

Figure I.5 *: Principe de la reconstruction d'un modèle 3D à partir d'image*…… 06

Chapitre II

Figure II.1 *: L'acquisition 3D mécanique par palpeur*…… 14

Figure II.2 *: Télémètre Laser* …… 14

Figure II.3 *: Projection du plan lumineux sur une scène*…… 15

Figure II.4 *: Projection du plan lumineux sur une scène. L'objet est placé sur une table tournante contrôlée par ordinateur et éclairée par deux raies lumineuses.* …… 15

Figure II.5 *: Schématisation de la reconstruction du relief à l'aide de lumière structurée.* 16

Figure II.6 *: Lumière structurée illuminant un visage*…… 16

Figure II.7 *: Impossibilité de reconstruire la scène à partir d'une seule image*…… 17

Figure II.8 *: Elimination de l'ambiguïté par deux images prise depuis des points de vue différents.* …… 18

Figure II.9 *: Système de la stéréovision*…… 18

Figure II.10 *: Schéma simplifier du système de stéréovision*…… 19

Figure II.11 *: système de coordonnées*…… 20

Figure II.12 *: les transformations possibles dans un espace euclidien*…… 23

Figure II.13 *: Un nouveau degré de liberté pour les transformations possibles dans un espace métrique.* …… 24

Figure II.14 *: Exemple de transformation affine d'un cube*…… 25

Figure II.15 *: un exemple d'image en représentation projective*…… 25

Figure II.16 *: (a) Le trou d'épingle et (b) son modèle géométrique*…… 26

Figure II.17 *: Formation de l'image*…… 28

Figure II.18 *: Géométrie épipolaire*…… 30

Figure II.19 *: Géométrie épipolaire et matrice essentielle*…… 33

Figure II.20 *: Une mire de calibrage photographiée sous plusieurs angles* ……..… 34
Figure II.21 *: La correspondance stéréo*……..…..…….....…… 35
Figure II.22 *: Difficultés d'établir des correspondances en stéréo*……...…..…… 35
Figure II.23 *: contrainte épipolaire*……..…..…….....….. 37
Figure II.24*: rectification de la paire stéréo*……..…..….....…… 37
Figure II.25 *: Contrainte d'ordre*……..…..…….....….. 38
Figure II.26 *: Contrainte d'unicité*……..…..…….....….. 39
Figure II.27 *: Contrainte de continuité figurale*…….....…..…… 40
Figure II.28 *: Corrélation : Au haut, les images gauche et droite avec la fenêtre de Référence. Au bas, une vue rapprochée des deux fenêtres appariées*…….... 42
Figure II.29 *: Exemple d'occultation.* ……..….....…….. 44
Figure II.30 *: Un cas où la contrainte de continuité n'est pas respec*……...…..…. 44
Figure II.31 *: Exemple de chemin minimal (en blanc) dans l'espace (x,d). Les niveaux de gris représentent les valeurs de dissimilarité.* ……..…..…..…… 45

Figure II.32 *:* Exemple de résultat obtenu par la méthode de programmation dynamique. On remarque les traits horizontaux irréguliers entre les lignes. ……..... 46
Figure II.33 : *Volume engendré en empilant les plans (x,d) pour chaque ligne de l'image.* ……..…..…….....…… 46
Figure II.34 *: Volume 3D défini en reliant chaque noeud à ses voisins*……..…..…. 47
Figure II.35 *: Volume 3D auquel on ajoute une source et un drain.* ……..…..…… 48
Figure II.36 *: La triangulation*……..…..……..... 49
Figure II.37 *: Etapes de 'tour into the picture'*……..... 52
Figure II.38 *: Les contours permettent souvent d'interpréter les objets d'une scène et sa structure 3D*……..…..…….....… 52
Figure II.39 *: Principe du shape from shading*……..….....…… 53

Figure II.40 *: Deux exemples de surfaces extraites à partir de l'illumination de la scène (shape from shading)* ……..…..…….....…. 53
Figure II.41 *: Principe de la reconstruction de l'enveloppe visuelle : Enveloppe visuelle d'une sphère obtenue avec 3 vues.* ……..……..…..…. 54

Figure II.42 *: (a) Une vue réelle de l'objet, (b) enveloppe visuelle reconstruite à partir*

de ses silhouettes, (c) modèle raffiné par une méthode de stéréovision, (d) modèle 3D reconstruit de l'objet après placage de texture [51]. …… 55

***Figure II.43** : (a) Texture extraite d'une robe, (b) surface reconstruite à partir de la texture*……..…..……..……. 56

***Figure II.44** : Le morphing : points de contrôles en blanc sur l'image initiale et l'image finale* ……..…..……..... 57

***Figure II.45** : résultat donné par « morphman » un logiciel du marché*……...…...… 57

***Figure II.46** : View morphing*……...…...……..... 58

***Figure II.47** :View Morphing de l'image I_0 vers l'image I_1. L'image intermédiaire Is est créée par post-warping de l'image I_s' elle même obtenue par transfert linéaire entre les images I_0' et I_1' .* …….....…..…… 58

***Figure II.48** : La fonction plénoptique décrit toutes les informations de l'image visibles A partir d'un point de vue particulier.* …….....…..... 59

***Figure II.49** : (a) caméra panoramique, (b) principe de la modélisation plénoptique*..... 60

***Figure II.50** : (a) vue panoramique de la scène, (b) nouvelles vues synthétisées*……. 60

***Figure II.51** : le light-slab*……...…...…….....…… 61

***Figure II.52** : Paramétrage de light field*……...….....……. 61

***Figure II.53** : (a)(b) Acquisition des échantillons d'images de référence pour la construction d'un light-slab ,(c) modèles 3D reconstruits.* ……...…..... 62

***Figure II.54** : (a)Prise de vues de la mosaïque concentrique,(b) deux vues rendues [54]*... 64

***Figure II.55** : images panoramiques*……...….....……... 64

***Figure II.56** : Calibrage et prise de vues [57]* ……..... 65

***Figure II.57** : (a) Voxel coloring : étant donnée un ensemble de vues (images) et une grille de voxels l'objectif est d'attribuer des valeurs de couleur aux voxels de telle sorte que ces derniers soient consistant avec toutes les images. (b) Exemple : Le voxel A est cohérent mais B ne l'est pas.* …… 66

***Figure II.58** : Reconstruction d'une fleur par voxel coloring utilisant 16 images : (a) vue réelle de la fleur, (b) trois vues du modèle 3D reconstruit[57]* …… 66

***Figure II.59**: Transfert d'images : (a)(b) images d'une paire stéréo,(c) carte de profondeur, (d)(e) deux nouvelles vues synthétisées.* ……..…..…… 67

***Figure II.60** : 3D image warping*……...…...……..... 68

***Figure II.61**: (a) image avec profondeurs : vue d'un modèle 3D d'une cathédrale, (b)*

(c) deux nouvelles vues obtenues par 3D image warping……..…..…. 69
***Figure II.62**: Layered Depth Images* ……..…..……..…….. 69
***Figure II.63**: Un modèle de dinosaure reconstruit par LDI à partir de 21 images*…… 70
***Figure II.64**: Les étapes du placage de texture en relief*……..…..……..... 71
***Figure II.65** : (a) une texture et sa carte de profondeur : texture en relief , (b) placage de texture classique, (c) Placage de texture en relief à partir du même point de vue que de (b).* ……..…..…….....…….. 72
***Figure II.66**: (a) Objet représenté par six textures en relief, (b) vue de l'objet reconstruit par placage de texture en relief.* ……..... 72
***Figure II.67**: des vues du Campanile à reconstruire*……..…..…..... 73
***Figure II.68**: éléments dans façade : les blocs*……..…..….....…… 73
***Figure II.69**: (a) modèles de bloc, (b) Modèle recouvré, (c) Marquage des bords et des conteurs sur la vue réelle, (d) projection des bords du modèle recouvré sur la vue réelle, (e) le modèle du campanile synthétisé.* …..……..…..… 74

Chapitre III

***Figure III.1** : À gauche scène réelle, à droite scène augmentée*……..…..…… 77
***Figure III.2** : (a) Annotations sur des voitures de course dans une diffusion en directe, (b) un simple scénario d'une table ronde dans le domaine de la planification urbaine* ……..…..……..... 78
***Figure III.3** : Affichage à base de moniteurs*……..….....…… 79
***Figure III.4** : Schéma d'un HMD optiques*……..…..……..... 80
***Figure III.5** : Schéma d'un HMD vidéo*……..…..……..... 80
***Figure III.6** : affichage des organes intérieurs d'un patient*……..…..…… 81
***Figure III.7** : estomac reconstruit à partir d'images*……..…..….....…….. 82
***Figure III.8** : Interface permettant d'insérer des meubles virtuels dans une photographie[30].* ……..…..…..... 83
***Figure III.9** : Système d'acquisition combinant les techniques passives et actives*…… 83
***Figure III.10** : (a) trois objets reconstruits, (b) objets reconstruits incrustés dans un environnement réel (L'appareil photo et la plante ne font pas partie des scènes originales)* ……..….....…….. 84
***Figure III.11** : Certains effets spéciaux de la trilogie Star Wars utilisent des images de synthèse superposées aux images réelles*…….....……..…..…….. 84

Figure III.12 *:* (a) montage du système de prise de vues, (b) deux vues de la scène reconstruite…….…..…….... 85

Figure III.13 *: Annotations sur des anciens objets réels*……..…..….....…… 86

Figure III.14 *: (a) la statue de David. (b) Modèles 3D des objets reconstruits par le projet Michelangelo* ……..…..…….....…….....…. 86

Chapitre IV

Figure IV.1 *: Principe de la reconstruction de l'enveloppe visuelle*……....…..…..… 90

Figure IV.2 *: principe de la stéréovision*……..…..…….....…….. 92

Figure IV.3 *: système de prise de vues*……..…..….....…….. 92

Figure IV.4 *: Restriction du champ de recherche Segment en rouge sur l'image de droite*……..…..…….....…….. 94

Figure IV.5 *: Direction de creusage pour un point reconstruit.* ……..…..…… 95

Figure IV.6 *:* Prise des vues avec 3D StudioMax.....…… 96

Figure IV.7 : Quatre vues de l'objet et les silhouettes correspondantes..... 97

Figure IV.8 *:* (a) Deux images stéréo, (b) carte des disparités, (c) carte des profondeurs correspondante, (d) représentation en fil de fer des données de la carte des profondeurs sous Matlab. 97

Figure IV.9 *: Enveloppe visuelle reconstruite*.....… 98

Figure IV.10 *: Résultat final de la reconstruction*.....…… 98

L i s t e d e s t a b l e a u x

Tableau II.01 *: Quelques mesures de similarité*..... 42

Tableau II.02 *: Types d'images et leurs composants : R=Rouge, V=Vert, B=Bleu, P=Profondeur ...* .....… 70

Résumé

La synthèse d'images a pour but de calculer des vues aussi réalistes que possible d'une scène tridimensionnelle définie par un modèle géométrique 3D, augmentée de certaines informations photométriques : couleurs, textures, matériaux, et nature de leurs interactions avec la lumière. Classiquement, pour ces applications, il est nécessaire d'effectuer une première étape de modélisation manuelle de chaque élément de la scène à synthétiser, puis une étape de rendu pour générer les images finales de cette scène. Ce type de synthèse présente des limitations en terme de temps de modélisation et de qualité des résultats.

Pour remédier à ces problèmes, ce qui est proposé est de définir la scène non pas par un modèle géométrique tridimensionnel, mais par des vues (bidimensionnelles) réelles de cette scène dans le but de synthétiser de nouvelles vues uniquement à partir des vues de départ en simulant le déplacement de la caméra qui a pris les vues réelles. Les techniques permettant de synthétiser de nouvelles vues à partir des vues réelles d'une scène sont appelés communément les méthodes de modélisation et rendu basés image (*Image Based Modeling and Rendering* : *IBMR*).

Les techniques d'IBMR ont trouvé des applications passionnantes dans plusieurs domaines, dont la réalité augmentée, qui consiste à augmenter la perception visuelle du monde réel par l'insertion réaliste d'objets visuels dans un environnement réel. Le but d'utiliser les techniques d'IBMR dans les applications de la réalité augmentée est d'améliorer la modélisation d'environnements augmentés, tant au niveau de la précision et de la rapidité de conception, qu'au niveau du réalisme.

Dans ce contexte de travail nous avons proposé une approche qui combine deux méthodes d'IBMR pour la reconstruction d'objets 3D réalistes. La première est la reconstruction à partir d'images stéréo et la deuxième est une technique appelée 'enveloppe visuelle'. Ces deux méthodes sont complémentaires en nature. la technique de l'enveloppe visuelle est une première forme de l'objet qui limite au minimum l'espace englobant cet objet, ce qui aide les algorithmes stéréo à éviter des calculs inutiles pour des endroits en dehors du volume de l'objet. Les méthodes de la stéréovision raffinent le modèle reconstruit de l'objet par la détection des points et des régions concaves sur la surface de l'objet.

Mots clés : reconstruction 3D, rendu et modélisation basés image, stéréovision, mise en correspondance, enveloppe visuelle, réalité augmentée.

Abstract

The aim of image synthesis is to compute realistic views of a three-dimensional scene, defined by a geometric 3D model, along with some photometric information: colour, texture, material, and their interaction with light.
Classically, for these applications, it is usually necessary to perform a first step consisting in manually modelling of every element of the modelled scene and then a rendering step witch generates the final views of this scene. This type of synthesis presents several limitations in term of modelling time and results quality.

To solve these problems, which is proposed to define a scene not by a 3D geometric model, but by real views (two-dimensional) of that scene in order to synthesize novel views using only the starting views by simulating the displacement of the camera which took the real views. Techniques that permit to synthesize new views using only photographs of a scene are commonly called Image Based Modelling and Rendering techniques: IBMR.

IBMR techniques have found many applications in several fields. Among them: the augmented reality which consists in increasing the visual perception of the real world by inserting visual objects in a real environment. The goal of using IBMR techniques in augmented reality applications is to improve the modelling of augmented environments, as well on the level of precision and the speed of design, as on the level of realism.

In the context of this work we propose an approach that combines two methods of IBMR techniques for the reconstruction of realistic 3D objects. The first is the reconstruction from stereo images and the second is a technique called 'the visual hull'. These two methods are complementary by nature. The visual hull technique reconstructs a first shape of the object that limits the space including this object and helps the stereo algorithms to avoid calculations out of the object volume. The stereovision method refines the resulting model of the object by detecting points and areas of concave surfaces of the object.

Key words: 3D reconstruction, Image based modelling and rendering, stereovision, stereo matching, visual hull, augmented reality.

CHAPITRE I

INTRODUCTION

1. Introduction

Les images, quelle que soit leur provenance, représentent un support considérable d'informations que ce soit comme support de réflexion (aide à la décision, conception, ... etc) ou comme support de communication (production, représentation, etc). L'émergence relativement récente des nouvelles technologies informatiques de traitement et d'acquisition de l'information a multiplié de manière importante les possibilités d'exploitation de ces images de manière presque exclusivement numérique, et le domaine de l'usage d'images de qualité photographique est particulièrement concerné. L'évolution des capacités de traitement, aussi bien matérielles que logicielles, a également permis de rendre abordable de nouveaux outils de manipulation permettant d'exploiter directement ces images au format numérique avec une qualité équivalente, sinon meilleure, que celle obtenue par les techniques plus traditionnelles pour certaines tâches bien définies.

Dans ce chapitre on essaye de définir très brièvement le principe de la synthèse d'images classique à base des modèles géométriques 3D, comme nous citons quelques limitations et difficultés de ce type de synthèse. Puis nous présentons le principe général de la synthèse d'images à partir d'images qui est une alternative à l'approche classique de synthèse d'images. Le chapitre suivant reprend en détails les notions et les bases de la synthèse d'images à partir d'images.

2. Synthèse d'images classique

La synthèse d'images classique a pour but de calculer des vues aussi réalistes que possible d'une scène tridimensionnelle définie par un modèle géométrique 3D, augmentée de certaines informations photométriques : couleurs, textures, matériaux, et nature de leurs interactions avec la lumière.
Les techniques classiques de synthèse d'images s'attachent toutes à produire des vues d'une scène de modèle géométrique 3D connu. Classiquement, pour ces applications, il est nécessaire d'effectuer une première étape de modélisation manuelle de chaque élément de la scène à synthétiser, puis une étape de rendu pour générer les images finales de cette scène.

2.1 Phase de modélisation

Lors de cette modélisation, le concepteur doit décrire à l'aide d'un formalisme approprié les caractéristiques géométriques (forme) et photométriques (couleurs, textures planes) des objets de la scène. Chacun de ces objets est alors habillé par une couleur (dans le cas le plus simple) ou d'une texture qui représente le matériel de cet objet, voire également sa rugosité et son relief. Puis, sont définies les différentes lumières éclairant la scène et les caméras desquelles sont prises les vues. Cette phase est réalisée généralement à la main, assistée par un logiciel spécialisé qui permet au concepteur de décrire rapidement et de placer les objets dans le volume de la scène.

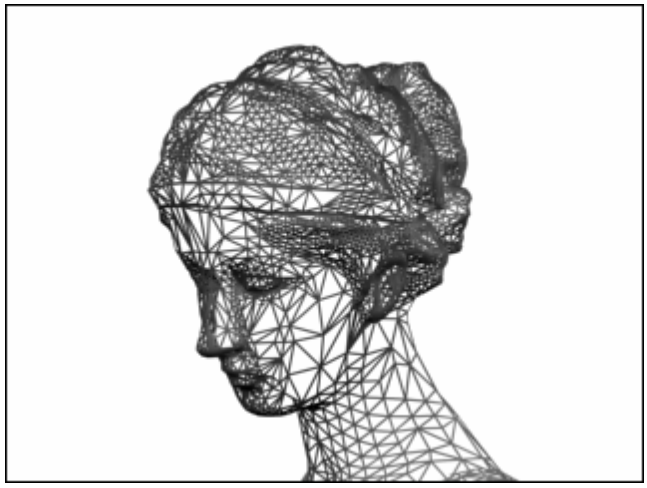


Figure I.1*: Modélisation d'un objet 3D*

Le modèle de l'objet ainsi constitué est appelé maquette numérique, qui est la représentation informatique de cet objet à partir d'informations géométriques. La méthode la plus classique de représentation consiste à raisonner en termes de surfaces. Chaque objet peut être décomposé en « facettes », ou polygones, qui, mis bout à bout, permettent de rendre compte de l'enveloppe extérieure d'un solide. Plus une maquette comporte de polygones, plus l'image qui en résulte est précise (figure I.2). Au moment de l'affichage, l'objet ainsi reproduit se présente sous la forme d'une juxtaposition de facettes, dite « structure en fil de fer ». Il s'agit d'une représentation purement géométrique qui ne prend pas en compte les caractéristiques optiques de l'objet.

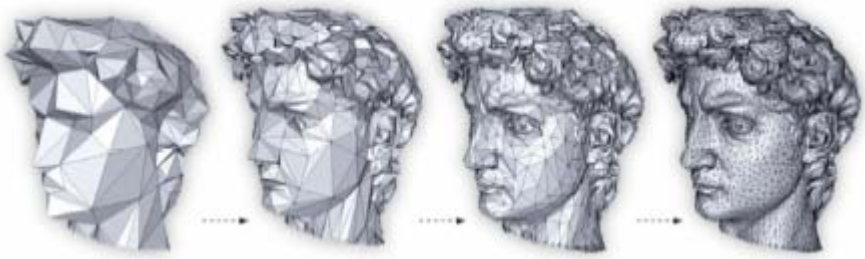


Figure I.2 : *effets du nombre de polygones sur la précision du modèle 3D*

2.2 Phase de rendu

C'est cette étape qui fabrique véritablement l'image de synthèse. Elle vise à transformer la description de la scène en 3D en une image 2D en fonction des différents éléments de la scène et du point de vue duquel elle est regardée. Le système de synthèse d'images génère des images d'un réalisme maximal vis-à-vis du modèle de la scène, et des lois physiques de propagation de la lumière, prenant en compte les réflexions, réfractions, diffusions, et interactions avec les matériaux composant la scène. Ceci est le calcul de rendu, et constitue l'axe essentiel de la recherche actuelle en synthèse d'images.

***Figure I.3**: Rendu de l'objet 3D modélisé*

Il existe diverses techniques de rendu, le « lancer de rayons » ou *ray tracing*, permet d'obtenir un « réalisme rutilant », par une simulation de l'optique géométrique, les rayons se reflètent ou se réfractent selon les matériaux. La « radiosité » est une autre technique qui permet d'obtenir un « réalisme feutré » calculé à partir des propriétés de réflectivité des matériaux. La lumière est analysée comme échange d'énergie entre surfaces, ce qui permet d'obtenir des lumières tamisées et des pénombres.

Bien que ces méthodes donnent des résultats très satisfaisants et de haute qualité (photographique et cinématographique) notamment pour les objets fabriqués par l'homme, elles sont inapplicables en l'état à des scènes réelles, qui sont vastes et complexes comme par exemple un paysage forestier, ou une ville entière. Les modèles décrivant ces scènes devraient en effet contenir un grand nombre de détails pour être réalistes (cas de la forêt), ou tout simplement un trop grand volume de données (cas de la ville), difficilement gérable en pratique. Enfin, nous pouvons évoquer aussi l'aspect du modèle 3D obtenu. Les objets produits ont souvent un aspect synthétique et artificiel, ce malgré des techniques de rendu de plus en plus réalistes, mais aussi de plus en plus complexes. Cela pose plusieurs types de problèmes.

1. Les méthodes de synthèse d'image sont gourmandes en temps de calcul. Si la scène est trop complexe, calculer une seule image peut déjà demander plusieurs minutes de calcul sur des machines spécialisées. Ceci peut être considéré comme un problème secondaire, car la puissance de calcul des machines évolue exponentiellement avec le temps. Cependant, des calculs temps-réel peuvent être dès à présent nécessaires pour, par exemple, la simulation à retour d'effort, ou les applications de la réalité virtuelle et de la réalité augmentée.

2. Une scène complexe demande une modélisation fastidieuse, pouvant se chiffrer en hommes-années. Ainsi, il est hors de question de modéliser chaque bâtiment et chaque rue d'une ville entière à la main, ou de modéliser les branches d'un arbre dépouillé en hiver. En revanche, la somme de données décrivant les détails d'une forêt au printemps est le plus souvent inutile, car un réalisme suffisant peut être obtenu par un modèle d'arbre relativement simple, pourvu d'une texture de feuillage faisant illusion.

3. Un autre constat porte sur l'aspect fastidieux de la tâche de modélisation. Si l'on souhaite, par exemple, reconstruire un immeuble existant, il nous faudra utiliser des

relevés métriques pris sur le terrain ou sur des plans afin de rester le plus fidèle possible à la réalité. L'obtention de ces données et leur saisie peut s'avérer être un travail long et difficile. De plus, si le site a été détruit, cette prise d'informations peut être impossible.

4. En plus une grande partie des environnements de modélisation traditionnels permettent l'utilisation de primitives de plus haut niveau, mais ils n'en restent pas moins difficiles à prendre en main du fait de la grande diversité de fonctions et d'options qu'ils proposent. Ainsi, l'utilisation de ces outils nécessite souvent une phase d'apprentissage plutôt longue de la part de l'utilisateur, et celui-ci ne maîtrise pas toujours toute la puissance du système. La phase de modélisation est, donc la plupart du temps, confiée à un spécialiste en modélisation 3D.

3. Synthèse de nouvelles vues à partir d'images réelles

Pour remédier à ces problèmes, ce qui est proposé est de définir la scène non pas par un modèle géométrique tridimensionnel, mais par des vues (bidimensionnelles) réelles de cette scène dans le but de synthétiser de nouvelles vues uniquement à partir des vues de départ en simulant le déplacement de la caméra qui a pris les vues réelles [5]. La scène doit donc exister réellement, et être physiquement accessible à la prise de vues.

Les techniques permettant de synthétiser de nouvelles vues à partir des vues réelles d'une scène sont appelés communément méthodes de modélisation et rendu basés image (*Image Based Modeling and Rendering : IBMR*), Mais certains auteurs font la distinction entre les méthodes de modélisation et les méthodes de rendu.

En général, le terme de "rendu basé images" (Image-Based Rendering : IBR) est utilisé pour décrire les techniques qui, à partir d'un ensemble d'images d'un environnement, produisent un nouvel ensemble d'images représentant des vues virtuelles de cet environnement (figure I.4).
Le terme de "modélisation basée images" (Image-Based Modeling IBM) est, quant à lui, utilisé lorsque l'on génère un modèle 3D à partir de photographies d'une scène (figure I.5). En effet, disposé d'un modèle 3D de la scène, permet de synthétiser de nouvelles vues depuis n'importe quels points de vue en simulant le déplacement de la caméra qui a pris les photographies. De plus, un modèle tridimensionnel permet des traitements géométriques plus généraux, comme des déformations, des simplifications, des augmentations ou des incrustations d'autres objets. Une représentation 3D peut en effet être manipulée par toutes sortes d'outils standard, comme des logiciels de synthèse d'images, ou des éditeurs de modèles.

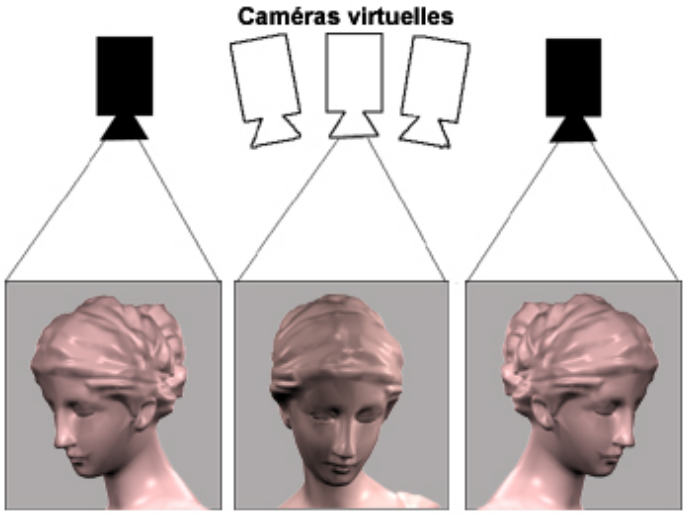


***Figure I.4** : Principe du rendu à base d'images*

Le problème principal posé par la synthèse d'images à partir d'images est qu'une image est une représentation bidimensionnelle d'un monde tridimensionnel, la troisième dimension est perdue au cours du processus de formation de l'image par projection. Tout l'art de la synthèse consiste alors à utiliser les informations présentes dans les photographies d'une scène pour inférer les informations pertinentes à la reconstruction de cette même scène.

Les informations concernant le monde réel que l'on veut extraire sont essentiellement de deux natures :

- Photométriques : la quantité de lumière (intensité lumineuse, la couleur, …) en chaque point de l'image.
- Géométriques : formes et positions dans l'espace des éléments composant la scène.

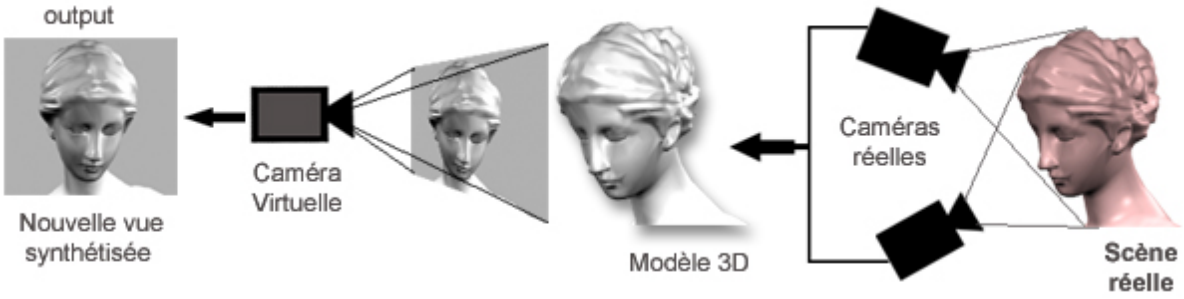


***Figure I.5**: Principe de la reconstruction d'un modèle 3D à partir d'image*

Les intérêts de la synthèse de nouvelles vues à partir de photographies sont multiples et dépendent du domaine concerné. Mais le point commun est tout de même le besoin d'améliorer la modélisation d'environnements en 3D, tant au niveau de la précision et de la rapidité de conception, qu'au niveau du réalisme. En effet, utiliser des images réelles pour créer des images de synthèse offre un double avantage : l'élimination du difficile problème de modélisation géométrique et photométrique complète du monde réel et l'accélération de l'étape de rendu. En fait, les vues disponibles de la scène contiennent des informations

géométriques et des informations de texture et couleurs sous une forme déjà rendue car les objets sont éclairés par une source de lumière réelle.

Les recherches dans la synthèse à partir d'images sont menées depuis longtemps dans le domaine de la photogrammétrie et de la vision par ordinateur et la robotique. Plus récemment, les communautés de l'intelligence artificielle en vision et de l'infographie s'y intéressent en particulier. Bien entendu, les préoccupations sont relatives aux besoins et aux outils que dispose chaque domaine de recherche. Pourtant, même si les problèmes rencontrés par ces communautés sont en général différents, les méthodes utilisées pour leur résolution sont souvent partagées.

- La *Photogrammétrie* emploie des techniques de photographie stéréoscopique [7]. Les formes sont déterminées à l'aide d'images qui se chevauchent prises avec des appareils calibrés. Ce sont des techniques apparues avant l'informatique. Les premiers travaux sur le sujet proviennent de l'inventeur français A. LAUSSEDAT en 1851 et furent formalisées par le mathématicien E. KRUPPA en 1913. L'idée de base est la suivante : la projection 2D d'un point 3D est sur le rayon formé par le centre de projection et ce point 3D. Si l'on a deux images de différents points de vues du même point 3D, celui-ci se trouve à l'intersection des deux rayons, c'est la *triangulation.*
Mais le problème majeur réside dans les nombreuses données qu'il est nécessaire de traiter (notamment les correspondances de points) et la nécessité d'utiliser des appareils calibrés avec précision.

- La *Vision par Ordinateur* et la *Robotique* font de la reconstruction 3D un de leurs principaux champs de recherche [33]. Il est en effet crucial en robotique de pouvoir reconstruire en temps réel un environnement 3D pour la navigation d'une machine. Les méthodes utilisées sont *l'inférence des structures à partir du mouvement* (*structure from motion*), la stéréovision, les cartes de profondeurs, ... Il n'y a pas besoin, en général dans ces applications, d'une reconstruction détaillée du fait que le modèle n'est pas destiné à être visualisé.
Ces deux domaines ont donc souvent recours à des algorithmes de mise en correspondance de points entre différentes images, et c'est cette partie du traitement qui pose le plus de problèmes [6]. Ces algorithmes de stéréo-correspondance sont très sensibles au bruit, aux déformations perspectives et aux éclairages, ce qui les rend moins fiables. Pour la photogrammétrie classique, ils peuvent servir d'aide à l'opérateur humain, mais en robotique, où tout doit être automatisé, de nombreuses recherches sont encore menées pour améliorer ces méthodes de correspondance, ainsi que celles de détection automatique de contours.

- *L'infographie,* les systèmes de modélisation sont souvent sophistiqués, complexes et longs à maîtriser, d'où un intérêt récent pour la reconstruction à partir d'images [1]. De plus, les applications nécessitent de plus en plus de réalisme (environnements immersifs, réalité augmenté, effets spéciaux, ...) et les modèles traditionnels sont souvent trop "propres", ou "simples". L'intérêt est donc pour l'infographie de pouvoir créer plus rapidement des modèles plus réalistes qui correspondent à des objets dont on a des photographies, des plans, des dessins... Dans ces approches, l'on vise en général à faire du système une aide au concepteur qui tient une place importante dans le processus de reconstruction.

4. Applications

Les applications de cette technique sont toutes les applications usuelles de la synthèse d'images : CAD, Réalité Virtuelle, Réalité Augmentée : entraînement, simulation, commerce, loisirs. On peut aussi l'appliquer à la compression de données, et des recherches sont en cours dans ce sens.

4.1 Entraînement

Pour toutes sortes d'interventions en milieu hostile : nucléaire, militaire, spatial, il est nécessaire de préparer et d'entraîner les hommes à évoluer dans leur futur milieu d'action. Ceci recouvre les simulateurs de vol et d'entraînement militaires, dont le but est de former les pilotes à leur mission ; en les plongeant dans une situation simulée le plus fidèlement possible, ils apprennent à repérer la topographie du site et l'emplacement des objectifs. Ceci recouvre également des missions d'intervention dans le nucléaire civil : des simulateurs ont pour but d'entraîner les agents à intervenir rapidement en cas d'incident, en leur permettant de se familiariser à la géographie du lieu à l'aide d'une représentation graphique tridimensionnelle réaliste.
Dans le cas des simulateurs de vol, le modèle de la scène est construit à partir de vues aériennes ou satellitaires. Les images constituent des couples stéréo, ou sont de simples vues monoculaires. Le travail de modélisation consiste à reconstruire la topologie du lieu observé (ainsi que les bâtiments), à partir de ces photos ; ce travail est encore très souvent manuel ou semi-manuel, et demande un temps considérable. Aussi, beaucoup de travaux ont pour but d'automatiser cette tâche.
Il faut remarquer que les images disponibles ne sont pas toujours bien adaptées à la tâche de reconstruction 3D. Ainsi, il est difficile d'obtenir un couple stéréo d'images satellitaires, car il faut que le satellite survole deux fois le même site, sous un angle différent, aux mêmes heures et sous les mêmes conditions météo, et à des instants assez proches (pour que les deux vues de la scène soient aussi semblables que possible). Aussi, certains satellites sont équipés de systèmes de prise de vues stéréoscopiques. Notons que dans le cas d'images monoculaires, une reconstruction tridimensionnelle automatique est bien sûr impossible, et il faut disposer d'autres sources d'information : cartographiques, ou suppositions sur la hauteur des bâtiments, par exemple. Enfin, la résolution, excellente pour des images aériennes (mais il faut que le site soit survolable à basse ou moyenne altitude), est moins bonne dans le cas d'images satellitaires : 1 pixel représente souvent 1 mètre au sol (20 cm pour les meilleurs satellites militaires d'observation).
Le cadre est un peu différent dans le cas des simulateurs civils (p. ex. pour le nucléaire), car les images sont disponibles en aussi grande quantité et précision que nécessaire. Ainsi, lors de la construction d'une centrale nucléaire, les architectes prennent plusieurs milliers de photographies du site, et les archivent sur un vidéodisque. Ces images permettent ensuite au personnel de la centrale de se familiariser avec tous les recoins du site, rendus désormais inaccessibles par les radiations, afin de pouvoir intervenir sans délai en cas d'incident, en se repérant facilement pour agir vite et subir une exposition minimale. De tels systèmes de formation sont opérationnels, mais ils ne présentent qu'une collection d'images fixes, depuis des points de vue figés. On pourrait imaginer les étendre pour synthétiser les images intermédiaires, ce qui permettrait de visualiser les déplacements de façon continue (images animées), rendant ainsi l'immersion de l'opérateur plus naturelle.

4.2 Simulation

La construction de modèles de villes à grande échelle [10] à partir de vues aériennes, par exemple, a d'autres applications que les simulateurs de vol. Par exemple, la très forte croissance des services de téléphonie mobile oblige les opérateurs à couvrir rapidement des zones urbaines denses. Il est donc nécessaire de savoir où placer les relais hertziens pour une couverture optimale, et ceci est réalisé par simulation, à partir d'un modèle 3D de la ville. Comme il est fastidieux de produire de tels modèles manuellement, des systèmes automatiques de génération à partir d'images aériennes ont ici tout leur sens.
La simulation d'environnements complexes est aussi nécessaire en robotique. Pour la simulation de la marche d'un robot martien, on pourrait envisager de modéliser le sol martien à partir de photographies, ce qui fournirait un environnement synthétique de complexité réaliste, tout en évitant une saisie manuelle fastidieuse. L'intérêt de telles simulations a été démontré dans des contextes industriels, parfois de façon spectaculaire. Ainsi au CERN, à Genève, la simulation en images de synthèse du site de construction du futur accélérateur de particules lHC dans le tunnel existant a permis aux ingénieurs de mieux appréhender sa configuration, et d'éviter la construction d'un puits supplémentaire, économisant plusieurs millions de dollars. Enfin, les problèmes de reconnaissance d'objets, de saisie et d'asservissement, sont aussi liés à l'extraction automatique de modèles à partir de vues.

4.3 Commerce, conservation, loisirs

Avec Internet s'est développée la possibilité d'effectuer des transactions financières à distance, et de pratiquer le commerce électronique. Aussi, les chaînes de distribution commencent à rendre disponibles leurs catalogues sur le Web, agrémentés de photos et de visualisations plus ou moins animées ou interactives de leurs produits. Certaines galeries marchandes virtuelles ont été créées (ibm, The Internet Mall, la Redoute...) où l'utilisateur pourra à terme se déplacer dans les magasins, afin de visualiser puis de choisir les produits. Certaines implémentations partielles sont actuellement basées sur QuickTime VR (hypermarchés Leclerc) [18].
La conservation est aussi une application de la capture automatique de modèles à partir d'images. Nous regroupons sous ce terme les activités cartographiques et de préservation du patrimoine. Le projet Michelangelo fait l'objet d'une telle étude, détaillée dans [26], qui consiste à pouvoir reconstruire et visualiser en 3D les sculptures de Michelangelo.
On peut encore citer le domaine des loisirs, et de toutes les applications exigeant un grand nombre d'images : visite de musées virtuels, visite de maisons ou d'équipements ménagers pour la vente, de lieux de vacances dans une agence de voyage, jeux vidéo. La visite virtuelle du Louvre, actuellement diffusée sur cédérom, utilise des films QuickTime VR créés à partir de photos.

4.4 Compression de données

À partir de quelques images d'une scène, nous pouvons calculer d'autres vues de cette scène. Cela a une application immédiate en compression de données : la compression à très fort taux de séquences vidéo. Des recherches sont actuellement poursuivies dans différentiels laboratoires afin d'utiliser la synthèse d'images à partir de vues pour la compression vidéo [39]. Le principe est de calculer une représentation géométrique grossière des objets filmés, ainsi que les transformations qu'ils subissent : changements de position, ou de point de vue. Pour transmettre le film, on transmet alors en une seule fois cette représentation géométrique. Chaque image successive est ensuite entièrement décrite par les paramètres décrivant le point

de vue courant, ce qui constitue un volume de données très faible et indépendant de la taille des images. Connaissant la représentation géométrique de la scène et la position d'observation, le récepteur peut alors reproduire chacune des images.
Notons que de façon similaire, la norme de compression vidéo à très bas débit mpeg4, explore les techniques de compression par modèles : les objets composant une image y seront décrits de façon individuelle par leur forme 2D, leur texture, peut-être même leur modèle 3D [39].

5. Organisation du mémoire

La synthèse de nouvelles vues à partir d'images réelles est une tache ardue et une large bibliographie traite se sujet. Dans le reste de ce mémoire nous essayons de d'introduire ce domaine et ses applications dans la réalité augmentée, la description du contenu de chaque chapitre est comme suit:

- Le chapitre II traite les principes de base de la synthèse de nouvelles vues à partir d'images réelles. La reconstruction de la géométrie de la scène est la clé pour la plupart des méthodes de reconstruction 3D à partir d'images.
  Parmi les nombreuses méthodes mises au point pour calculer ces informations 3D, on peut distinguer deux grandes classes : Les méthodes actives qui utilisent des capteurs ayant une action directe sur l'environnement étudié et les méthodes passives qui n'ont aucune action sur l'environnement. Premièrement nous faisons un survole des méthodes actives de reconstruction 3D sans entrer dans les détailles, puis nous nous plaçons dans le domaine de la stéréovision (méthode passive) qui nous intéresse et nous développons en détail les étapes classiques qui entrent en jeu dans la reconstruction 3D par stéréovision. Les trois étapes majeures sont :
  - Définition géométrique du processus de prise de vues (Formation de l'image, modèle de la caméra, calibrage,...) ;
  - La mise en correspondance ;
  - La reconstruction 3D.

  Et finalement nous passons en revue les méthodes d'IBMR les plus connues dans la littérature illustrées par des exemples des résultats les plus significatifs. Ces méthodes sont diverses et dépendent souvent des moyens mis en oeuvre.

- Le chapitre III traite le problème de l'application des méthodes d'IBMR dans le domaine de la réalité augmentée. Cette dernière consiste à augmenter la perception visuelle du monde réel par l'insertion réaliste d'objets visuels synthétiques. L'approche traditionnelle consiste à générer des images d'un objet ou d'une scène 3D à l'aide d'un algorithme de rendu appliqué à un modèle tridimensionnel construit à l'aide d'un logiciel de modélisation. Les limitations des techniques de synthèse d'images classique ont encouragé le développement de nouvelles techniques basées sur les images pour accroître le réalisme et simplifie la modélisation.
  Dans ce chapitre nous faisons une brève introduction des notions de la réalité augmentée, puis nous exposons l'aidée de l'utilisation des méthodes d'IBMR dans les applications de la réalité augmentée. Pour cela nous présentons des applications où les méthodes d'IBMR sont appliquées avec un succès prouvé.

- Dans le chapitre IV, et du fait que notre objectif dans ce travail est la synthèse de nouvelles vues pour les applications en réalité augmentée, où cette dernière exige que les

objets utilisés pour l'augmentation soient en 3D. Une nouvelle approche de reconstruction d'objets 3D est présentée. L'approche proposée pour la reconstruction d'objets 3D combine deux techniques d'IBMR. La première est la reconstruction par stéréovision et la deuxième est celle de l'enveloppe visuelle.

Et enfin, une conclusion et des perspectives possibles pour notre travail seront exposées.

CHAPITRE II

SYNTHÈSE DE NOUVELLES VUES À PARTIR D'IMAGES RÉELLES (IBMR): *Principes et Méthodes*

1. Introduction

La synthèse de nouvelles vues à partir d'images du monde réel est une tache de la vision par ordinateur. Cette dernière est une discipline qui tente de simuler la vision humaine en établissant des modèles qui possèdent des propriétés proches de la perception visuelle humaine. Le processus de reconstruction des formes qui, chez l'homme est réalisé de manière plus ou moins consciente, s'avère un problème assez délicat lorsqu'il s'agit de faire ce travail par une machine.

Le problème principal posé par la synthèse d'images à partir d'images est qu'une image est une représentation bidimensionnelle d'un monde tridimensionnel, la troisième dimension est perdue au cours du processus de formation de l'image par projection. Tout l'art de la synthèse consiste à utiliser les informations présentes dans les photographies d'une scène pour inférer les informations (en particulier la dimension perdue) pertinentes à la reconstruction de cette même scène.

Les informations concernant le monde réel que l'on veut extraire sont essentiellement de deux natures :

- Photométriques : la quantité de lumière (intensité lumineuse, la couleur, …) peuvent être pertinente à connaître ;
- Spatiales ou géométriques: les coordonnées spatiales des éléments composant la scène sont fondamentales pour la majorité des applications.

Parmi les nombreuses méthodes mises au point pour calculer les informations 3D, on peut distinguer deux grandes classes :

• Les méthodes actives qui utilisent des capteurs ayant une action directe sur l'environnement étudié [40] : systèmes acoustiques (relatifs au son), tactiles (qui réagit au toucher), télémètres, …etc. De cette caractéristique fondamentale découle leur nom : méthodes actives. Ils acquièrent généralement l'information spatiale.

• Les méthodes passives qui n'ont aucune action sur l'environnement. Les capteurs utilisés sont généralement des caméras comme mode d'acquisition des informations élémentaires (images), desquelles sont extraites les informations photométriques (qui mesure l'intensité d'une source lumineuse) et spatiales. Ces méthodes font recours aux techniques de la stéréovision pour inférer la dimension perdue.

L'objectif commun des méthodes actives et des méthodes passives est la reconstruction de la géométrie 3D de la scène observée c'est-à-dire le calcul des coordonnées 3D de chaque point de l'ensemble des points constituant la scène dans l'espace.

1.1 Les méthodes actives

L'acquisition d'objets en vue d'une reconstruction 3D a débuté grâce à des équipements de "copie mécanique", ou plus communément connus sous le nom de Machines à Mesurer Tridimensionnelles (M.M.T.). Ces outils s'appuient sur la technologie alors bien maîtrisée de la détermination de points dans l'espace. En effet, les coordonnées d'un point quelconque se déterminent aisément par la géométrie d'un palpeur et les codeurs des axes de mouvement de la machine d'acquisition.

Pour le simple copiage, un palpeur à balayage continu ou analogique sera utilisé, ou pour des applications plus complexes, un palpeur point par point, sur bras fixe ou pantin mobile (bras articulé). Le coût de ces appareils est très élevé surtout lorsqu'ils sont constitués de pièces mécaniques de haute précision.

L'acquisition des données 3D nécessite énormément de temps puisqu'il faut amener le palpeur sur chaque point voulu et mesurer. Les palpeurs modernes sont connectés à des ordinateurs et les mesures sont faites automatiquement. Même si cette méthode est la seule permettant de numériser tout type d'objet, indépendamment de leur forme ou de leur matière, cette technologie est de moins en moins utilisée au profit des systèmes de télémétrie laser pour différentes raisons : Sans contact, plus rapide, et permettant une plus grande flexibilité, le laser présente aussi un spot de dimension réduite permettant l'appréhension de détails très fins.

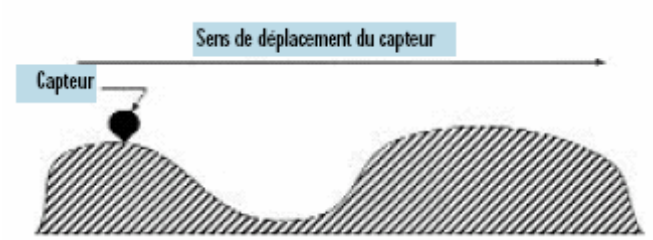


Figure II.1 : *L'acquisition 3D mécanique par palpeur*

Les méthodes de télémétrie laser utilisent en général le temps de vol de la lumière pour faire l'aller retour du laser au point visé pour déterminer la distance du point au laser. Chaque relevé 3D d'un point nécessite le positionnement du laser, l'émission de lumière et la mesure du temps de vol. La méthode, même si elle se révèle être la plus précise nécessite un temps énorme pour scanner des objets complexes et n'est donc utilisée que dans les cas où une précision extrême est exigée et lorsque l'objet est immobile. Ce type de scanner est utilisé dans un très grand nombre de domaines, dans l'industrie, par les géomètres, par les architectes, …etc. le schéma de principe du télémètre laser est donné figure II.2.

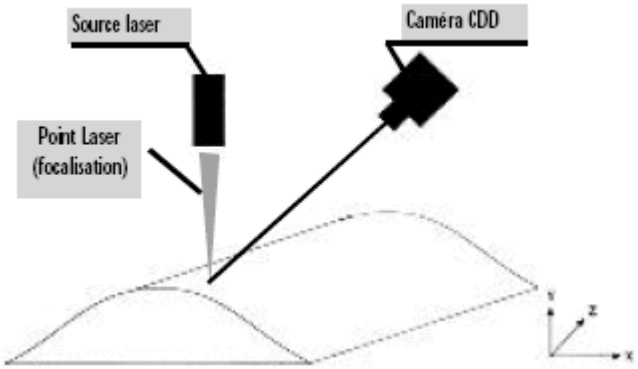


Figure II.2 *: Télémètre Laser*

Certains systèmes actifs intègrent un système de vision et font alors généralement intervenir une source lumineuse supplémentaire. On utilise une lumière donnant un sur-éclairement de la scène et permettant alors une détection plus facile des points ainsi marqués ; les lasers sont les plus couramment employés. La figure II.3 illustre ce principe.

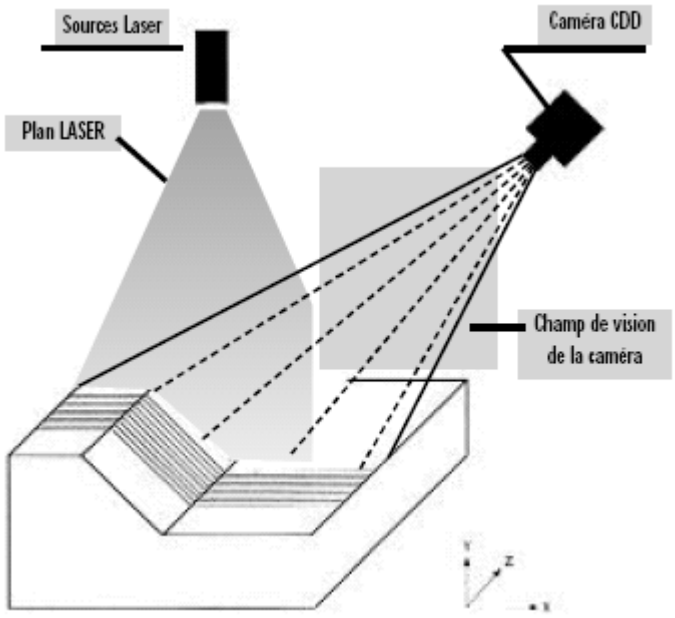


Figure II.3 *: Projection du plan Lazer lumineux sur une scène*

Il est alors possible de calculer les coordonnées tridimensionnelles des points de la scène ainsi éclairés par simple triangulation à condition de connaître avec précision la position de la source de lumière et le modèle de la caméra utilisé. L'ensemble de la scène peut être étudié si l'on peut faire bouger soit la source de lumière, soit la scène. Le mouvement de la source de lumière est généralement obtenu en faisant pivoter un miroir réfléchissant les rayons lumineux. Déplacer la scène revient par exemple à placer les objets à étudier sur un plateau tournant (voir figure II.4).

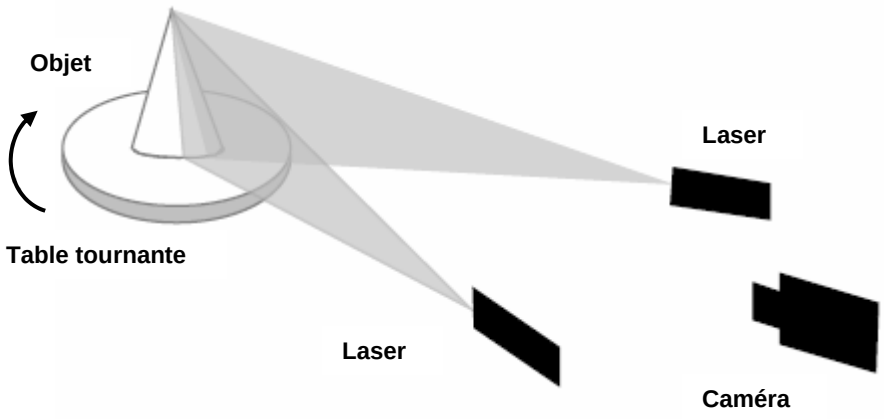


Figure II.4 *: Projection du plan lumineux sur une scène L'objet est placé sur une table tournante contrôlée par ordinateur et éclairée par deux raies lumineuses.*

Un autre type de méthodes utilise un éclairage contrôlé ou une lumière structurées illuminant la scène d'une manière particulière. Le principe de fonctionnement est assez simple puisqu'il consiste à projeter une lumière de forme connue sur l'objet et de mesurer les déformations de cette forme à la surface de l'objet grâce à une caméra ou un appareil photo. On observera alors non plus la scène dans son ensemble mais le résultat de l'éclairage de la scène par cette

source particulière. Ceci permet d'en extraire des informations tridimensionnelles sans tenir compte des caractéristiques photométriques propres de la scène puisque celles-ci sont imposées par la source lumineuse.

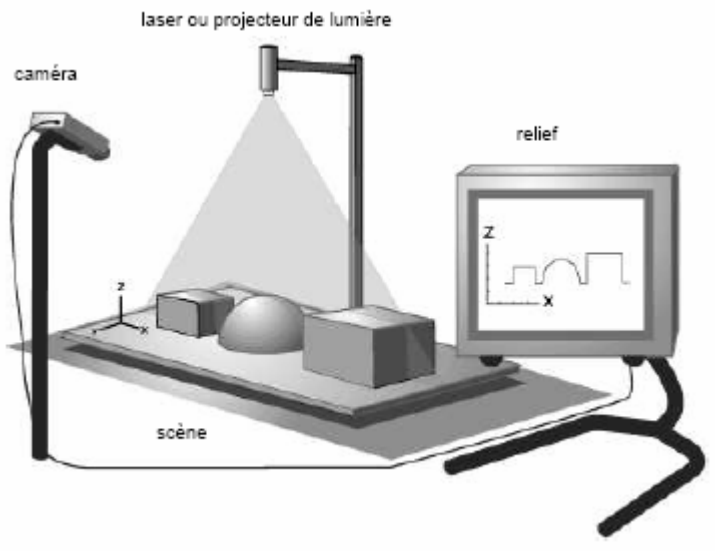


***Figure II.5** : Schématisation de la reconstruction du relief à l'aide de lumière structurée*.

Les méthodes de reconstruction basées sur la lumière structurée englobent de nombreuses techniques dont, en particulier, la triangulation laser. Néanmoins, il nous semble intéressant de nous attarder sur l'une de ces méthodes, à savoir, la projection sur la scène d'un motif structurant. Comme cela se fait par la triangulation laser plane, on peut envisager plutôt que de balayer la scène avec un faisceau plan, de projeter sur la scène entière un motif structurant (souvent une grille, ou des franges d'interférence). Le processus de triangulation est donc ici un peu plus complexe puisqu'il faut apparier la grille apparente (celle projetée sur la scène) avec la grille théorique (le motif structurant proprement dit).

En fait, on peut considérer les méthodes, qui utilisent la projection d'une grille sur la scène comme un homologue de la stéréovision dans lequel on aurait remplacé une des deux caméras par un projecteur. Un processus d'appariement est alors nécessaire mais sa complexité est uniquement dépendante de la complexité du motif structurant projeté sur la scène et, de ce fait, n'est pas dépendant des artefacts qu'engendrent de faux appariements entre les deux images.

Figure II.6: *Lumière structurée illuminant un visage*

Les méthodes actives fournissent des mesures précises et peuvent être s'avérer très efficaces, mais la description du monde obtenu est généralement très locale. Les systèmes utilisant des

sources lumineuses supplémentaires imposent en outre de travailler sous des conditions d'éclairage très contraignantes et ne conviennent pas aux surfaces non réfléchissantes.
Néanmoins, les méthodes actives sont très utilisées dans de nombreuses applications, et particulièrement dans le domaine de la robotique.
Il est cependant des cas où certains systèmes actifs sont inutilisables du fait soit de leur interférence avec le milieu (scène d'intérieure : le laser peut s'avérer dangereux d'utilisation, les ultrasons nocifs aux animaux, …) soit de la profondeur de la scène observée (images satellites).

1.2 Les méthodes passives

Les méthodes passives [7] n'ont aucune influence sur l'environnement, leur rôle est simplement de recevoir les signaux lumineux émis ou réfléchis par la scène. Les capteurs utilisés sont principalement les Caméras. L'information directement acquise n'est pas tridimensionnelle, mais seulement bidimensionnelle. Une ou plusieurs images sont traitées afin d'extraire l'information tridimensionnelle souhaitée.
Parmi ces méthodes, la stéréovision est sans doute la plus utilisée. Elle nécessite l'emploi de plusieurs vues prises depuis des points de vue différents du même objet. L'appariement de points provenant des différentes caméras permet ensuite d'appliquer le principe de la triangulation, et donc de positionner dans l'espace des points mis en correspondance dans une paire ou une séquence d'images. En ce sens la stéréovision représente le système minimal pour réaliser cette triangulation.
On est donc confronté directement à la complexité apparente des images. Il s'agit alors de déterminer quelles sont les informations pertinentes à extraire de ces images pour en reconstruire la géométrie de la scène.

Dans ce travail on s'intéresse à la reconstruction 3D par les méthodes passives et en particulier par la stéréovision. Pour cette raison nous détaillons dans ce qui suit les notions et la démarche suivie pour la reconstruction 3D à partir d'images par stéréovision.

1.2.1 La stéréovision

Le monde réel perçu par une machine de vision artificielle est un univers en deux dimensions. En effet, une caméra effectue une opération de projection perspective qui transforme le monde tridimensionnel de la scène en une représentation bidimensionnelle, causant la perte de l'information de profondeur. Si cette transformation de projection est connue, il est possible, étant donné un point physique de la scène, de connaître précisément la position de sa projection dans l'image. En revanche, étant donné un point de l'image, il existe une infinité de points de la scène portés par une droite qui vérifie la transformation inverse.

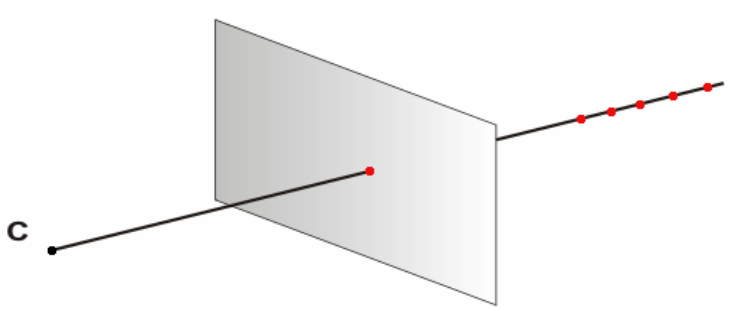


***Figure II.7** : Impossibilité de reconstruire la scène à partir d'une seule image*

Comment, dans ces conditions, retrouver la dimension perdue ? Sans elle, nous perdons nous-mêmes. Un complément d'information est nécessaire pour déterminer la coordonnée tridimensionnelle manquante. La solution se trouve dans la combinaison de plusieurs images prises de points de vue différents. En particulier, la vision binoculaire ou stéréovision est l'interprétation de deux vues distinctes de la scène afin de résoudre l'ambiguïté de la profondeur. Connaissant ainsi le modèle de projection de chaque caméra et la relation spatiale entre elles, il s'agit de calculer les coordonnées 3D d'un point à partir de ses deux projections dans les deux images.

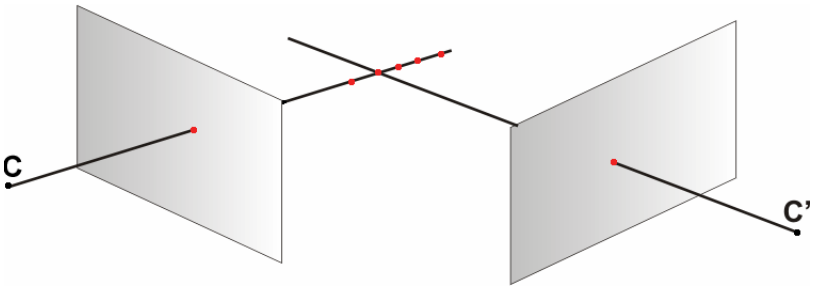


***Figure II.8** : Elimination de l'ambiguïté par deux images prise depuis des points de vue différents*

1.2.2 Système de stéréovision

La stéréovision vise à réaliser une tâche similaire à la vision humaine, à l'aide d'un ou plusieurs périphériques de capture d'image (par exemple un appareil photo ou un caméscope numérique), reliés à un ordinateur.
On utilise deux caméras, vidéo ou CCD, disposées comme les yeux humains, qui vont donner deux images d'une *scène* (figure II.8). Ces images sont formées de pixels (ou points image) et constituent une *paire stéréoscopique*. On considère que l'aspect géométrique de l'espace qui nous entoure est une représentation tridimensionnelle d'entités physiques. Une image prise par une seule caméra est alors considérée comme une représentation bidimensionnelle de cet espace. Il y a donc perte d'information durant le processus de formation d'une image. En particulier, la troisième dimension. La récupération de cette dernière est le but de la stéréovision.

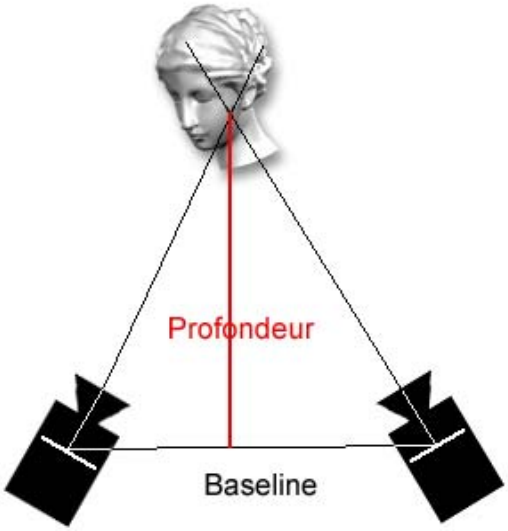


***Figure II.9** : Système de stéréovision*

Du fait de l'écartement des caméras qui les ont prises, les images de cette paire ne sont pas sans relation. Ainsi un observateur de la scène verra ses images décalées d'une image de la paire sur l'autre, d'un certain nombre de pixels. Ce décalage est appelé *disparité*. Si on est capable d'attribuer une disparité à chaque pixel d'une image de la paire stéréoscopique, on attribuera par extension une disparité à tous les points d'un objet sur l'autre image. On sera alors en mesure de replacer tous les points de cet objet dans l'espace, donc de reconstruire l'objet dans la scène à partir des deux images est l'information de profondeur inférée.

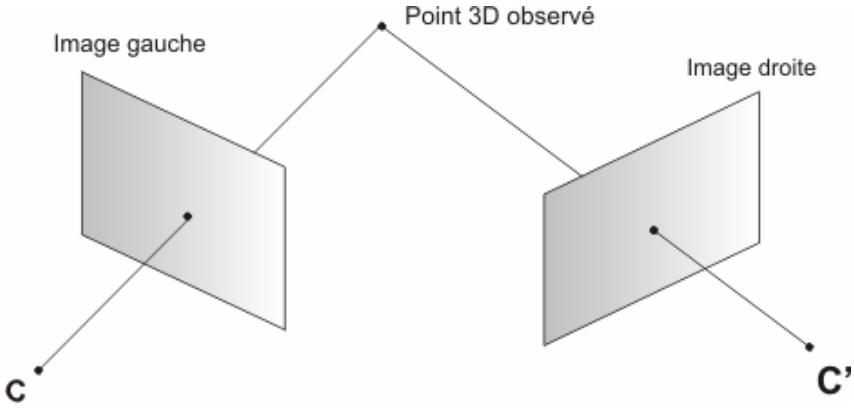


***Figure II.10** : Schéma simplifier du système de stéréovision*

Présenté ainsi, le principe de la stéréovision parait relativement simple, mais en on doit affronter plusieurs problèmes dès lors que l'on souhaite appliquer ce procédé de manière automatique sur une paire d'images :

- Dans un premier temps il faut savoir comment sont formées les images prises par les caméras du système stéréoscopique, par d'autre terme il faut connaître le modèle géométrique (de projection) des caméras qui ont pris les vues ainsi que leurs paramètres intrinsèques et extrinsèques. On doit donc modéliser la transformation perspective subie par les points du monde réel dans le monde projectif des images, ainsi que le déplacement relatif entre les deux points de vue. Ce problème constitue l'objet de la section 2 (*géométrie de la vision)*.
- Armé de ces informations, on peut alors aborder le problème de la mise en correspondance entre les points des deux images. Pour un point donné dans une image, on doit trouver son homologue dans l'autre image de la paire stéréoscopique pour que le calcul de la dimension perdue devienne possible. Ce problème constitue l'objet de la section 3 : (*La mise en correspondance)*.
- Pour tous les couples de points mis en correspondance calculer la position 3D du point de la scène réelle par triangulation. C'est le problème de la reconstruction. Ce problème constitue l'objet de la section 4 *(Reconstruction)*

2. La Géométrie de la vision

La caméra est l'outil essentiel autour duquel se développe la vision. Elle est l'interface entre deux espaces. Le premier est un ensemble de données inconnues évoluant dans un espace inconnu, il s'agit de la scène. Le deuxième est l'espace image, il s'agit des données perçues qui évoluent dans l'espace particulier qu'est l'image. La caméra se définit comme l'outil géométrique (projectif) qui traduit les relations existantes entre ces deux espaces.

Si le but de la reconstruction à partir d'images est de reconstruire la scène tri-dimensionnelle projetée sur une ou plusieurs images, il est important de comprendre et de bien modéliser les transformations géométriques qui permettent la formation des images, pour enfin pouvoir extraire des informations métriques sur la scène à partir de ces images.

Dans cette partie, on va présenter quelques notions de la géométrie de la vision. Notre intention n'est pas ici de faire un état de l'art de ce domaine mais plutôt de présenter les bases minimales nécessaires à la compréhension des méthodes qui seront présentées plus loin dans ce travail.

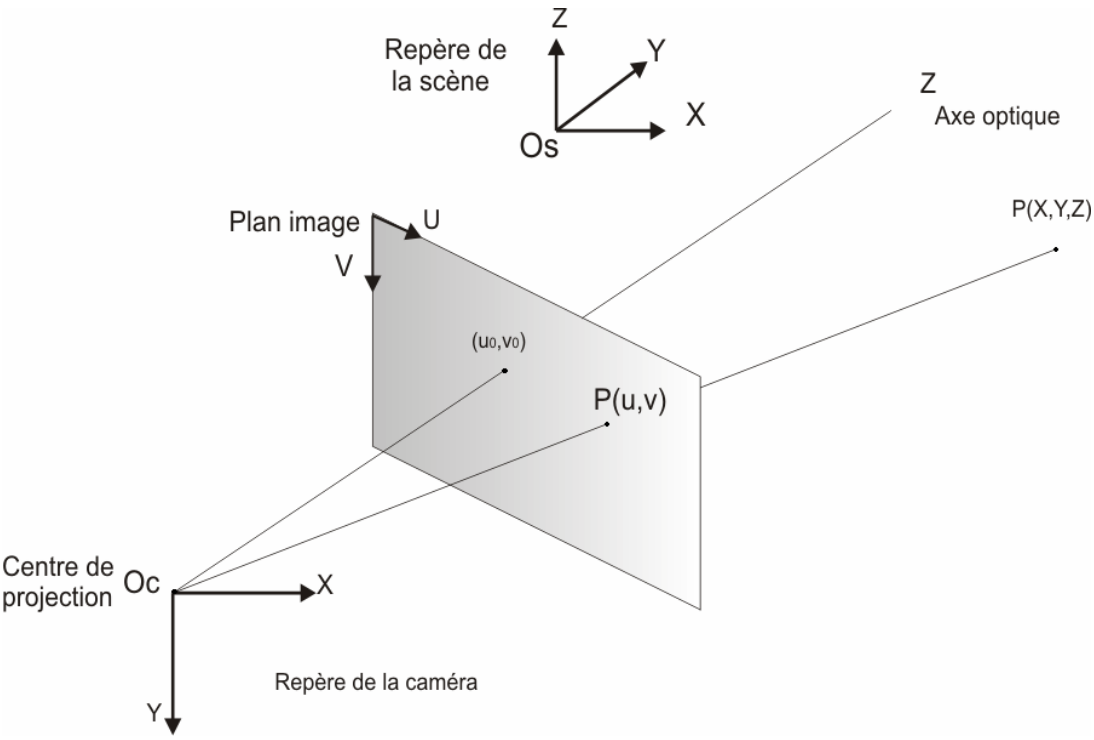


***Figure II.11** : système de coordonnées*

2.1 Systèmes de coordonnées

Considérons les positions respectives d'une caméra et d'un objet, comme dans la *figure II.10*. Nous pouvons définir plusieurs systèmes de coordonnées pour représenter cette situation. Tout d'abord, il y a un référentiel "monde " fixe, lié à la scène, ***Rs= (Os,X,Y, Z)***, par rapport auquel la caméra pourrait être en mouvement.
Ensuite, nous pouvons associer un référentiel à la caméra : $\boldsymbol{R_c = (O_c , x,y, z)}$.
Finalement, nous pouvons définir un référentiel lié à l'image : ***Ri= (Oi, u, v)*** .
Le plus souvent, nous chercherons à obtenir nos résultats dans le système $\boldsymbol{R_c}$, mais il peut parfois être utile de les exprimer dans le système ***Rs,*** comme par exemple quand on veut construire une carte de l'espace de travail.

2.2 Coordonnées homogènes et transformations géométrique

La géométrie projective simplifie l'étude des processus de vision [4]. Les coordonnées homogènes permettent en effet de décrire les différentes transformations impliquées sous

forme matricielle. De plus, la géométrie Euclidienne, qui nous intéresse, constitue un cas particulier de la géométrie projective.
L'objet de base "point" est représenté par un vecteur de ***n*** coordonnées (***n*** est la dimension de l'espace de travail): **(x,y)** en 2D et **(x,y,z)** en 3D.
Un objet graphique pourra être modélisé par un ensemble de facettes elles-mêmes modélisées par des sommets (points).
Une transformation géométrique (translation, rotation, changement d'échelles….) appliquée aux points prend la forme d'une matrice carrée **M** de dimension ***n*** (***n*** est la dimension de l'espace de travail). Cette matrice appliquée à un point **P** donnera le nouveau point **P'** par le produit matriciel :

$$\mathbf{P' = M * P}$$

Problème: Les transformations translation d'une part, et rotations et mises à l'échelle d'autre part ne sont pas uniformément représentables au moyen de cette définition. On peut représenter toutes ces transformations par des matrices, mais il n'existe pas d'opération mathématique de composition de matrices permettant de déterminer la matrice modélisant la réalisation successive de deux transformations. Ce modèle n'est pas capable de représenter deux ou plus transformations géométriques successives. Pour remédier à ce problème un nouveau système de coordonnées est utilisé : *les coordonnées homogènes.*

2.2.1 Coordonnées homogènes

On utilise des vecteurs à **n+1** coordonnées dans un espace de représentation de dimension **n** : **n** coordonnées + une coordonnée supplémentaire.
Les transformations géométriques sont réalisées à partir de matrices carrées de dimension **n+1**.
Les translations sont modélisées à partir des valeurs contenues sur la **n+1^{ième}** colonne de la matrice de transformation tandis que les rotations et mises à l'échelle utilisent plus classiquement la sous-matrice 3x3 supérieure gauche.
La **n+1ème** coordonnée d'un vecteur est initialisée à **1** pour représenter un vecteur position, ou à **0** pour initialiser un vecteur déplacement invariant par translation (**0** désigne aussi les points à l'infini.).
Les points de l'espace projectif **3D *P³*** sont représentés par un vecteur homogène de dimension **4 : P' = (X1; X2; X3; X4)^{t}**. Ce point **(X; Y; Z)** est défini par la relation :
X = X1/X4
Y = X2/X4
Z = X3/X4

2.2.2 Les transformations élémentaires que subissent les objets dans l'espace 3D

Les transformations sont représentées sous forme matricielle.

Translation :

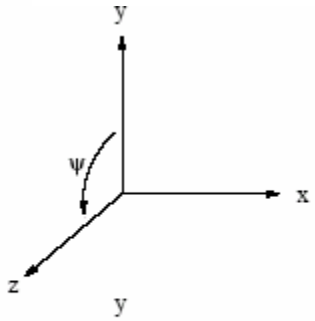
Translation dans l'espace 3D selon le vecteur (Tx,Ty,Tz).

$$T = \begin{bmatrix} 1 & 0 & 0 & T_x \\ 0 & 1 & 0 & T_y \\ 0 & 0 & 1 & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

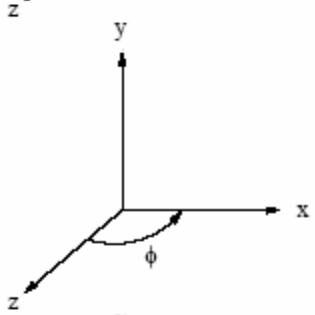
Rotations : Représentation par les angles d'Euler :

$$R = R_z \cdot R_y \cdot R_x,$$

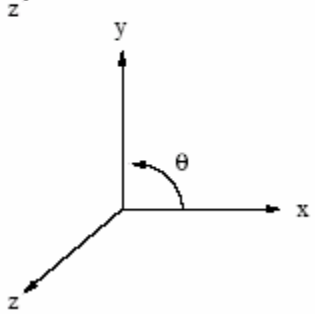
$$R_x = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\psi & -\sin\psi & 0 \\ 0 & \sin\psi & \cos\psi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



$$R_y = \begin{bmatrix} \cos\phi & 0 & \sin\phi & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\phi & 0 & cos\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



$$R_z = \begin{bmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



Changements d'échelles **:**

$$T = \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

2.2.3 Les Homographies

Les homographies sont des applications linéaires de $\mathbf{P^3}$, elles sont décrites par l'ensemble des matrices **H** 4x4 non singulière et possèdent 15 degrés de liberté (nombre de translations, rotations, changement d'échelle...). Le 16 terme traduit l'aspect homogène de ces matrices, les homographies sont définies à un facteur d'échelle prés. Nous décrivons ici les différents types de transformations observables. Dans le cas le plus général, la matrice H est décomposée en 4 parties.

$$H = \begin{bmatrix} A & t \\ v^t & s \end{bmatrix}$$

Cette matrice est une homographie lorsque A est une matrice 3x3 non singulière, ***v*** et ***t*** deux vecteurs et ***s*** un scalaire. **H** étant une matrice homogène, on peut toujours se ramener au cas ou ***s*** = 1 en divisant chacun des termes de la matrice par ***s***. Dans le cas particulier où **v** = 0; **H** représente une transformation affine de l'espace, les affinités présentent la propriété de conserver le parallélisme des plans. L'application possède alors 12 degrés de liberté.
Supposons que l'on puisse de plus décomposé A tel que:

$$H = \begin{bmatrix} \alpha R & t \\ 0^t & 1 \end{bmatrix}$$

Où ***R*** est une matrice de rotation. Alors **H** est une similarité. Physiquement, un volume qui subit une telle opération est translaté suivant ***t***, subit la rotation ***R*** et le facteur d'échelle ***α***. Cette transformation possède 7 degrés de liberté.
Enfin si ***α* = 1** l'homographie correspond à une transformation euclidienne.

Les modèles qui permettent de décrire la scène ou l'image sont basés sur la géométrie euclidienne, affine ou projective. On distingue autant de modèles que de types de relations entre l'espace scène et l'espace image. Chacun fournit une interprétation différente.

2.3 Espaces Euclidien, métrique, affine et projectif

L'espace Euclidien est celui qui nous entoure : les transformations sont rigides et les grandeurs physiques sont exprimées dans une unité de mesure. En déplaçant un objet non déformable dans son environnement comme par exemple une chaise, il est possible de lui faire subir uniquement des composées de translations dans trois directions, ainsi que des composées de rotations autour de ces trois directions tel que représenté dans la *figure II.11*. Tout déplacement réalisable "physiquement dans la réalité" sur cet objet est une transformation qui correspond au groupe euclidien et l'environnement dans lequel nous évoluons peut donc ainsi être décrit comme "euclidien".

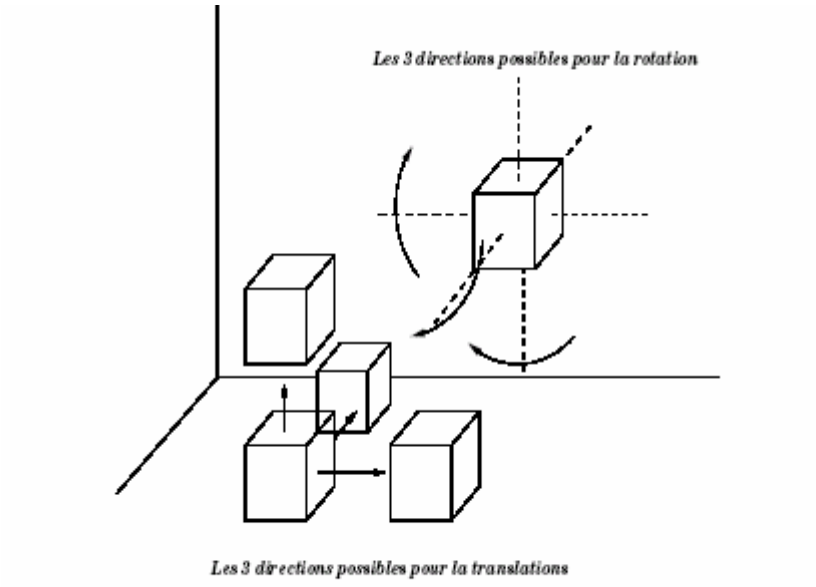


***Figure II.12** : les transformations possibles dans un espace euclidien*

L'espace métrique est l'espace Euclidien où l'échelle globale a été changée. En particulier, il n'est possible d'obtenir une reconstruction Euclidienne que lorsque la mesure d'une longueur réelle est introduite. En l'absence d'une telle mesure, seule une reconstruction métrique peut être obtenue à partir de caméras.

Ses transformations possèdent un degré de liberté supplémentaire puisque les distances absolues ne sont plus nécessairement conservées. Cependant la *figure II.12* montre que les distances relatives sont conservées : un cube reste un cube dans une transformation du groupe métrique, mais sa taille peut varier. Cela reviendrait à avoir la possibilité de pouvoir modifier la taille de notre chaise, en plus de pouvoir lui appliquer des translations et des rotations. Ses transformations possèdent un degré de liberté supplémentaire puisque les distances absolues ne sont plus nécessairement conservées. Cependant la *figure II.12* montre que les distances relatives sont conservées : un cube reste un cube dans une transformation du groupe métrique, mais sa taille peut varier. Cela reviendrait à avoir la possibilité de pouvoir modifier la taille de notre chaise, en plus de pouvoir lui appliquer des translations et des rotations.

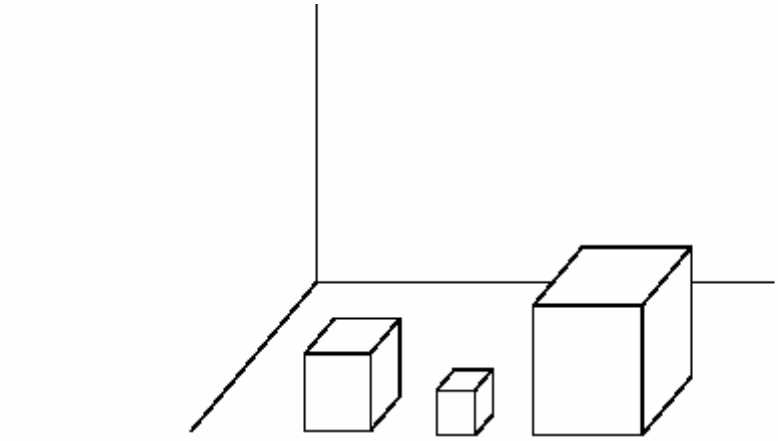


Figure II.13 : *Un nouveau degré de liberté pour les transformations possibles dans un espace métrique*

L'espace affine est un espace dans lequel les contraintes sont moins importantes que dans l'espace métrique, les rapports de longueurs et les angles n'ont pas de sens, mais le parallélisme en a un, et dans lequel les transformations obtiennent des degrés de liberté supplémentaires. Il est plus général que l'espace métrique car les transformations affines n'y imposent plus la conservation des angles. Ainsi, le cube des *figures II.11* et *II.12* pourra devenir un des formes de la *figure II.13* au travers d'une transformation de l'espace affine.

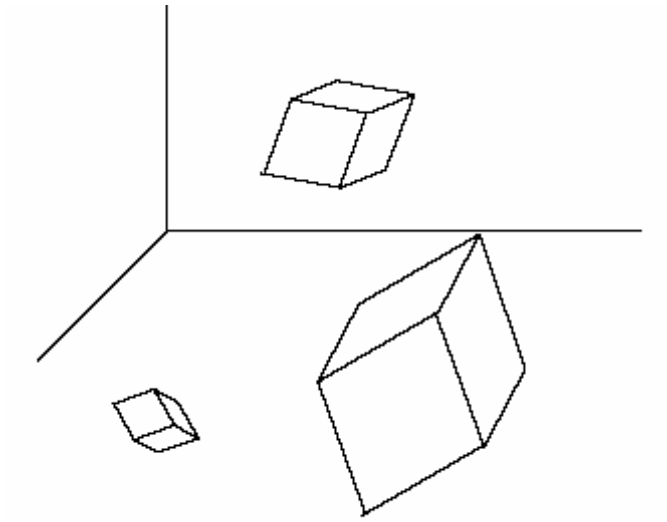


***Figure II.14** : Exemple de transformation affine d'un cube*

En espace projectif, toutes ces caractéristiques sont perdues. Cet espace comporte encore plus de degrés de liberté que l'espace affine. En effet, les transformations des différents espaces évoqués ci-dessus conservaient le parallélisme alors que ce n'est pas toujours le cas pour les transformations de la couche projective. Lorsqu'on trace deux droites parallèles dans le plan affine, métrique ou euclidien, les deux droites ont la particularité de ne jamais se couper. Mais dans l'espace projectif, ces deux droites se coupent en un point que l'on appellera à l'infini. Puisque des droites parallèles sont définies comme convergentes vers un point à l'infini, les deux droites restent parallèles mais leur représentation projective ne le met pas en évidence. Le phénomène est bien connu des dessinateurs puisque la réalisation d'une vue en perspective impose de définir un point de fuite (*vanishing point*). Ce point de fuite n'est rien d'autre qu'un point à l'infini se retrouvant dans le plan affine à la suite de la projection (en perspective) du modèle en 3D sur un plan en 2D. Ce là est le principe du modèle géométrique utilisé en vision pour décrire une projection centrale de la scène 3D sur un plan 2D qui est l'image.

***Figure II.15** : un exemple d'image en représentation projective*

2.4 Modèle de la camera et formation de l'image

2.4.1 Modèle de la caméra

Le modèle géométrique le plus couramment employé en vision pour décrire la formation de l'image est le modèle du « trou d'épingle » ou « sténopé » (*pinhole),* illustré à la *figure II.15(a)*. Les rayons lumineux provenant des objets de la scène passent à travers un trou d'épingle dans une boîte (modélisant la caméra) et se projettent sur une surface plane. On notera que la projection obtenue est inversée par rapport à la scène et que ses dimensions sont proportionnelles à la distance entre le trou d'épingle et la surface de projection.
Cette surface sera appelée le plan image de la projection. Le point O_c coïncidant avec la position du trou d'épingle est le centre optique, comme il est montré à la *figure II.15(b)*. La distance f entre O_c et le plan image est dans ce modèle la distance focale de la caméra. L'axe O_cz est l'axe optique, perpendiculaire au plan image et pointant vers la scène dont il mesure la profondeur Z. Enfin, on définit les axes O_cx et O_cy de manière à ce que :
(1) ces axes soient parallèles aux axes de l'image (réelle),
(2) le référentiel $R_c = (O_c, x, y, z)$ soit direct.

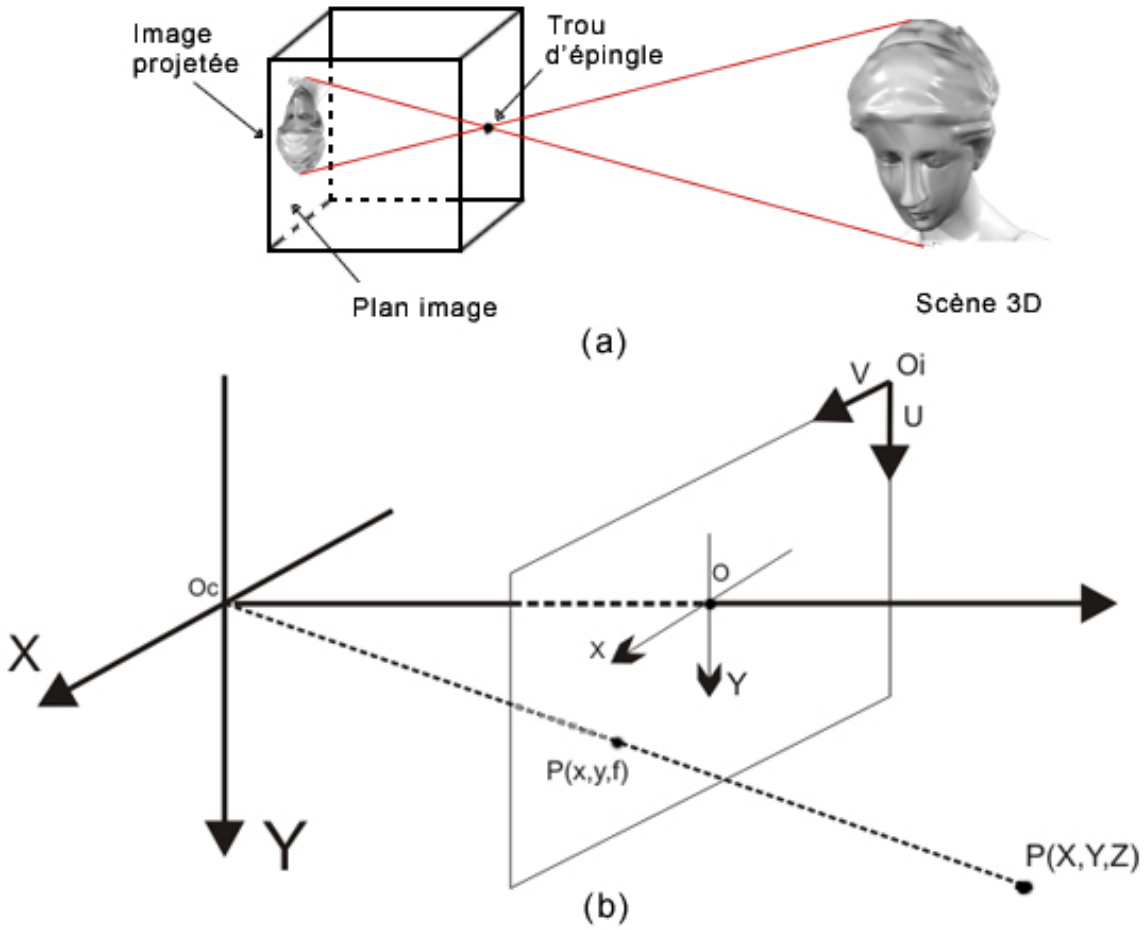


Figure II.16 *: (a) Le trou d'épingle et (b) son modèle géométrique*

On peut remarquer trois choix qui ont été faits dans ce modèle géométrique :

- *Le plan image a été placé en avant du centre optique* ***Oc****.*
La raison pour cela est tout simplement que ce modèle ne change pas la géométrie de la projection et permet d'avoir simultanément un axes ***z*** pointant vers la scène et une distance focale ***f*** positive.

- *L'origine du référentiel de la caméra a été placée au centre optique* O_c*.*
On aurait pu choisir le centre du plan image, ***o***, mais cela aurait compliqué les expressions de la projection.

- *L'axe* ***y*** *pointe vers le bas*
Cela nous permet d'avoir un trièdre direct et en même temps que les axes dans le plan image aient les orientations classiques gauche-droite et haut-bas.

2.4.2 Limitations du modèle de trou d'épingle

Le modèle du trou d'épingle n'est qu'une approximation imparfaite du processus de formation de l'image. En particulier,

- Le modèle ne prend pas en compte les problèmes de mise au point. En pratique, avec une optique réelle, il ne sera pas possible d'obtenir une image nette que pour un intervalle de profondeur donné. Un point situé hors de cet intervalle sera alors observé dans l'image sous la forme d'un petit disque flou (dont le rayon dépend de la distance focale et de la distance *z*).

- Le modèle suppose que la projection géométrique est parfaite. En pratique, on observe toutes sortes de distorsions dans l'image (en particulier, les projections de lignes droites apparaissent courbes dans l'image).

2.5 Formation de l'image

En se limitant à l'aspect géométrique, une image obtenue avec une caméra de type sténopé est le résultat d'une transformation géométrique. Cette dernière fait passer d'une représentation tridimensionnelle de la scène à une représentation bidimensionnelle (image).
Soit **P** un point de l'espace de la scène et **p** sa projection sur l'image par le modèle du trou d'épingle selon les repères :
Repère de la Scène :

Point Scène : $\mathbf{P_s = (X,Y,Z,1)^T}$ (point scène en repère scène)

Repère de la Caméra :

Point Caméra : $\mathbf{P_c} = (x,y,z,1)^T$ (point scène en repère caméra)

Point Image : $\boldsymbol{p_r} = (x_r,y_r,1)^T$ (point image en repère caméra)

Repère de l'Image :

Point Image : $\boldsymbol{p_i} = \mathbf{(u,v,1)^T}$ (point image en repère image)

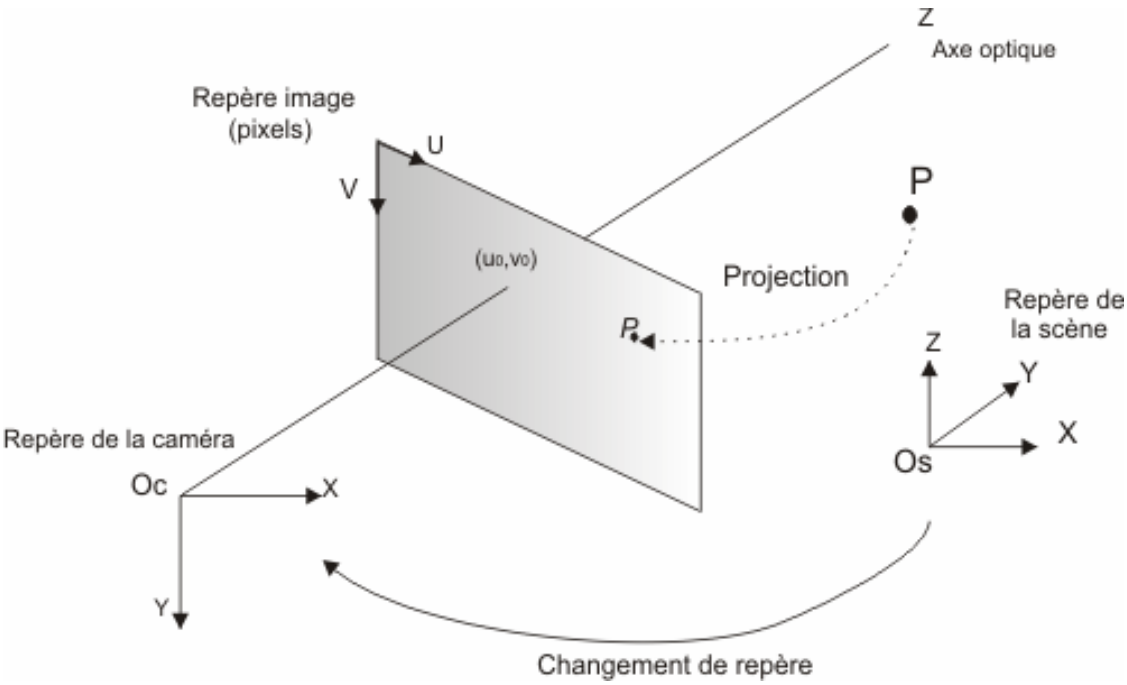


Figure II.17 : *Formation de l'image*

Pour passer des coordonnées définies dans le repère de la scène aux coordonnées images exprimées en pixels, trois phases sont nécessaires [33]:

(1) Un déplacement tridimensionnel: les points tridimensionnels exprimés dans un repère de la scène subissent un changement de repère pour passer au repère de la caméra. Ce changement de repère comporte donc 6 paramètres : 3 pour la rotation et 3 pour la translation. Ces paramètres ne sont autres que la position et l'orientation de la caméra, ils sont appelés *paramètres extrinsèques*.

$$\mathbf{P_c} = (\mathbf{RT})\ \mathbf{P_s}$$

(2) Une projection 3D-2D : après le changement de repère de la phase précédente, les points tridimensionnels exprimés dans le repère de la caméra sont projetés sur le plan image. Les nouvelles coordonnées ainsi obtenues sont appelées coordonnées normalisées (ils sont toujours exprimées dans le repère caméra). Cette projection est une projection perspective dans le modèle de trou d'épingle.

$$\boldsymbol{p_r} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{F} & 0 \end{pmatrix} \mathbf{P_c}$$

(3) Changement de coordonnées : pour passer aux coordonnées pixels, les coordonnées normalisées subissent une transformation du plan. Cette dernière, comporte 5 paramètres appelés *paramètres intrinsèques* de la caméra. Ces paramètres sont :

- La distance focale ***f*** qui est la distance entre le centre optique et le plan image, ***f*** est donnée en millimètres.
- $(\mathbf{u_0},\mathbf{v_0})$ sont les coordonnées en pixels du centre image, c'est à dire les coordonnées du point d'intersection de l'axe optique avec le plan image.
- ***ku*** *et* ***kv*** taille d'un pixel image en pixels/mm (pixels non carrés).

La transformation du repère de la caméra vers le repère image s'écrit donc:

$$\boldsymbol{p_i} = \begin{pmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \boldsymbol{p_r}$$

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

Les trois transformations citées ci-dessus que les coordonnées tridimensionnelles des points d'une scène subissent pour arriver aux coordonnées pixels peuvent être écrites :

$$\begin{pmatrix} wu \\ wv \\ w \end{pmatrix} = \begin{pmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 0 \end{pmatrix} \cdot \begin{pmatrix} & R & & T \\ & & & \\ & & & \\ 0 & & 1 & \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

soit encore :

$$\begin{pmatrix} wu \\ wv \\ w \end{pmatrix} = \begin{pmatrix} k_u f & 0 & u_0 \\ 0 & k_v f & v_0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} & R & & T \\ & & & \\ & & & \\ 0 & & 1 & \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} wu \\ wv \\ w \end{pmatrix} = {}^{i}_{r}C \cdot {}^{c}_{s}T \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

Posons : $$\boldsymbol{M = {}^{i}_{r}C\ {}^{c}_{s}T}$$

Finalement : $$\boxed{\boldsymbol{p_i = M\ \mathbf{P_s}}}$$

${}^{i}_{r}\boldsymbol{C}$ est la matrice des paramètres intrinsèques.
${}^{c}_{s}\boldsymbol{T}$ est la matrice des paramètres extrinsèques.
$\boldsymbol{M}$ est la matrice de projection perspective d'une image appelée aussi matrice de transformation rigide.

2.6 Géométrie épipolaire et relation entre deux images

Considérons deux images **A** et **B** provenant de deux caméras perspectives observant la même scène. Soient $\mathbf{M^A}$ et $\mathbf{M^B}$ les matrices de projections correspondantes à ces deux images. Un point **P** de la scène se projette en $\mathbf{X^A = M^A\ P}$ et $\mathbf{X_B = M_B\ P}$.
Nous allons nous intéresser ici aux liens qui existent entre $\mathbf{X^A}$ et $\mathbf{X^B}$. Ces liens sont caractérisés par la *géométrie épipolaire.*

2.6.1 La contrainte épipolaire

La contrainte épipolaire caractérise le fait que le correspondant $\mathbf{X^B}$ d'un point $\mathbf{X^A}$ (i.e., $\mathbf{X^A}$ et $\mathbf{X^B}$ sont les projections images du même point ***X***) se situe sur une droite $\boldsymbol{l_{XA}}$ dans l'image **B**. En effet, le point $\mathbf{X^B}$ appartient nécessairement au plan défini par $\mathbf{X^A}$,$\boldsymbol{O^A}$ et $\boldsymbol{O^B}$. La droite $\boldsymbol{l_{XA}}$ est appelée *droite épipolaire* du point $\mathbf{X^A}$ dans l'image **B**.

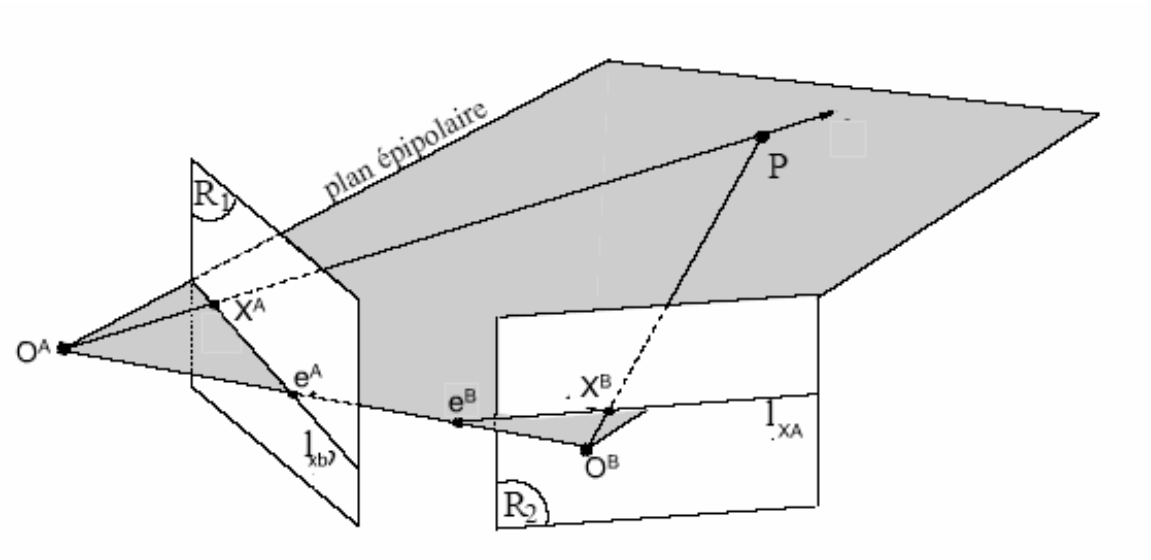


Figure II.18 : *Géométrie épipolaire*

La contrainte épipolaire est symétrique.
Les droites épipolaires dans une image s'intersectent toutes en un point appelé *épipole*. Ce point correspond à la projection du centre de projection de l'autre image considérée :

$$\boldsymbol{e^A = M^A\ O^B}$$
$$\boldsymbol{e^B = M^B\ O^A}$$

Les épipoles jouent un rôle fondamental dans la vision stéréoscopique.

2.6.2 La matrice fondamentale

Nous allons exprimer ici, sous une forme algébrique, la contrainte épipolaire. La relation épipolaire entre les deux images **A** et **B** est défini par une matrice notée **F** appelée *matrice fondamentale*. Etant donné un couple de correspondance $\mathbf{X^A \leftrightarrow X^B}$ où $\mathbf{X^A}$ et $\mathbf{X^B}$ sont les coordonnées homogènes des pixels d'un même point sur les images **A** et **B**, la matrice fondamentale vérifie la formule :

$$\mathbf{X^A\ F\ X^B = 0}$$

où **F** est une matrice 3×3. En géométrie projective à 2D ($\boldsymbol{P^2}$), une droite et un point sont de même nature. On peut donc considérer le vecteur $\mathbf{FX^A}$ comme une droite et le fait d'avoir $\mathbf{X^A\ F\ X^B = 0}$ signifie que le point $\mathbf{X^B}$ se trouve sur la droite $\mathbf{FX^A}$. En pratique, si l'on connaît le pixel $\mathbf{X^A}$ (ou $\mathbf{X^B}$) on obtient les droites épipolaires $\mathbf{L^B}$ (et respectivement $\mathbf{L^A}$) de la façon suivante :

- $\mathbf{L^B = F\ X^A}$
- $\mathbf{L^A = F\ ^t X^B}$

- Calcul de F

La matrice **F** est de la forme :

$$F = \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix}$$

Pour calculer la matrice **F**, on part de l'équation $\mathbf{X^A\ F\ X^B = 0}$. Elle doit être vérifiée pour tous les couples ($\mathbf{X_i^A}$, $\mathbf{X_i^B}$), on obtient donc l'équation suivante :

$$x_i^A x_i^B f_{11} + x_i^A y_i^B f_{12} + x_i^A w_i^B f_{13} + y_i^A x_i^B f_{21} + y_i^A y_i^B f_{22} + y_i^A w_i^B f_{23} + w_i^A x_i^B f_{31} + w_i^A y_i^B f_{32} + w_i^A w_i^B f_{33} = 0$$

Il faut donc pour ***n*** correspondances résoudre le système :

$$\begin{pmatrix} x_1^A x_1^B & x_1^A y_1^B & x_1^A w_1^B & y_1^A x_1^B & y_1^A y_1^B & y_1^A w_1^B & w_1^A x_1^B & w_1^A y_1^B & w_1^A w_1^B \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^A x_n^B & x_n^A y_n^B & x_n^A w_n^B & y_n^A x_n^B & y_n^A y_n^B & y_n^A w_n^B & w_n^A x_n^B & w_n^A y_n^B & w_n^A w_n^B \end{pmatrix} \begin{pmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Une solution évidente est $\boldsymbol{F = 0_{3\times3}}$ mais elle n'est pas très intéressante. Pour éviter de trouver ce résultat, nous allons utiliser une astuce qui consiste à supposer ***f33*** non nul. Ce n'est pas forcement très stable numériquement mais c'est pratique. Puisque ***F*** est invariant par facteur d'échelle (la magie des coordonnées homogènes), nous posons ***f33 =1***. Le système à résoudre devient alors :

$$\begin{pmatrix} x_1^A x_1^B & x_1^A y_1^B & x_1^A w_1^B & y_1^A x_1^B & y_1^A y_1^B & y_1^A w_1^B & w_1^A x_1^B & w_1^A y_1^B \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^A x_n^B & x_n^A y_n^B & x_n^A w_n^B & y_n^A x_n^B & y_n^A y_n^B & y_n^A w_n^B & w_n^A x_n^B & w_n^A y_n^B \end{pmatrix} \begin{pmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \end{pmatrix} = \begin{pmatrix} -w_1^A w_1^B \\ \vdots \\ -w_n^A w_n^B \end{pmatrix}$$

Pour résoudre ce système, il faut donc un ensemble de 8 correspondances de points. Cette méthode fonctionne bien mais 8 points ne permettent pas toujours une très bonne précision. Pour être plus précis, il suffit de prendre plus de points mais le système à résoudre devient surdéterminé.
La matrice fondamentale caractérise toute la géométrie entre deux images hors paramètres intrinsèques.
Les droites épipolaires peuvent êtres calculées à partir de ***F*** :
Les épipoles sont les projections du centre des caméras sur l'autre caméra. Chaque droite épipolaire passe par son épipole. Ceux-ci se calculent avec les relations suivantes :

$$Fe^{A}=0$$
$$F^{t}e^{B}=0$$

2.6.3 La matrice essentielle

Lorsque les paramètres intrinsèques des caméras sont connus, il est possible d'exprimer la contrainte épipolaire dans le repère de la caméra. C'est à dire, pour un point image $\mathbf{X^A}$ de l'image **A**, on peut exprimer la contrainte épipolaire comme suit :

$$\mathbf{X_c^A} = {}^{A}_{r}\boldsymbol{C}^{-1}\mathbf{X^A}$$

Où ${}^{A}_{r}\boldsymbol{C}$ est la matrice des paramètres intrinsèques de l'image ***A*** et $\mathbf{X_c^A}$ les coordonnées du point $\mathbf{X^A}$ dans le repère caméra. Par simple analogie avec ce qui a été vu précédemment, la contrainte épipolaire devient :

$$(\mathbf{X_c^B})^{t} . {}^{B}_{r}\boldsymbol{C}^{t}.\mathbf{F}. {}^{A}_{r}\boldsymbol{C}. \mathbf{X_c^A}=0$$

Soit :

$$\boxed{(\mathbf{X_c^B})^{t} . \mathbf{E}. \mathbf{X_c^A}=0}$$

Avec :

$$E= {}^{B}_{r}C^{t}.F. {}^{A}_{r}C$$

La matrice ***E*** de l'expression précédente est appelée *matrice essentielle*. Cette matrice dépend uniquement du déplacement entre les deux positions des caméras (*paramètres extrinsèques : R et T).*

La matrice essentielle est de rang deux, de plus les deux valeurs propres non nulles de ***E*** ont même valeurs. La matrice essentielle peut être estimée, théoriquement, à partir de cinq correspondances.

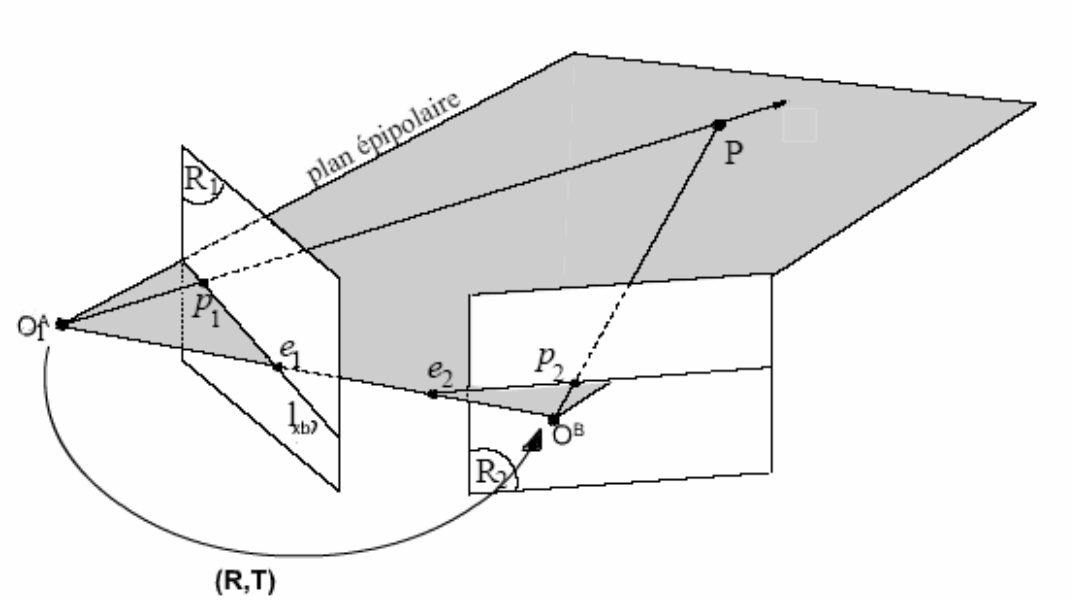


***Figure II.19** : Géométrie épipolaire et matrice essentielle*

2.7 Le Calibrage

Si le but de la reconstruction stéréo est de reconstruire la scène tridimensionnelle projetée sur une paire d'images, il est important de comprendre et de bien modéliser les transformations géométriques qui permettent la formation des images, pour enfin pouvoir extraire des informations métriques sur la scène à partir des images. Le calibrage répond à ce problème, elle permet de déterminer les paramètres intrinsèques et extrinsèques des caméras [32].

2.7.1 Calibrage d'une caméra

Pour reconstruire un modèle 3D, il faut d'abord *calibrer* la caméra. Le Calibrage d'une caméra consiste à estimer M : la matrice de transformation rigide d'un point de l'espace 3D (repère scène) en un point image (repère image). Pour estimer M il faut connaître au moins 6 points non coplanaires de l'espace et ces projections dans une image. Pour calculer la matrice des paramètres intrinsèques et la matrice des paramètres extrinsèques il faux décomposer M ($M = {}^{i}_{r}C\ {}^{c}_{s}T$).
Le calibrage de caméra est un domaine de recherche à part entière, très vaste est le nombre des travaux qui ont été publiés [32,33] et il ne serait pas possible de les citer ici, mais nous pouvons les classer en deux catégories : La première, et la plus ancienne, est la calibration à partir de l'analyse d'images d'une mire dont on connaît précisément la géométrie. La seconde est une calibration automatique (self-calibration) sans connaissance a priori des éléments composant l'image, à partir de paires stéréo, de mouvements de caméras, ou en retrouvant dans l'image des éléments du monde dont on connaît les propriétés projectives, comme par exemple des faisceaux de droites parallèles.

2.7.1.1 Calibrage classique

Ces méthodes, qui nécessitent d'avoir la caméra en sa possession ainsi qu'une mire. Le principe de base est de photographier une mire *(figure II.18)* dont on connaît les formes et la

position dans l'espace, et de détecter ces formes sur les images afin de calculer les paramètres intrinsèques. Chaumette et Rives montrent dans [41] que la connaissance des coordonnées 3D de six points non coplanaires et des coordonnées images de leurs correspondants 2D suffisent à résoudre le problème par un système linéaire. Mais, il est évident qu'il est nécessaire d'utiliser plus de correspondances afin d'améliorer la précision. Des méthodes de minimisation non-linéaires, en général plus robustes et précises sont aussi utilisées [41]

***Figure II.20** : Une mire de calibrage photographiée sous plusieurs angles*

Elles utilisent souvent les résultats d'une résolution linéaire comme valeurs de départ. Les données utilisées en général sont des points mais certaines méthodes utilisent aussi des lignes [42] ou des ellipses ; mais il y a besoin d'extraire ces données des images produites par l'appareil. Dans le cas ou l'image est bruitée ou que les distorsions n'ont pas été calculées, les résultats peuvent être assez surprenants.

2.7.2.2 Calibrage automatique ou auto-calibrage

La calibration automatique est surtout issue de recherches en robotique. La calibration des paramètres intrinsèques d'une caméra peut être faite une fois pour toute, mais les paramètres extrinsèques doivent être recalculés à chaque mouvement, afin de connaître la position de l'appareil.
Il n'y a plus besoin dans ces méthodes de photographier une mire. En effet, elles utilisent des correspondances de points entre deux ou plusieurs images au cours d'un déplacement ou un changement d'orientation afin de déterminer les paramètres à l'aide de la géométrie épipolaire. Il est montré par Faugeras et Luong dans [43] que au moins trois vues différentes, gardant quand même des correspondances, sont nécessaires pour résoudre les équations de Kruppa qui mènent aux paramètres.

3. La mise en correspondance

La mise en correspondance, est le premier pas vers la reconstruction. En effet quel que soit le type de reconstruction envisagé il est nécessaire de trouver dans les deux images les primitives homologues c'est-à-dire qui correspondent à la même entité physique du monde réel.

A partir des positions respectives des deux points dans les deux images on peut reconstruire un point dans l'espace par triangulation. Mais le problème principal de la mise en correspondance ne se trouve pas dans le calcul qui à partir de la disparité permet de replacer un point dans l'espace de la scène. Le problème se trouve principalement dans la mise en correspondance elle même. En effet, l'être humain n'éprouve aucune difficulté à apparier des

points, en revanche cette correspondance peut s'avérer relativement difficile à réaliser d'un point de vue purement algorithmique.

3.1 Correspondance stéréo

Une des explications “de bon sens” les plus communes de notre perception de la profondeur dans une scène est que nous utilisons la différence entre ce que nous voyons avec l’oeil gauche et avec l’oeil droit.
A la base de la stéréovision est la notion de correspondance (*figure II.20*).

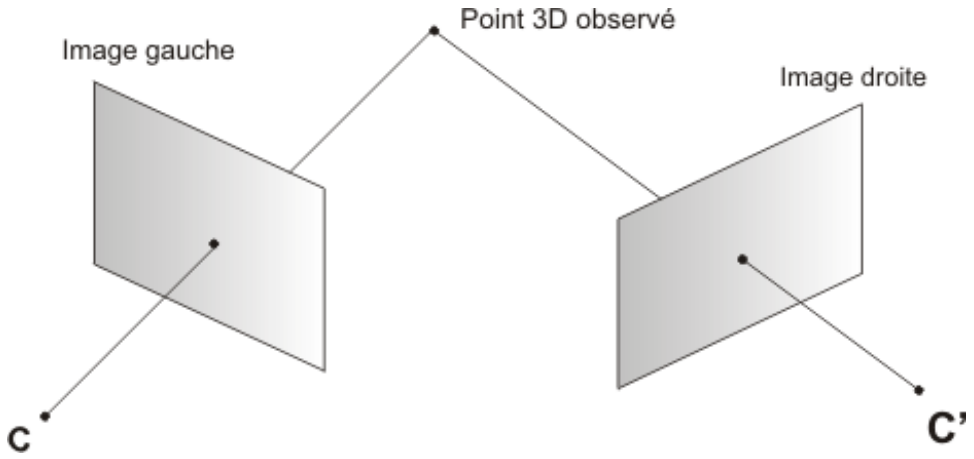


***Figure II.21** : La correspondance stéréo*

En théorie, tout semble très simple : il suffit donc d’établir des correspondances (entre points, lignes, régions,...). En pratique, c’est bien sûr plus compliqué et il n’existe pour le problème de la mise en correspondance aucune théorie dont les implantations donnent des résultats acceptables dans le cas général (voir *figure II.21*).

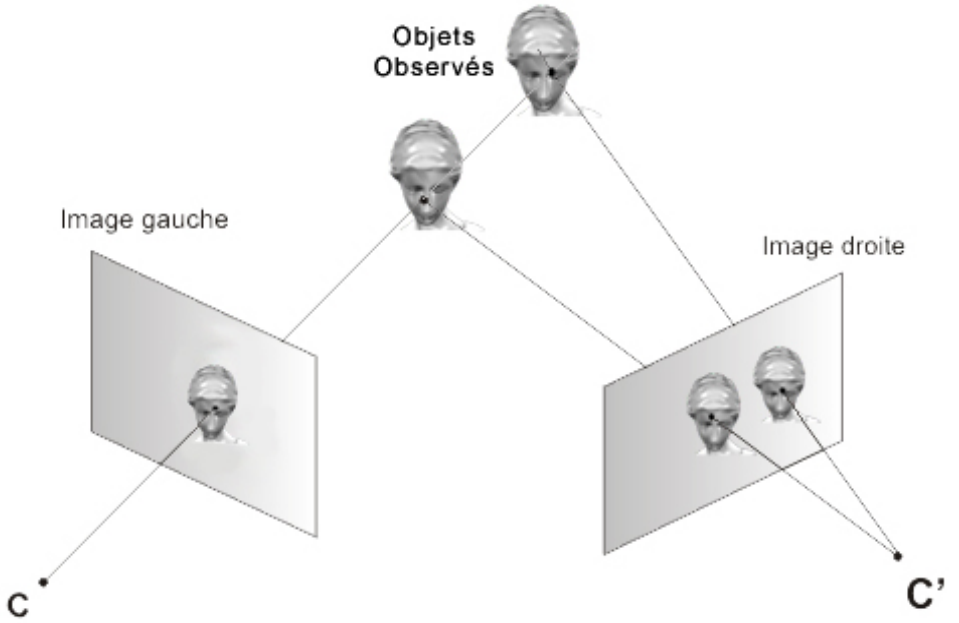


***Figure II.22** : Difficultés d’établir des correspondances en stéréo*

En effet quel que soit le type de reconstruction envisagé il est nécessaire de trouver dans les deux images les primitives homologues c'est-à-dire qui correspondent à la même entité physique du monde réel. Une façon de classer grossièrement les méthodes de mise en correspondance consiste à examiner les primitives qui doivent être associées :

3.2 Primitives stéréoscopiques

Le choix des éléments (ou primitives) image à mettre en correspondance est crucial. La primitive "idéale" est telle que [34]:

(i) ces propriétés intrinsèques permettent une mesure de ressemblance fortement discriminante,
(ii) elle permette la mise en oeuvre efficace de contraintes stéréoscopiques,
(iii) la reconstruction 3D soit possible.

Les techniques de segmentation d'images permettent l'extraction d'une gamme variée de primitives telles que :

- Des points (pixels, points d'intérêt, éléments de contour, points caractéristiques le long d'un contour, jonctions, etc.) ;
- Des segments (segments de droite, arcs de cercle, portions de conique, etc.);
- Des régions.

Les segments de droite ont été utilisés avec quelque succès mais leur utilisation est limitée aux scènes polyédriques [34]. Il est à noter qu'un segment de droite a la même représentation géométrique qu'un point de contour (position dans l'image et direction) et les mêmes contraintes stéréoscopiques s'appliquent également aux segments de droite et aux points de contour.
Bien que très "pauvres" au contenu sémantique, les points s'avèrent finalement être les primitives les mieux adaptées pour la stéréovision. La plupart des contraintes que nous allons étudier s'appliquent aux points.

3.3 Contraintes de la mise en correspondance

On utilise des critères de mise en correspondance pour limiter l'espace de recherche des correspondants dans une paire d'images ou bien renforcer et/ou éliminer certaines correspondances déjà établies. Certaines de ces contraintes concernent les relations qui existent d'une image à l'autre comme la contrainte épipolaire. D'autres concernent plus les éléments à mettre en correspondance comme des points, des contours ou des régions. On peut donc distinguer deux types de critères : Les critères géométriques et les critères morphologiques ou figuraux.

3.3.1 Contraintes géométriques

3.3.1.1 Contrainte épipolaire

Nous avons déjà abordé précédemment le principe de la géométrie épipolaire. La contrainte épipolaire est une contrainte géométrique qui réduit l'ensemble des correspondants

potentiels d'un point à une droite dans l'image. Considérons le point X^A de l'image gauche. Les points de l'espace ayant pour image le point X^A sont situés sur la ligne de vue de direction O^AP. Les correspondants potentiels de X^A dans l'image de droite sont donc nécessairement situés sur la projection de la droite O^AP dans l'image de droite.

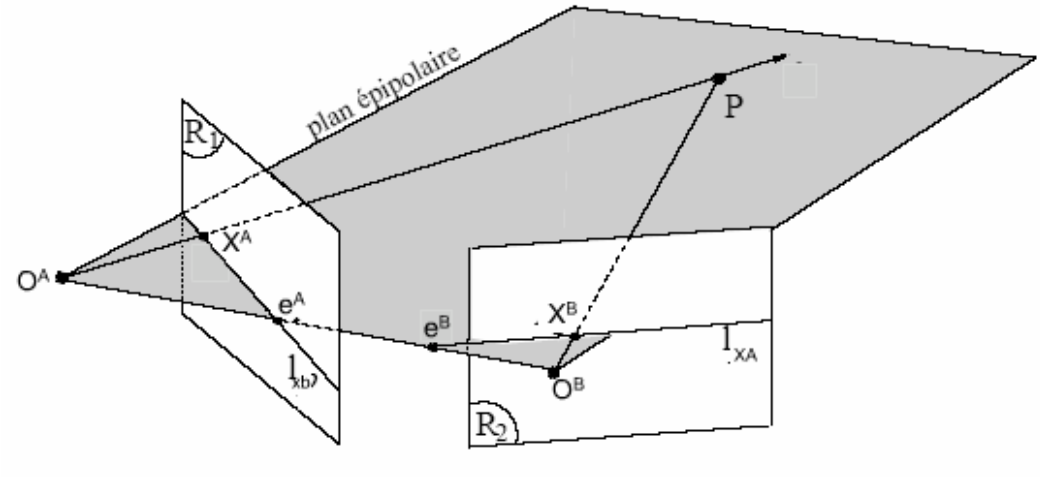


***Figure II.23** : contrainte épipolaire*

Pour renforcer encore cette contrainte les deux images stéréo sont rectifiées. La rectification d'images consiste à recalculer, pour deux images en position générale, deux nouvelles images telles que la géométrie épipolaire de ces deux images soit simple; c'est à dire que les droites épipolaires sont horizontales, ce qui implique que les deux nouveaux épipoles soient à l'infini.

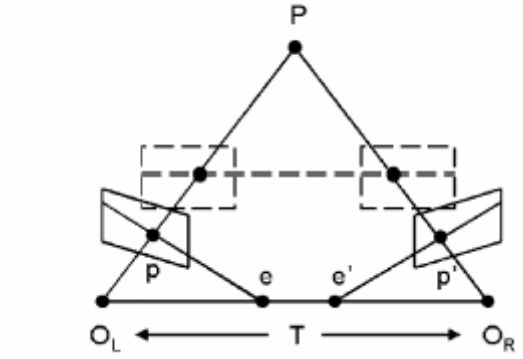


***Figure II.24**: rectification de la paire stéréo*

Une solution consiste à garder les deux centres de projection comme nouveaux centres de projection et à utiliser comme nouveaux plans images un seul et même plan contenant la direction de la droite liant les deux centres de projection. Ce plan n'étant pas défini de manière unique, il faut choisir une orientation. On peut alors considérer que le nouveau plan rétinien contient la direction de la droite intersection des plans rétiniens des images originales.

3.3.1.2 Contrainte de limites de disparité

La contrainte de disparité est issue de la calibration de la caméra. Lorsqu'on dispose d'un ensemble statique caméra - scène on peut évaluer à partir d'images test (avec une mire

par exemple) les disparités minimums et maximums qui apparaîtront dans une paire d'images de cette scène. On pourra ainsi limiter l'espace de recherche d'un point homologue sur la ligne épipolaire au segment de droite centré sur la position virtuelle du point p dans l'image droite.

3.3.1.3 Contrainte d'ordre

Les contraintes épipolaires et de disparité permettent de réduire le nombre d'appariements possibles entre les primitives de l'images gauche et les primitives de l'image droite. Il s'agit maintenant d'éviter les aberrations liées aux configurations particulières du capteur stéréoscopique ou des objets de la scène. La contrainte d'ordre implique que la projection des objets d'une scène conserve le même ordre dans les deux projections images. Si un point p se trouve à gauche d'un point q dans l'image de gauche, la contrainte d'ordre exige que les points correspondants p' et q' dans l'image droite soient dans le même ordre.

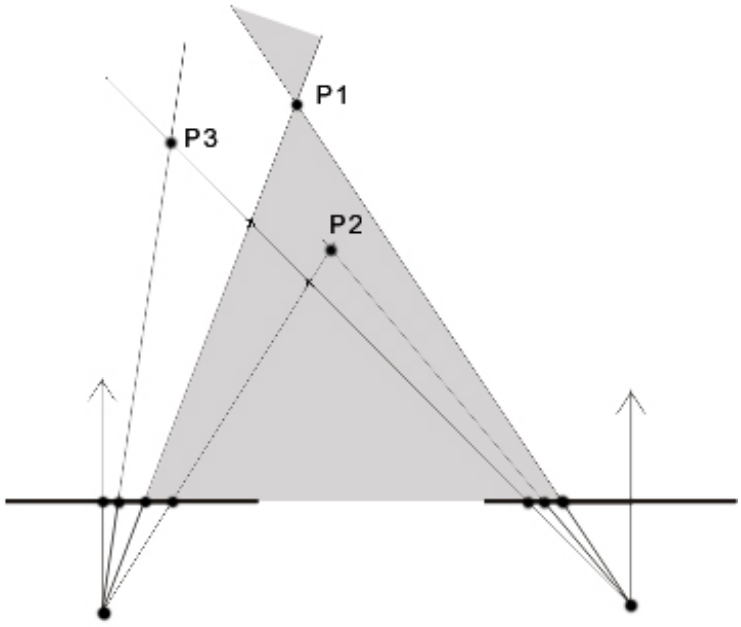


***Figure II.25** : Contrainte d'ordre : P3 et P1 ont le même ordre mais P1 et P2 sont inversés dans l'image de gauche*

3.3.1.4 Contrainte d'unicité

La contrainte d'unicité est vérifiée si tout point de l'image gauche possède au plus un correspondant dans l'image droite. La contrainte d'unicité découle directement de la contrainte d'ordre dans le sens ou elle décrit un cas limite de la contrainte d'ordre. Lorsque deux points P1 et P2 se projettent en un même point p dans l'image de gauche et en deux points différents p'1, p'2 dans l'image de droite.

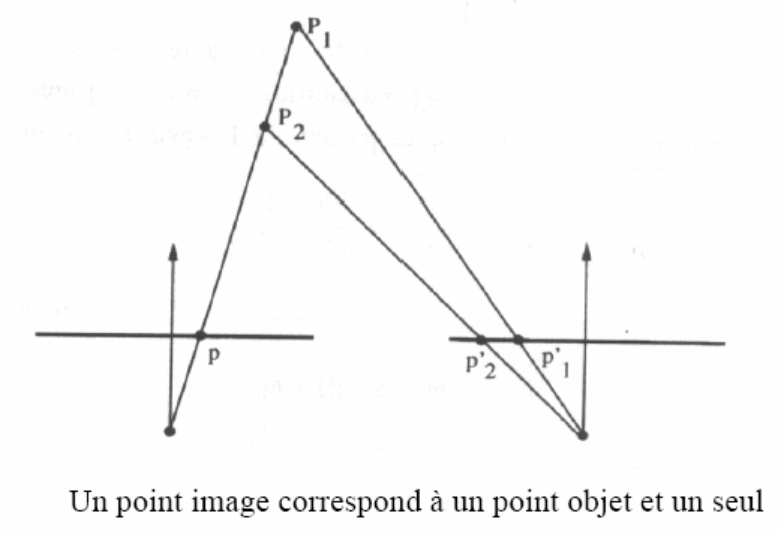


***Figure II.26** : Contrainte d'unicité*

3.3.2 Contraintes figurales

Les contraintes figurales sont définies comme telles car elles n'ont pas à l'instar des contraintes géométriques de base théorique, néanmoins, elles peuvent elles aussi imposer une contrainte d'ordre géométrique. On trouvera aussi dans cette catégorie toute contrainte spécifique à une primitive particulière.

3.3.2.1 Disparité locale constante

Lorsqu'une scène comporte des objets proéminents dont la surface est fortement inclinée par rapport aux plans des deux images, il y a un risque que la contrainte d'ordre ne soit pas respectée. Pour éviter cela on peut imposer une limite au gradient de disparité entre deux paires de points appariés consécutifs. Ce faisant on limite, en fait, l'inclinaison des surfaces des objets présents dans la scène, et de fait, on limite par là même la variété des objets à reconstruire.

3.3.2.2 Continuité figurale

Cette contrainte limite les variations de disparité le long des contours. On évite ainsi autant que faire se peut que les points d'un contour dans l'image gauche soient appariés avec des points de plusieurs contours dans l'image droite. Cela suppose aussi qu'un contour peut appartenir à plusieurs surfaces mais ne doit pas traverser le bord d'un objet. Ce qui peut arriver malgré tout.

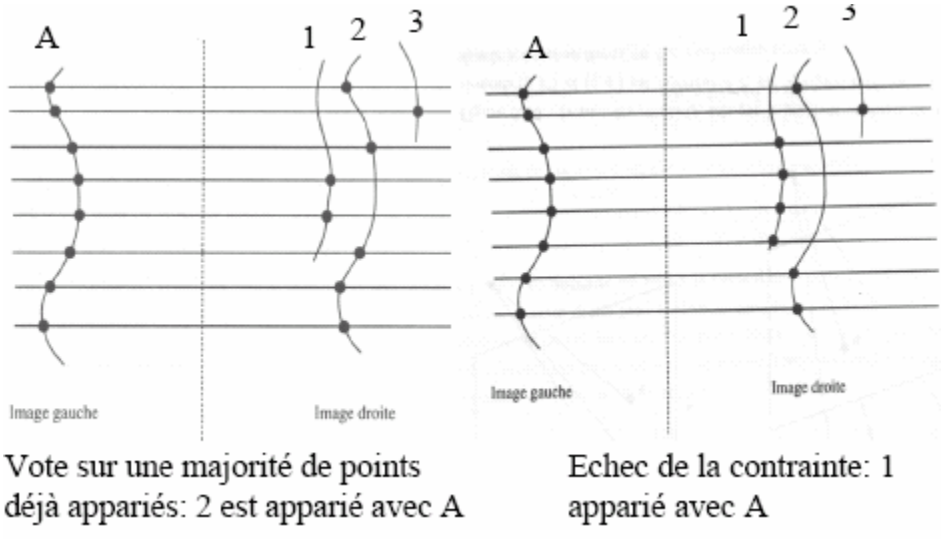


***Figure II.27** : Contrainte de continuité figurale*

Seuls les critères d'ordre géométrique possèdent une véritable base théorique, issue de la calibration. Les critères d'ordre figural n'ont qu'une base empirique et reposent principalement sur la ressemblance entre les éléments que l'on peut trouver dans la paire d'images.
Néanmoins, les critères géométriques ne peuvent s'appliquer que de manière ponctuelle ou sur des alignements de pixels. Il est en effet difficile d'appliquer une contrainte d'ordre géométrique sur un contour ou une région. A l'inverse les critères morphologiques s'appliquent particulièrement bien sur des primitives d'un niveau supérieur au pixel comme les contours, ou les régions, mais sont dépendants de la ressemblance apparente qui peut exister entre deux primitives. Notamment, lorsque la distorsion perspective devient importante du fait que les prises de vue sont relativement éloignées l'une de l'autre (aussi bien en translation qu'en rotation), ces critères morphologiques ne sont plus valables car la ressemblance apparente entre les deux représentations devient trop faible.

3.4 Les méthodes de la mise en correspondance

La mise en correspondance est un domaine vaste, où deux approches différentes se confrontent en traitement d'images [6,7,44].

- Les uns cherchent à mettre en correspondance des primitives d'images (segments, régions, contours…), à la suite d'une segmentation. La méthode consiste alors à trouver des points d'intérêt dans les images stéréo, et à déterminer le déplacement de ceux-ci, s'ils se conservent, d'une image à la suivante. Ces méthodes de mise en correspondance "éparse"appelées communément *méthodes locales* ont un atout, qui est la faible complexité algorithmique au regard de la quantité d'information à traiter.

- Une deuxième classe de méthodes, permet d'obtenir d'une information "dense" par la mise en correspondance de la réflectance de tous les éléments (points) de la scène. Lorsque l'on utilise des images en niveaux de gris, c'est la luminance de chaque pixel qui est comparée d'une image à l'autre (avec l'hypothèse que la luminance se conserve). Ces méthodes sont appelées communément *méthodes globales*.

3.4.1 Méthodes locales d'appariement

Les méthodes présentées dans cette section utilisent des primitives pour évaluer les appariements [44]. Des contraintes locales prennent en considération un voisinage plus ou moins restreint autour des pixels de la primitive considérée, ce voisinage étant contraint à l'intérieur d'une frontière définie, une fenêtre la plupart du temps. On observe une fenêtre autour d'un pixel de l'image de référence et on balaie la droite épipolaire correspondante de la seconde image avec une fenêtre de mêmes dimensions et on cherche à y identifier la position pour laquelle la fenêtre glissante est la plus similaire. Dans le cas d'images rectifiées, cette recherche s'effectue le long de la ligne horizontale correspondante. Deux familles de méthodes seront présentées ici : celles utilisant une mesure de similarité et celles utilisant le gradient de l'intensité dans l'image, aussi appelées méthodes de flux optique.

3.4.1.1 Mesures de similarité

Une première approche classique est d'utiliser une mesure de similarité entre les deux voisinages comparés. Un pointage est attribué à chaque appariement possible et celui auquel le pointage le plus élevé est attribué est l'appariement retenu. On considère que les surfaces observées sont Lambertiennes, c'est-à-dire que l'illuminance perçue n'est pas fonction de la direction d'observation, ce qui permet de considérer qu'un point réel présentera la même illuminance dans les deux images (si on suppose les deux caméras identiques). Dans l'immense majorité des cas, les mesures de ressemblances sont implantées par des mesures de corrélations.
Une première mesure de similarité est la corrélation normalisée, dont les valeurs sont comprises dans l'intervalle [-1, 1]. Soit deux images ***a*** et ***b*** de dimensions identiques, soit ***Ii(u,v)*** l'intensité du pixel ***(u,v)*** dans l'image ***i, k*** un voisinage du pixel ***(u,v)***, **d** la disparité entre les deux pixels pour laquelle on veut comparer les voisinages et soit ***μi*** la moyenne de l'intensité dans la fenêtre de l'image ***i***, le coefficient de corrélation normalisée [21] est donné par :

$$C(u,v) = \frac{\sum_k (I_1(u,v) - \mu_1(u,v)) * (I_2(u+d,v) - \mu_2(u,v))}{\sqrt{\sum_k (I_1(u,v) - \mu_1(u,v))^2 * \sum_k (I_2(u+d,v) - \mu_2(u,v))^2}}$$

Le coefficient de corrélation mesure la similarité entre deux régions : plus les régions sont similaires, plus le coefficient aura une valeur près de 1 ; plus les régions sont différentes, plus le coefficient aura une valeur près de 0. Des valeurs négatives de ce coefficient (entre *0* et *-1*) indiquent une similarité 'opposée' entre les régions. Par exemple, une fenêtre blanche (255) et une fenêtre noire (0) auraient un coefficient de corrélation normalisée de -1 : une forte relation entre les deux fenêtres, mais avec des intensités opposées. On calcule ce coefficient pour chacune des combinaisons entre la fenêtre dans l'image de référence et toutes les fenêtres de mêmes dimensions le long de la droite épipolaire dans la seconde image. Après avoir calculé chacun de ces coefficients, on choisit celui correspondant à l'appariement ayant la valeur de similarité la plus élevée (Figure 4).

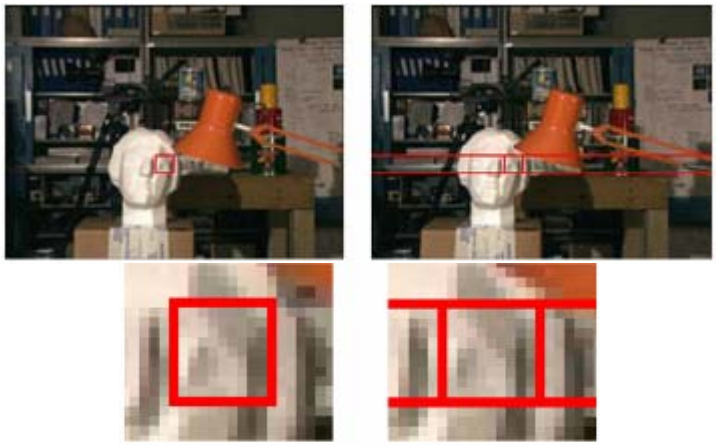


***Figure II.28** : Corrélation : Au haut, les images gauche et droite avec la fenêtre de Référence. Au bas, une vue rapprochée des deux fenêtres appariées*

Le coefficient de corrélation normalisée n'est pas la seule mesure de similarité pouvant être utilisée pour effectuer la comparaison de voisinages. D'autres mesures [7] telles que la somme des différences absolues ou la somme des différences carrées sont également utilisées. Celles-ci sont plutôt appelées mesure de dissimilarité puisqu'elles évaluent la différence entre deux régions et non pas la similarité comme le coefficient de corrélation normalisée. Dans ces deux cas, on calcule la différence d'intensité entre les pixels correspondants dans la fenêtre encadrant les voisinages comparés. On effectue ensuite la somme de ces différences (ou du carré de ces différences) pour obtenir le coût associé à cet appariement.
On peut également, de la même manière, utiliser la variance comme mesure de dissimilarité. Comme pour la corrélation normalisée, on calcule des coefficients pour tous les appariements possibles et on retient celui ayant le coût le moins élevé (différence la plus faible). Il est intéressant de constater que même si cette approche simple fut parmi les premières à être proposées, elle est encore parmi les plus utilisées, même au niveau de produits commerciaux. On constate donc la complexité du problème et on comprend pourquoi le sujet est encore très actif au niveau de la recherche.

MATCH METRIC	DEFINITION
Normalized Cross-Correlation (NCC)	$\frac{\sum_{u,v}\left(I_1(u,v)-\bar{I}_1\right)\cdot\left(I_2(u+d,v)-\bar{I}_2\right)}{\sqrt{\sum_{u,v}\left(I_1(u,v)-\bar{I}_1\right)^2\cdot\left(I_2(u+d,v)-\bar{I}_2\right)^2}}$
Sum of Squared Differences (SSD)	$\sum_{u,v}\left(I_1(u,v)-I_2(u+d,v)\right)^2$
Normalized SSD	$\sum_{u,v}\left(\frac{\left(I_1(u,v)-\bar{I}_1\right)}{\sqrt{\sum_{u,v}\left(I_1(u,v)-\bar{I}_1\right)^2}}-\frac{\left(I_2(u+d,v)-\bar{I}_2\right)}{\sqrt{\sum_{u,v}\left(I_2(u+d,v)-\bar{I}_2\right)^2}}\right)^2$
Sum of Absolute Differences (SAD)	$\sum_{u,v}\left\|I_1(u,v)-I_2(u+d,v)\right\|$

***Tableau II.01** :Quelques mesures de similarité*

3.4.1.2 Méthodes de gradient

Une seconde approche tente plutôt d'estimer le mouvement entre les deux images de la paire stéréo pour en déduire une mesure de disparité [44]. Il s'agit des méthodes de gradient, également appelées méthodes de flux optique. En estimant le mouvement horizontal entre les deux images (rectifiées), c'est en fait la disparité que l'on estime.
On considère toujours les surfaces observées comme étant lambertiennes. On tente alors de déterminer la translation horizontale d'un même point entre les deux images. On compare en premier lieu la valeur de l'intensité d'un même pixel (même position) dans les deux images et on note la différence entre ces deux valeurs (**Et**). On calcule ensuite le gradient d'intensité de l'image à ce pixel dans l'image de référence. On pose ensuite l'équation différentielle suivante :

$$\nabla_x(I) * d + E_t = 0,$$

où ∇_x est la composante horizontale du gradient. Selon cette équation, on cherche donc la valeur du déplacement **d** qui, selon la composante horizontale du gradient, engendrera la différence d'intensité **Et**. On utilise donc le gradient de l'intensité de l'image pour estimer le déplacement dans la direction de la composante horizontale de celui-ci qui engendrerait l'intensité du même pixel dans la seconde image.
On émet l'hypothèse que l'intensité varie de manière constante dans le voisinage du pixel à apparier. La valeur du gradient pouvant être largement différente pour des pixels voisins, cette approche ne peut être utilisée que pour l'estimation précise de très faibles déplacements. En théorie, cette approche ne peut estimer que des valeurs de déplacement inférieures à 1/2 pixel car les dérivées locales ne sont valides que sur cette plage. Certaines approches peuvent être utilisées pour estimer de plus grands déplacements, par exemple des approches de traitement hiérarchique [45]. De plus, le fait d'utiliser uniquement la composante horizontale du gradient peut engendrer des imprécisions.

3.4.2 Forces et lacunes des méthodes locales

Ces méthodes présentent des forces et des lacunes similaires. Au niveau des forces, mentionnons la simplicité et la rapidité des calculs. Ces caractéristiques se prêtent bien aux applications en temps réel et aux implémentations matérielles d'algorithmes de stéréo dans les travaux de recherche récents. Au niveau des lacunes, mentionnons en premier lieu que ces algorithmes performent moins bien sur des zones de texture uniforme. Prenons par exemple le cas extrême où les deux images représentent un mur blanc mât et sans ombrage : dans un tel cas, les mesures de similarité auront pratiquement la même valeur partout dans l'image, ce qui rendra impossible de discerner le meilleur appariement possible pour un pixel. Pour ce qui est des méthodes de gradient, la valeur du gradient sera pratiquement nulle en tout point de l'image, tout comme celle du mouvement estimé à partir de celle-ci.
Une seconde lacune qui affecte les méthodes locales et tous les algorithmes stéréo actuels est celui des occultations. Une occultation est décrite comme étant une section de la scène qui est visible dans une seule des deux images. Les occultations se produisent habituellement le long de discontinuités de profondeur de la scène comme dans la *figure II.29*. On observe qu'une section, dont la profondeur est moindre que celle du reste de la scène, cache certaines parties de celle-ci pour une ou l'autre des caméras. Au moment de calculer la similarité entre les voisinages, la fenêtre sur laquelle le calcul est effectué contient des pixels qui sont présents dans une seule image, ce qui vient perturber la valeur de similarité.

Dans le cas des méthodes de gradient, une discontinuité de profondeur ou une occultation engendre une valeur de gradient très élevée, ce qui produit une très faible valeur de disparité ne correspondant pas à la réalité. On risque ainsi de se retrouver avec des appariements choisis n'étant pas représentatifs de la scène réelle. Certaines techniques ont été développées afin de contrer ces problèmes. Une première technique consiste à détecter les occultations pour ensuite effectuer une interpolation dans les zones de disparités inconnues en fonction des disparités voisines. L'interpolation s'effectuant pour des zones inconnues de la scène, cette méthode vise plus à produire un résultat esthétique que conforme à la scène 3D.

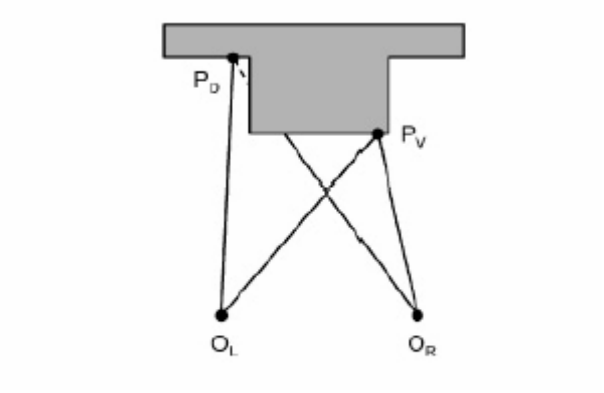


***Figure II.29** : Exemple d'occultation*

Une seconde approche consiste à imposer des contraintes supplémentaires aux appariements. Une de ces contraintes est appelée 'contrainte de continuité' et indique que : soit b' le correspondant de b et a' le correspondant de a, si b est à la droite de a, alors b' sera également à la droite de a'. Pour cette contrainte, on assume que la scène observée est plutôt planaire et régulière et que la situation illustrée à la *figure II.30* ne se produit pas.

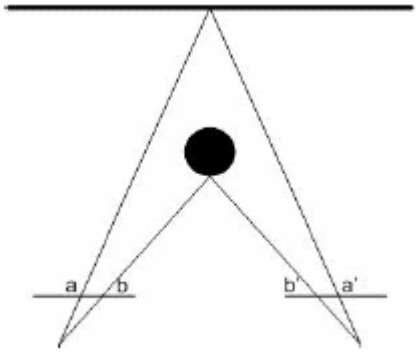


***Figure II.30** : Un cas où la contrainte de continuité n'est pas respectée*

3.4.3 Méthodes globales d'appariement

Les méthodes locales ont pour limitations principales le fait de produire une carte de disparités incomplète et d'offrir de faibles performances dans certaines régions problématiques. Afin de contourner ces problèmes, il faut appliquer des contraintes tenant compte d'informations plus complètes que le simple voisinage du pixel à apparier. Certaines de ces contraintes tiennent compte d'une ligne complète dans l'image, d'autres de l'image entière. Deux méthodes seront présentées dans cette catégorie.

3.4.3.1 Programmation Dynamique

Une première approche utilise la programmation dynamique pour résoudre un problème de minimisation sur une ligne complète de l'image [46]. On utilise une mesure de similarité afin de calculer un coût d'appariement pour toutes les combinaisons (x,d) possibles le long d'une ligne de l'image de référence. La première étape est donc identique à celle des méthodes locales. La différence se situe au niveau des contraintes : en assumant la scène observée comme étant plutôt régulière (peu de discontinuités de profondeur) et sachant que celles-ci engendrent des discontinuités au niveau des disparités, on peut imposer une contrainte de continuité de la disparité le long de cette ligne.

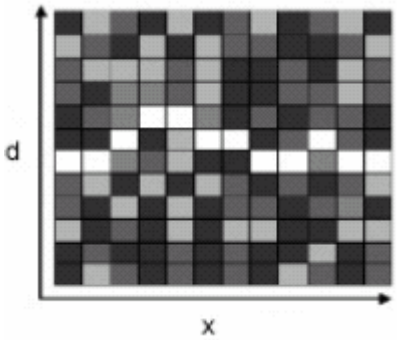


***Figure II.31** : Exemple de chemin minimal (en blanc) dans l'espace (x,d). Les niveaux de gris représentent les valeurs de dissimilarité.*

Avec cette contrainte, on peut favoriser (défavoriser) la présence de discontinuités de profondeurs dans les appariements retenus. Cette contrainte consiste en un coût de pénalité attribué pour chaque discontinuité dans la disparité. Lorsque deux pixels voisins n'ont pas la même disparité, on ajoute au coût d'appariement une valeur kP où P est le coût de pénalité et k correspond à la différence de disparité entre les deux pixels. En modifiant la valeur de P, on peut jouer sur la régularité du résultat : une valeur plus grande donnera une solution plus régulière tandis qu'une valeur plus petite engendrera une plus grande présence de discontinuités. A l'extrême, une valeur nulle produira un chemin comportant un maximum de discontinuités tandis qu'une valeur infinie produira un chemin droit (disparité constante). On utilise généralement une valeur de P constante pour une même paire d'images. Avec cette approche, plutôt que de considérer chaque pixel individuellement, on recherche la série d'appariements dont la somme des erreurs est minimale sur la longueur de la ligne. On peut formuler le problème comme étant la recherche du chemin dans l'espace (x,d) (*Figure II.31*) dont la somme des coût s d'appariement est minimisée. On cherche donc à minimiser la fonction suivante :

$$\phi = \sum_{u=1}^{x} d(u) + Pk(u) \qquad (*)$$

où d(u) est la mesure de dissimilarité du pixel u et Pk(u)le terme représentant le coût de pénalité pour l'appariement retenu. Ce chemin dans l'espace (x,d) pourrait se rechercher récursivement, mais on utilise plutôt les techniques de programmation dynamique [46], d'où le nom, pour des questions de performance.
Contrairement aux méthodes locales, cette approche assure une valeur de dis parité en tout point de l'image de référence, parfois au détriment de la précision de celles-ci.

Figure II.32 : Exemple de résultat obtenu par la méthode de programmation dynamique. On remarque les traits horizontaux irréguliers entre les lignes.

Bien qu'assurant une cohérence horizontale, les méthodes effectuant des optimisations sur une seule ligne ne parviennent pas à produire des résultats cohérents entre chacune de ces lignes puisque aucune contrainte ne permet d'utiliser les informations des lignes voisines. Chacune des lignes est donc traitée indépendamment de ses voisines, ce qui produit parfois des résultats insatisfaisants : une erreur locale à un pixel se retrouve propagée à la grandeur de la ligne, ce qui engendre des séquences horizontales de valeurs de disparité erronées *(Figure II.32)*. Pour éviter ce problème, il faut être en mesure d'appliquer des contraintes sur l'ensemble de l'image, ce qui est possible de faire en utilisant une approche de flux maximal ('max-flow').

3.4.3.2 Max-flow / min-cut

Afin de corriger le problème d'incohérence interligne rencontré avec la programmation dynamique, il semble logique d'empiler plusieurs plans (x,d) pour ensuite rechercher une surface dans le volume ainsi créé (*Figure II.33*). En utilisant une telle structure, il est possible d'imposer une contrainte de cohérence locale tenant compte du voisinage d'un pixel dans toutes les directions et non plus le long d'une seule ligne comme précédemment. Cette nouvelle contrainte permet d'attribuer une pénalité pour les discontinuités de disparité entre lignes voisines. On est ainsi en mesure de minimiser la même fonction (équation (*)) à l'intérieur de ce volume.
Il est par contre excessivement complexe de minimiser cette fonction en utilisant la programmation dynamique. La contrainte d'ordre sur une même ligne épipolaire, sur laquelle cette méthode est basée, ne peut plus s'appliquer dans ce volume : il n'y a pas d'ordre précis pour la construction de la surface recherchée, contrairement au chemin recherché sur une seule ligne (i.e. gauche à droite). Il faut donc formuler le problème autrement.

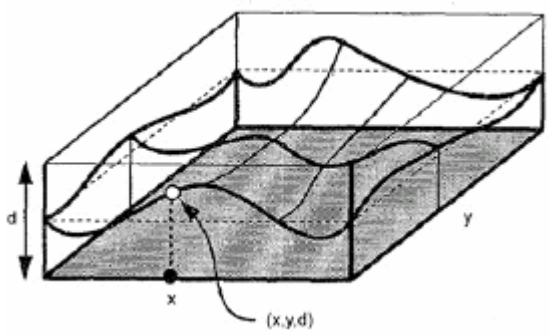


Figure II.33 : *Volume engendré en empilant les plans (x,d) pour chaque ligne de l'image*

Roy et Cox [47][48] ont proposé de le formuler en tant que calcul d'un flux maximal à travers un volume présenté sous forme de graphe (*Figure II.34*). Le volume en question englobe la surface à reconstruire et est en fait un échantillonnage de l'espace 3D ayant comme référence la caméra (image) pour laquelle on désire construire la carte de profondeur. à noter qu'on parle désormais de carte de profondeur et non de carte de disparité. Selon cette approche, la valeur calculée est la profondeur z du point et non une valeur de disparité. La différence est mineure étant donné que la disparité et la profondeur sont reliées à une constante près (la longueur f * base de triangulation). Cette différence offre cependant un avantage sur les autres méthodes : étant donné que l'on considère la profondeur du point et non plus une valeur de correspondance entre deux images, il est possible d'utiliser plusieurs images afin d'identifier la position du point 3D. On projette celui-ci dans chacune des images et on effectue la comparaison entre les voisinages de chaque pixel selon une mesure de dissimilarité appropriée (e.g. la variance). On peut ainsi utiliser l'information de plusieurs images sans se soucier de la géométrie épipolaire et sans avoir a rectifier les images : il suffit de calibrer les caméras par rapport au référentiel de la scène.

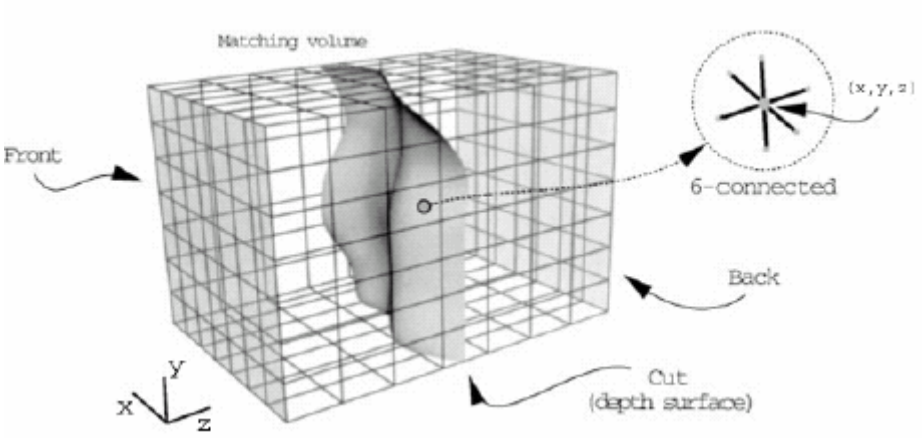


***Figure II.34** : Volume 3D défini en reliant chaque noeud à ses voisins.*

On ajoute une source s et un drain t au volume *(Figure II.35)*. Le flux quitte la source, traverse le volume et rejoint le drain. à l'intérieur du graphe, chaque noeud est relié à ses six voisins immédiats. La source est reliée à tous les nœuds du plan z = 0 et le drain est connecté à tous les noeuds du plan z = zmax. Le volume est donc défini par l'ensemble des noeuds V :

$$V = V^* \cup \{s, t\},$$

où s est la source, t le drain et V* est l'ensemble des autres noeuds :

$$V^* = \{(x', y', z') : x' \in [0...x'_{max}], y' \in [0...y'_{max}], z' \in [0...z'_{max}]\}, \qquad (6)$$

où (x'_{max} +1, y'_{max} +1) sont les dimensions de l'image et z'_{max} +1 est la valeur maximale en profondeur du volume observé. L'ensemble des arcs reliant chacun des noeuds du graphe est défini par :

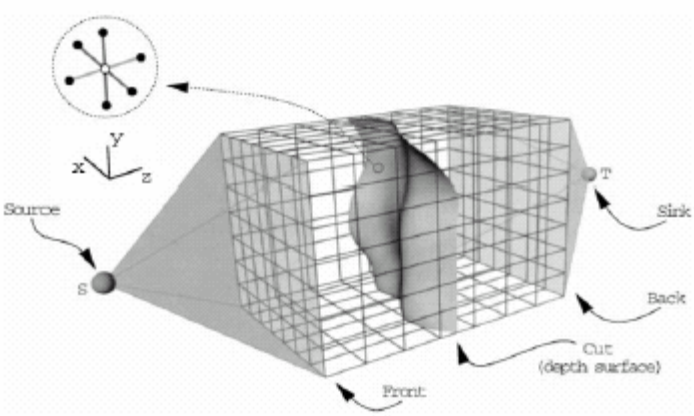


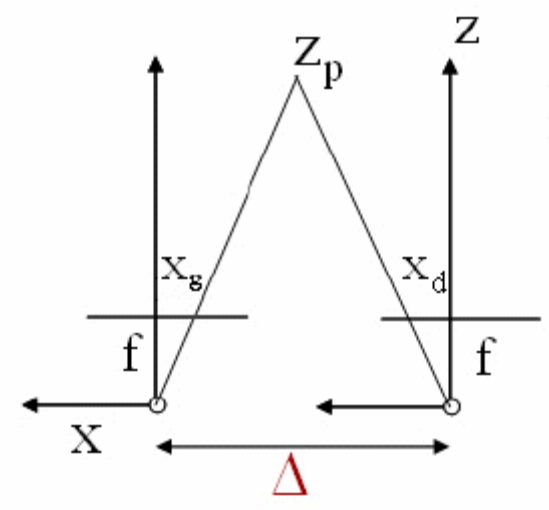
***Figure II.35** : Volume 3D auquel on ajoute une source et un drain.*

4. La reconstruction 3D

La reconstruction 3D est une interprétation géométrique particulière de ce qui est observable dans un espace image caractérisé par sa géométrique épipolaire [5]. Le capteur stéréoscopique constitué de deux caméras projectives fournit des paires d'images stéréo. La reconstruction 3D à partir d'une paire d'images suppose que l'on dispose de correspondances entre ces deux images. La reconstruction est le processus inverse de la formation d'images.

En lançant, pour chaque caméra, un rayon passant par le centre de projection et par le plan image à la position du pixel donné, il suffit de trouver le point d'intersection de ces deux rayons qui correspond à la position 3D du point observé. Si les images sont rectifiées, les droites épipolaires étant horizontales et à la même hauteur dans les deux images, la correspondance entre les deux projections d'un même point est donc réduite à une valeur unidimensionnelle, l'obtention de la position en z du point est obtenue par simple triangulation *(Figure II.36)*. La profondeur **z** n'est alors fonction que de la disparité **d.** Ainsi, pour chaque point d'une image dont on connaît le correspondant dans la seconde image, nous sommes en mesure de retrouver sa position 3D dans un référentiel donné (caméra gauche, caméra droite, référentiel positionné au milieu de la base de triangulation (baseline)). Si la correspondance entre tous les pixels d'une paire d'images est connue, on peut reconstruire un modèle 3D complet de la scène observée. La liste de ces correspondances est appelée 'carte de disparités'. Pour une paire d'images rectifiées, cette carte peut être présentée

Soient deux images $\mathbf{I_g}$ et $\mathbf{I_d}$, $\mathbf{x_g}$ et $\mathbf{x_d}$ les coordonnées en x des pixels correspondant à la projection d'un même point de la scène dans les deux images, cette correspondance est la disparité d par rapport à l'image de référence $\mathbf{I_g}$: $\mathbf{d = x_g - x_d}$.



$$\frac{\Delta + x_g - x_d}{z_p - f} = \frac{\Delta}{z_p}$$

$$\frac{\Delta - d}{z_p - f} = \frac{\Delta}{z_p}$$

$$z_p = \frac{\Delta f}{d}$$

Où **Δ** est la base de triangulation
d la disparité

***Figure II.36** : La triangulation*

Le type de reconstruction qu'on peut alors effectuer dépend du type de calibrage dont on dispose. On peut distinguer les cas suivants [3] :

1. le capteur stéréoscopique est calibré et on dispose alors de paramètres intrinsèques de chaque caméra ainsi que de la transformation rigide entre les deux caméras (paramètres extrinsèques): dans ce cas on obtient une reconstruction euclidienne dans le repère de calibrage ;
2. les paramètres internes de chaque caméra sont connus mais la transformation rigide entre les deux caméras n'est pas connue. Il faut alors estimer la matrice essentielle [42] à partir de laquelle on peut extraire la transformation rigide entre les deux caméra à un facteur d'échelle prés : dans ce cas on obtient une reconstruction euclidienne dans le repère de l'une ou l'autre des deux caméras ;
3. si aucun calibrage n'est disponible (ni les paramètres intrinsèques ni les paramètres extrinsèques des deux caméras) il faut alors estimer la matrice fondamentale (calibrage) comme il a été expliqué auparavant. A partir de la matrice fondamentale on peut obtenir une reconstruction projective tri-dimensionnelle [43].

Si on se place dans le premier de ces cas et si le point **p** de l'image de gauche a été mis en correspondance avec le point **p'** de l'image de droite, on a alors deux ensembles d'équations:

$$\begin{aligned}
u &= \frac{m_{11}X + m_{12}Y + m_{13}Z + m_{14}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \\
v &= \frac{m_{21}X + m_{22}Y + m_{23}Z + m_{24}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \\
u' &= \frac{m'_{11}X + m'_{12}Y + m'_{13}Z + m'_{14}}{m'_{31}X + m'_{32}Y + m'_{33}Z + m'_{34}} \\
v' &= \frac{m'_{21}X + m'_{22}Y + m'_{23}Z + m'_{24}}{m'_{31}X + m'_{32}Y + m'_{33}Z + m'_{34}}
\end{aligned}$$

Les coordonnées X, Y et Z du point **P** reconstruit, dans le repère de calibrage, se calculent en résolvant ce système de 4 équations linéaires. On peut également reconstruire le point **P** dans le repère de la caméra de gauche en utilisant les équations de triangulation *Figure II.36*. Les coordonnées X, Y et Z du point **P** seront données dans ce cas par :

$$\begin{aligned} x' &= \frac{(r_{11}x + r_{12}y + r_{13})Z + b_x}{(r_{31}x + r_{32}y + r_{33})Z + b_z} \\ X &= xZ \\ Y &= yZ \end{aligned}$$

Ces équations nous permettent de constater que le déplacement entre la caméra gauche et la caméra droite doit compter une translation. Si la transformation gauche-droite est une rotation pure, la reconstruction n'est pas possible.
Finalement on peut remarquer que lorsque les images ont été rectifiées la matrice décrivant la transformation gauche-droite est réduite à une translation.
Le point **P** peut alors être reconstruit dans le repère rectifié de la caméra gauche grâce aux équations suivantes :

$$\begin{aligned} Z &= \frac{b}{y' - y} \\ X &= xZ \\ Y &= yZ \end{aligned}$$

et on remarque que, d'un point de vue géométrique, la reconstruction n'est rien d'autre qu'une triangulation.

5. Les méthodes de rendu et de modélisation à base d'images (IBMR)

Les techniques de rendu et de modélisation à base d'images (*IBMR)* utilisent principalement les informations photométriques (couleur et intensité en chaque pixel) et géométriques (profondeur de chaque pixel, paramètres des caméras) d'une image pour synthétiser de nouvelles vues.

La façon dont sont utilisées ces informations permet de distinguer les méthodes à base d'images. En effet, ils utilisent, déforment et combinent une ou plusieurs images selon différentes techniques qui vont du simple placage de texture au transfert épipolaire en passant par la déformation 3D ainsi que différentes méthodes d'interpolation. Il existe trois approches principales au problème de la synthèse de nouvelles vues à partir des images réelles [1,36] :

- Les techniques fondées exclusivement sur le rendu par les images. Ces techniques essaient de générer des vues de synthèse à partir d'un ensemble d'images originales. Elles n'estiment pas la vraie structure 3D à partir des images, mais utilisent, interpolent ou déforment directement l'ensemble des vues originales pour générer une nouvelle vue.

- Les techniques basées images / basées géométrie qui mélangent le rendu par les images avec une reconstruction 3D partielle utilisant la géométrie (implicite ou explicite) de la scène. Dans ce type d'approches l'objectif est de générer des vues cohérentes de la scène réelle, pas d'avoir des mesures précises, excepter le cas du 3D scanning qui essaient de retrouver complètement la structure 3D sous-jacente.

- Les techniques hybrides combinent les techniques de modélisation standard avec des vues de la scène réelle, soit pour accélérer certaines techniques de rendu existantes, soit pour tirer parti des avantages des deux approches. Ces techniques peuvent être utilisées pour la modélisation de scènes complexes.

5.1 Techniques de rendu purement à base d'images

5.1.1 Imposteurs

Dans le but d'obtenir une grande complexité visuelle au moindre coût, il existe une classe de techniques regroupées sous le terme d'*imposteurs*. Toutes ces méthodes effectuent un pré-rendu hors-ligne d'une partie de la scène. L'image créée est utilisée par la suite en temps que texture plaquée sur un rectangle positionné au sein de la scène. Ainsi, à la place de l'objet complexe est affiché son image : c'est un imposteur.

Les plus simples des imposteurs sont les « Panneaux d'affichage » (*billboards)* et les sprites. Un Panneaux d'affichage est un plan texturé placé dans la scène, représentant un objet complexe. Typiquement, les Panneaux d'affichage sont employés pour représenter des objets tels que des nuages ou des arbres, très difficiles à modéliser explicitement. On distingue les Panneaux d'affichage toujours parallèle à l'écran, ceux toujours orientés vers l'observateur et ceux qui sont de simples plans fixes dans l'espace de la scène. Chacun de ces types est adapté au rendu de certains types d'entités. Ceux toujours parallèles à l'écran ou orientés vers l'observateur servent par exemple pour représenter des nuages ou de la fumée. Un arbre, par contre, est souvent approximé par deux panneaux d'affichage perpendiculaires positionnés dans la scène.

Les sprites se différencient du panneau d'affichage par l'utilisation qui en est faite et leur mise en oeuvre : un sprite est généralement plus petit, en mouvement dans la scène et appliqué directement à l'écran. En résumé, un panneau d'affichage est un plan 3D tandis qu'un sprite est un rectangle 2D toujours parallèle à l'écran quelle que soit l'orientation de la caméra. Un exemple d'utilisation des imposteurs est celui de Horry *et al.* [25] qui utilisaient des panneaux d'affichage extraits de photographies dans le cadre de leur méthode nommée "Excursion dans une image" (*Tour Into the Picture)*. Le principe est de déformer une image en fonction de son point de fuite. Le point de fuite sert de repère pour la création (par l'utilisateur) d'un maillage recouvrant la scène, et la séparant en différentes zones (arrière plan, sol, bords, ...). Les panneaux d'affichage sont utilisés pour modéliser les objets au premier plan de l'image et sont déplacés lors de l'animation afin d'accompagner le mouvement de pénétration dans l'image. Leur technique, très artistique, a été utilisée dans de nombreux vidéoclips : le déplacement des panneaux d'affichage, combiné à une légère déformation de l'image donne l'impression de pénétrer dans l'image.

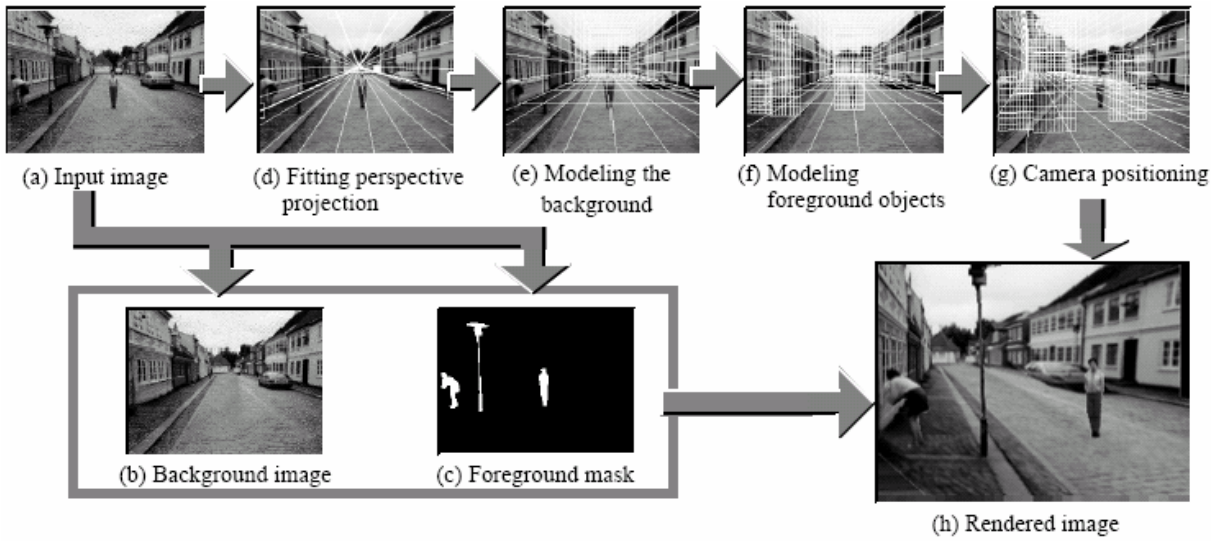


***Figure II.37** : Etapes de "tour into the picture"*

5.1.2 Forme à partir de X (Shape from X)

Cette catégorie de techniques permet de reconstituer la forme des surfaces par l'utilisation d'hypothèses fortes sur la nature des objets ou les conditions d'observations : illumination, textures et contours.

5.1.2.1 Forme à partir de contour (Shape from contour)

Les hypothèses peuvent porter sur les contours des objets observés. Le principe est de reconstituer les normales aux surfaces visibles à partir de leurs contours. Il est alors possible de reproduire la surface à partir d'un champ de normales. Cette approche donne d'assez bons résultats dans le cas d'objets dont la structure est connue à l'avance, comme par exemple des objets de révolution ou bien à base de cylindres généralisés [50].

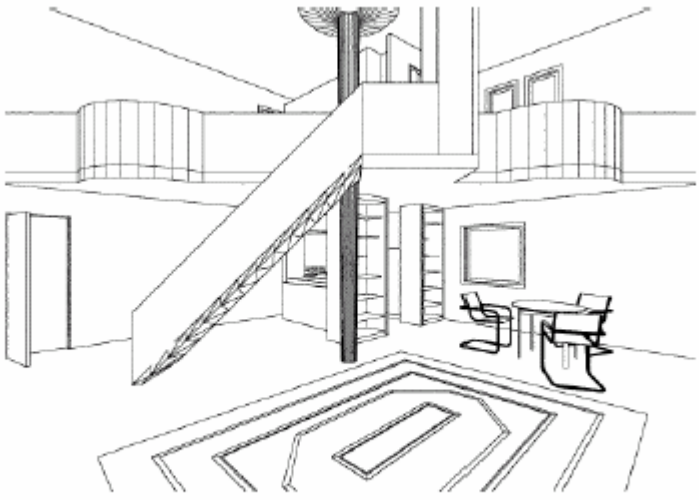


***Figure II.38 :** les contours permettent souvent d'interpréter les objets d'une scène et sa structure 3D*

5.1.2.2 Forme à partir de l'ombrage (shape from shading)

Cette catégorie de techniques permet de reconstituer la forme des surfaces par l'utilisation d'hypothèses fortes sur la nature des objets ou les conditions d'illumination.
Il est possible de calculer la structure tridimensionnelle d'un objet en observant uniquement la distribution de l'intensité lumineuse qu'il réfléchit [33]. Si les positions de la caméra et de la source lumineuse sont connues, ainsi que les lois de réflectance de l'objet observé, il est possible de calculer la forme de l'objet. Par exemple, si l'intensité lumineuse réfléchie en un point est $I = f(\vec{n}; \vec{s}; \vec{r})$, où $\vec{n}$ est la normale à l'objet en ce point, $\vec{s}$ la direction de la source lumineuse (rayon incident), et $\vec{r}$ la direction du rayon réfléchi (direction de vue), alors il est possible de calculer les normales $\vec{n}$ à la surface de l'objet, en tous les points où I, f, $\vec{s}$ et $\vec{r}$ sont connus, donc de calculer une équation de la surface de l'objet *(voir figure II.39)*. Les vecteurs $\vec{s}$ et $\vec{r}$ sont donnés par les positions de la source lumineuse et de la caméra, I est donné par la caméra, et f est issue d'un modèle de réflexion à priori.

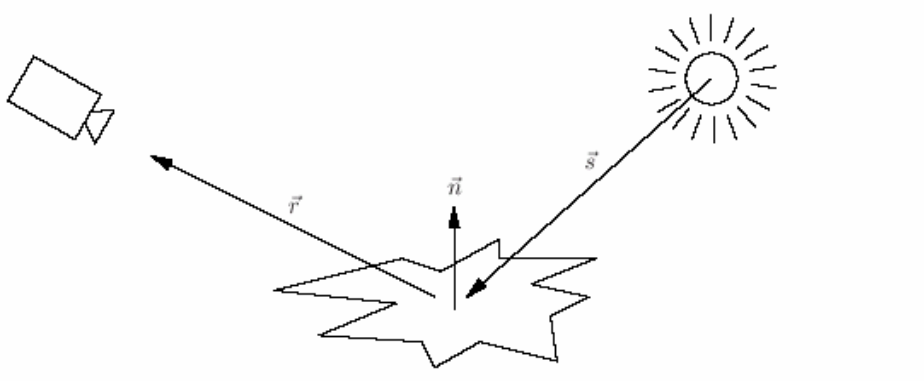


Figure II.39 : *Principe du shape from shading*

Cette approche donne d'assez bons résultats dans le cas d'objets dont la structure est connue à l'avance comme par exemple des objets de révolution ou bien à base de cylindres généralisés (*voir figure II.40*).
Même si elles donnent d'assez bons résultats dans des situations bien précises les hypothèses demandées par ces méthodes sont en général trop sévères pour une utilisation générale.

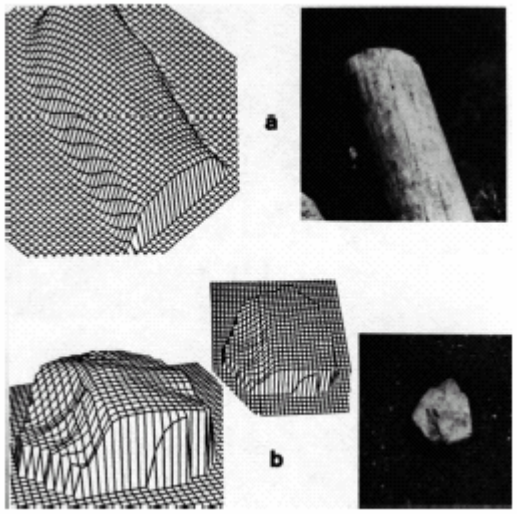


Figure II.40 : *Deux exemples de surfaces extraites à partir de l'illumination de la scène (shape from shading)*

5.1.2.3 Forme à partir de silhouette (shape from silhouette)

Une méthode appelée le 'shape from silhouette' utilise les silhouettes de l'objet pour reconstruire une forme approximative de ce dernier [37,38,51] .Une séquence d'images d'un objet disposé devant un fond connu est acquise depuis des points de vue entourant cet objet. Chaque image de la séquence est ensuite traitée afin d'extraire le fond de la forme. Cette information de contour est appelée silhouette qui permet de définir une sorte d'inéquation en disant que l'objet se trouve à l'intérieur du contour. Grâce aux différents points de vue on obtient un système complexe d'inéquations permettant d'englober l'objet dans un volume appelé enveloppe visuelle [37].

L'*enveloppe visuelle* est l'intersection des cônes des silhouettes d'entrée. Le volume qui résulte de l'intersection est une limite approximative de la forme de l'objet. Cet forme est cohérente avec les silhouettes de l'objet, c'est-à-dire que si l'on se place à chacun de ces points de vue, l'enveloppe visuelle donne la même vue que celle qui a servi à la construire en ce point. Plus de silhouettes utilisées, plus l'enveloppe visuelle convergera vers un volume qui est plus serré et plus proche de la forme de l'objet réel, mais ce volume ne convergera pas nécessairement à la géométrie réelle de cet objet. Ce là est dû à la présence potentielle des régions concaves sur la surface de l'objet qui sont difficiles, sinon impossible, à détecter en utilisant seulement des silhouettes. En dépit de cette imperfection, l'enveloppe visuelle est toujours une bonne approximation de la géométrie réelle de l'objet.

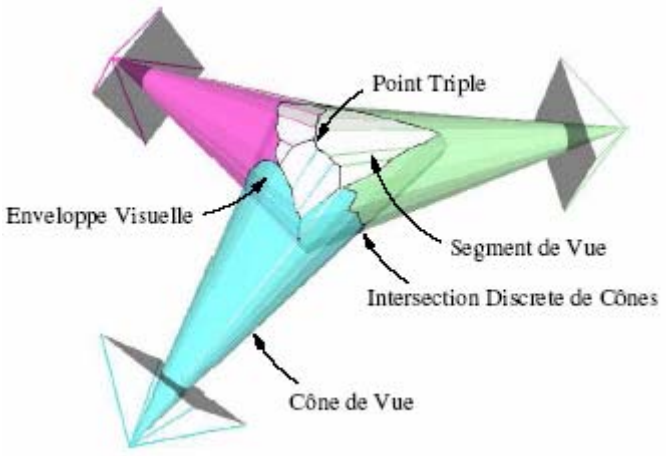


***Figure II.41** : Principe de la reconstruction de l'enveloppe visuelle :*
Enveloppe visuelle d'une sphère obtenue avec 3 vues.

L'enveloppe visuelle a été très étudiée, de manière implicite et explicite, dans les communautés de la vision par ordinateur et de l'image de synthèse. Certains s'intéressent au volume délimité par l'enveloppe visuelle et se basent sur des discrétisations de l'espace (approches volumétriques). D'autres visent à reconstruire la surface de l'enveloppe visuelle en fournissant des points isolés ou un maillage (approches surfaciques).

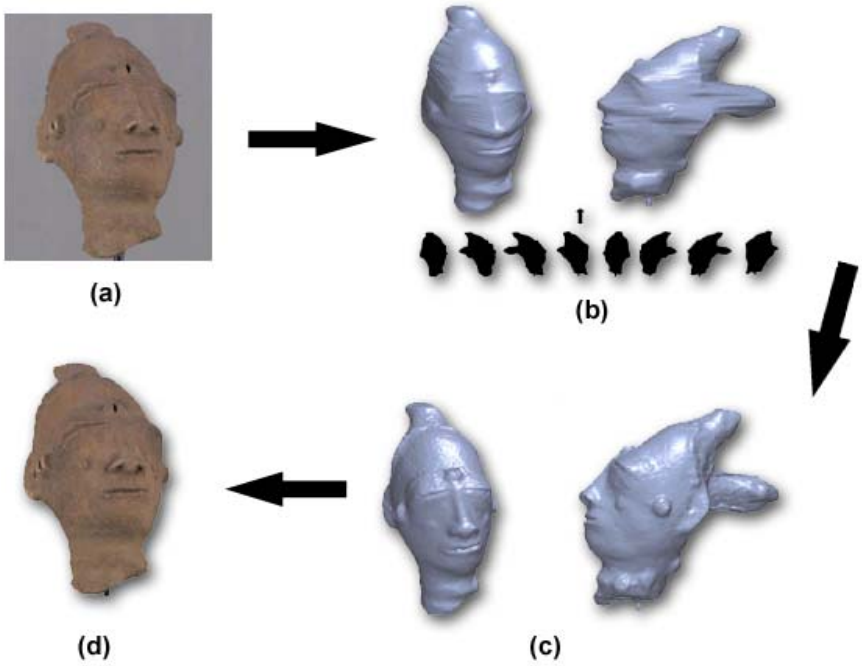


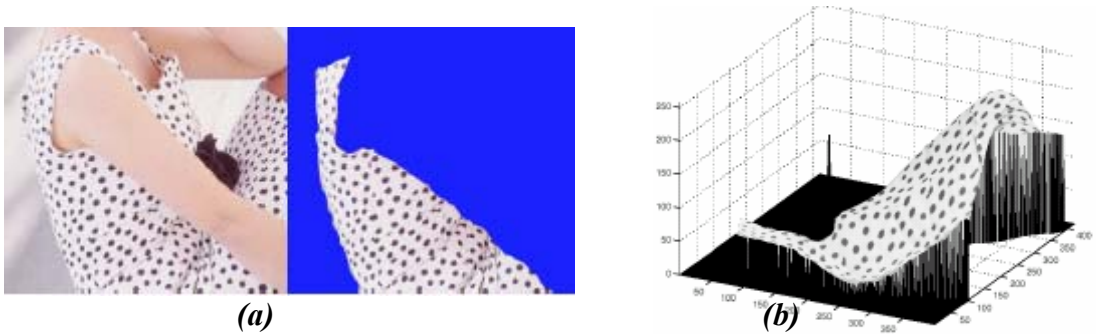
***Figure II.42** : (a) Une vue réelle de l'objet, (b) enveloppe visuelle reconstruite à partir de ses silhouettes, (c) modèle raffiné par une méthode de stéréovision, (d) modèle 3D reconstruit de l'objet après placage de texture [51].*

5.1.2.4 Forme à partir de texture (shape from texture)

On désigne par "texture" une caractéristique matérielle d'une surface constituée d'un motif plus ou moins régulier (grain d'un bois, trame d'un tissu, rugosité d'une pierre, poli d'un métal, etc...). La texture joue un rôle important, par exemple, en segmentation, où elle peut permettre de séparer certains objets du fond. Pour ce qui est de la reconstruction 3D, les variations de texture sur l'image peuvent donner une bonne indication du relief [24], mais seulement dans le cas où cette texture est homogène.

Supposons qu'une surface S soit uniformément texturée : S est couverte d'éléments de texture ou texel dont la taille, la forme et la distribution spatiale sont stationnaires (même densité de probabilité dans le voisinage d'un point quelconque P de S) et isotropique (toute section, dans toute direction, passant par P à les mêmes densités de probabilité en taille et distribution spatiale).
Un motif périodique est un cas particulier de texture uniforme. Dans une image de S, dans le voisinage d'un point P où S est perpendiculaire à la direction de vue, l'image de la texture apparaîtra uniforme. Par contre, si S est inclinée par rapport à la direction de vue en P, l'image de la texture ne sera pas localement isotropique (les tailles et espacements seront raccourcis dans la direction d'inclinaison). De plus, si la profondeur augmente, les éléments de texture deviennent petits. Donc, une variation de profondeur est détectée par une non stationnarité de la texture. Si la taille (ou espacement) moyenne des éléments de texture peut être calculée avec précision dans une direction donnée, on a alors accès à l'amplitude de l'inclinaison par l'amplitude de la variation de la moyenne en fonction de la direction. La direction d'inclinaison est associée à la direction dans laquelle la variation est minimale. Dans une image réelle, il est nécessaire de procéder à une segmentation des éléments de texture pour

pouvoir avoir accès à leur taille et espacements. Les imprécisions liées à cette étape sont souvent génératrices d'erreurs.
On peut bien sur employer d'autres effets d'anisotropie produits par une inclinaison. Par exemple, si la densité de probabilité des orientations des contours des éléments de texture est uniforme, une inclinaison va faire apparaître un pic dont la détection peut permettre la détermination des paramètres de l'inclinaison.
Si la surface ne présente pas texture régulière, il est possible de créer celle-ci grâce à une illumination structurée. Par exemple, si l'on illumine une scène à travers une forme composée de barres noires et blanches, l'inclinaison de la surface dans une direction non parallèle aux barres provoque une réduction des espacements entre les barres. Grâce à une grille, on peut déterminer l'amplitude et la direction d'inclinaison.



(a) ***(b)***

Figure II.43 *: (a) Texture extraite d'une robe, (b) surface reconstruite à partir de la texture*

5.1.3 Morphing/Interpolation

Les méthodes décrites dans cette section permettent de calculer des images intermédiaires entre deux points de vue de référence de la scène.

5.1.3.1 Le morphing

Le *morphing* est le plus simple de ces techniques [52]. Il permet de créer des images intermédiaires entre deux images par interpolation des couleurs et des formes. Son principe est simple, il suffit de disposer de deux vues de la scène, une vue initiale et une vue finale, sur lesquelles des points de contrôle en correspondance sont établies entre les deux images (segments de droite ou des sommets de maillage), puis ces éléments de contrôle sont interpolés, généralement linéairement, entre les deux points de vue. Les autres points de la nouvelle image sont interpolés de façon plus complexe (bilinéaire, splines) entre les éléments de contrôle les plus proches.

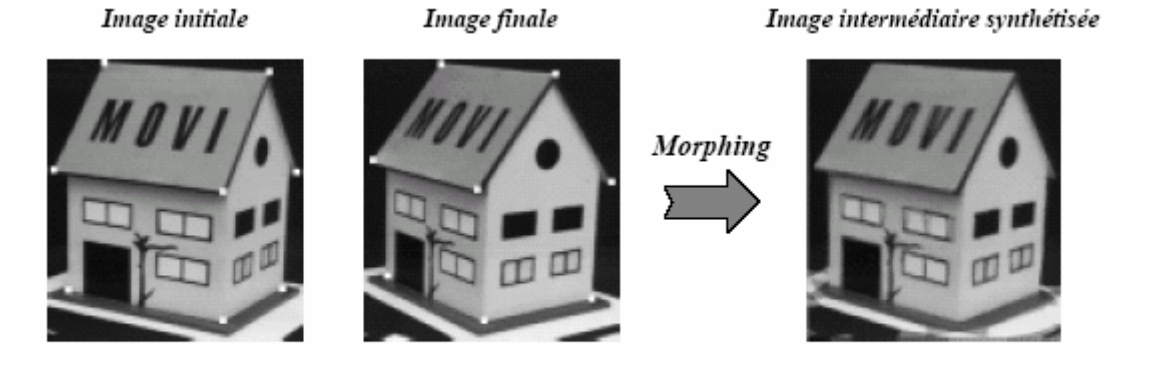


Figure II.44 *: Le morphing : points de contrôles en blanc sur l'image initiale et l'image finale*

Si les deux images représentent deux points de vue d'une scène, cette technique ne fonctionne que si les caméras de référence sont parallèles, sinon l'interpolation est peu réaliste.

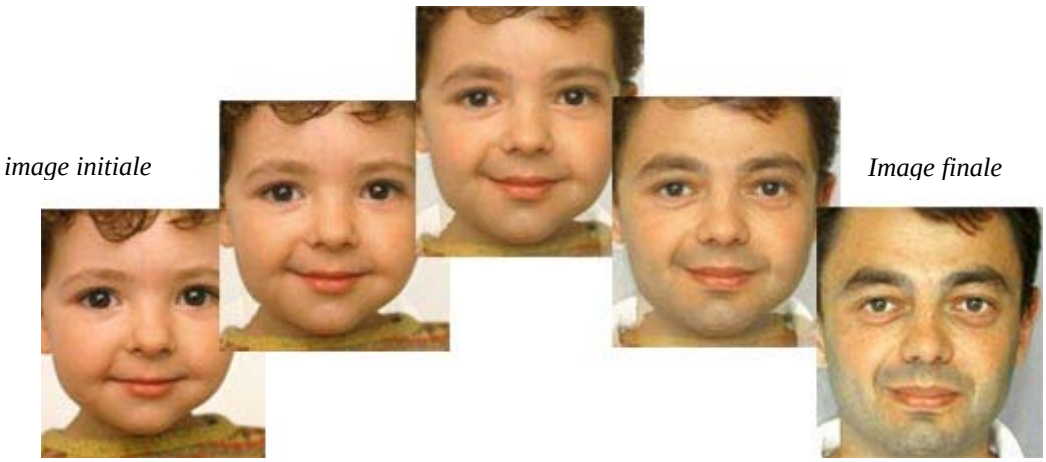


Figure II.45 *: résultat donné par « morphman » un logiciel du marché*

5.1.3.2 Interpolation de point de vue (View Interpolation)

Chen et Williams [19] ont présenté une méthode d'interpolation de point de vue (*View Interpolation)* basée sur le principe du transfert linéaire. Leur méthode synthétise de nouvelles vues à partir d'images de référence pour lesquelles ils établissent des correspondances permettant de déplacer les points d'une image à l'autre. Chaque pixel se déplace donc dans l'image de son emplacement d'origine à son emplacement d'arrivée en suivant une ligne droite. Ainsi, ils génèrent toutes les images intermédiaires entre deux images. D'une façon générale, les pixels en mouvement ne sont pas exactement au bon endroit (par rapport à la scène réelle), mais cette interpolation linéaire donne de bons résultats tant que les deux points de vue de référence ne sont pas trop éloignés. Par ailleurs, si les plans de projections des caméras des points de vue de référence et du point de vue à générer sont parallèles, leur interpolation est exacte. Etant donné que plusieurs pixels peuvent être déplacés au même endroit, un tampon de profondeur doit être utilisé. Les trous qui inévitablement apparaîssent sont remplis en interpolant les couleurs des pixels adjacents différents du fond.

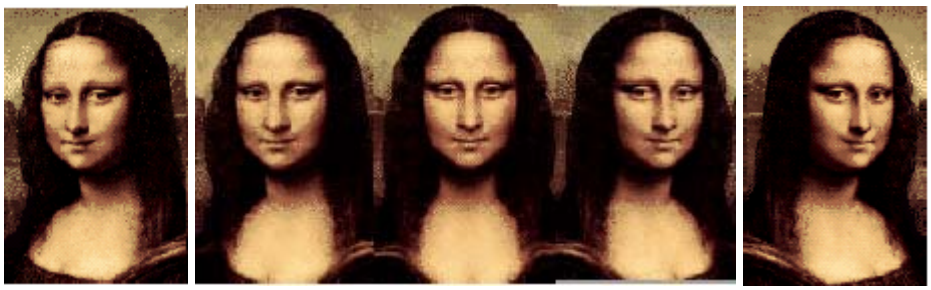


Image de dépare *Image d'arrivée*

Figure II.46 : *View morphing*

5.1.3.3 Déformation de point de vue (View Morphing)

Seitz et Dyer [27] ont introduit la *déformation de point de vue* (*View Morphing)* qui peut être vues comme une extension des travaux de Chen et Williams. Leur méthode consiste à prédéformer les images de référence avant de les interpoler linéairement. La prédéformation permet d'obtenir des images parallèles pour lesquelles l'interpolation par transfert linéaire est exacte. Une fois l'interpolation effectuée et l'image intermédiaire aux points de vue calculée, l'image est redéformée pour donner l'image finale (*post-warping*). Voir la *figure II.47*

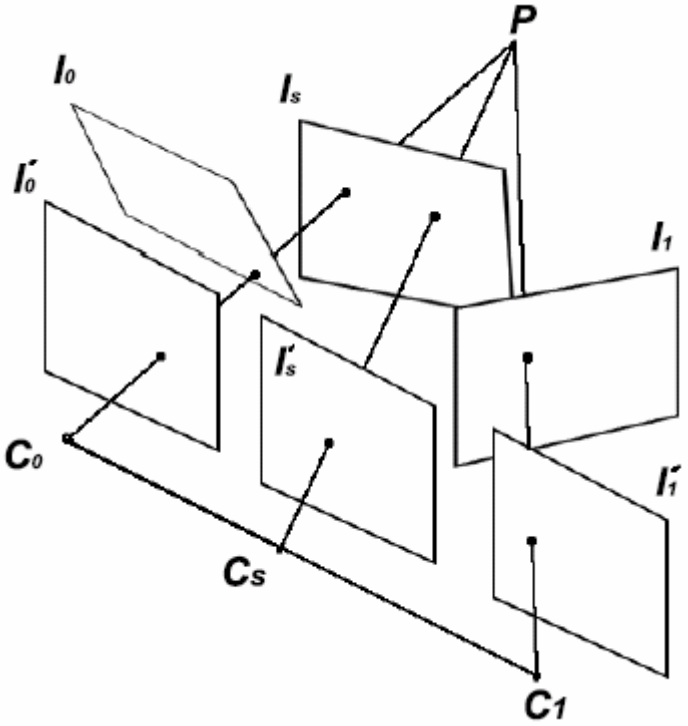


Figure II.47 :*View Morphing de l'image* I_0 *vers l'image* I_1*. L'image intermédiaire* I_s *est créée par post-warping de l'image* I_s*' elle même obtenue par transfert linéaire entre les images* I_0*' et* I_1*' .*

5.1.4 Interpolation de rayons lumineux et fonction plénoptique

Le principe de l'interpolation de rayons lumineux est de considérer l'espace à cinq dimensions de tous les rayons lumineux possibles traversant une scène. Cet espace est dénommé *lightfield* (champ lumineux). Un rayon ***y*** est défini par sa position (3 coordonnées)

et sa direction (2 angles). Une image est alors un plan 2D plongé dans cet espace et "interceptant" ces rayons lumineux. Les méthodes d'interpolation de rayons visent à acquérir ce champ lumineux (à partir d'images réelles), à le stocker, et à l'utiliser afin de produire des images [11,12,21,23]. La géométrie de la scène n'intervient alors pas dans les calculs, seul le champ lumineux est reconstruit. C'est donc idéalement l'approche parfaite pour la synthèse d'images. Bien sûr, il est impossible d'acquérir intégralement ce champ lumineux, cela demanderait un espace de stockage (et un temps d'acquisition) infini. Ces méthodes se basent sur une fonction définie par Adelson et Bergen [21] appelée *fonction plénoptique*.

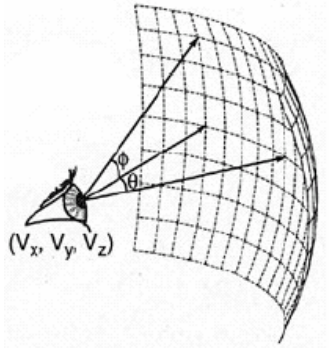


Figure II.48 : *La fonction plénoptique décrit toutes les informations de l'image visibles à partir d'un point de vue particulier*

5.1.4.1 Fonction plénoptique

La *fonction plénoptique* [21] est une fonction 7D qui modélise un environnement 3D dynamique par l'enregistrement des rayons lumineux (énergie radiante) en chaque point (V_x,V_y,V_z) de l'espace dans toutes les directions (θ,φ), à tout instant t et pour toute longueur d'onde λ, soit :

$$(V_x, V_y, V_z, \theta, \varphi, \lambda, t)$$

En ne tenant pas compte du temps (scène statique), ni de la longueur d'onde (éclairement constant), la fonction se réduit à 5 dimensions [23] :

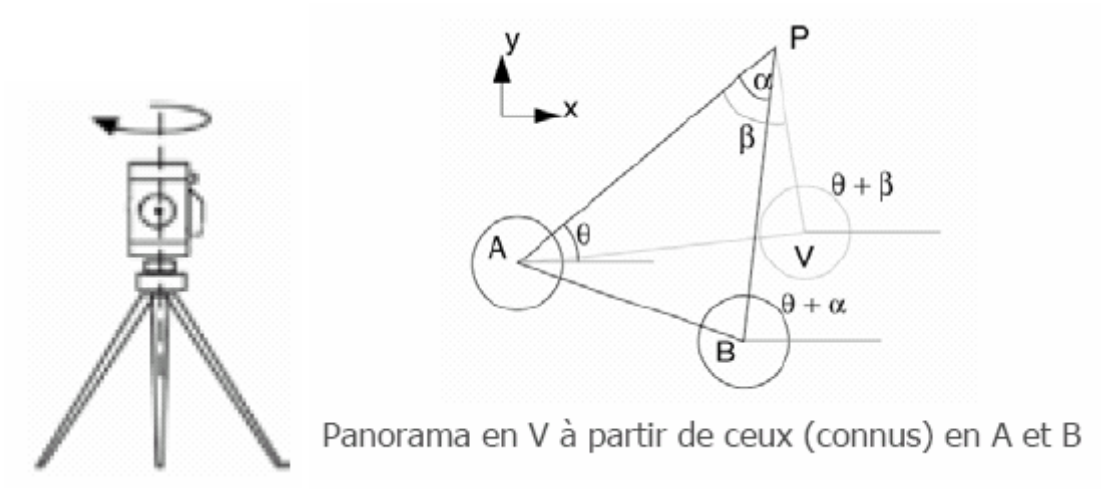
$$(V_x, V_y, V_z, \theta, \varphi)$$

Malgré ces simplifications, la quantité de données nécessaire pour décrire une telle fonction est tellement grande que les méthodes que nous allons décrire restreignent les positions (ou les orientations de vue) pour lesquelles le champ lumineux pourra être reconstruit. Ainsi, si la scène peut être contenue dans une boîte englobante, les *Light Fields* [12] et *Lumigraphes* [11] permettent de construire le flot lumineux *autour* de la scène. Si l'espace des points de vue possibles est contenu dans un cercle 2D, les *Mosaïques Concentriques* [53] peuvent simuler la fonction plénoptique. Si le point de vue est fixé, la fonction plénoptique peut alors être représentée par un *panorama 2D*.

5.1.4.2 Modélisation plénoptique (Plenoptic Modeling)

Dans [23], McMillan et Bishop choisissent d'ignorer le temps t et la langueur d'onde λ et d'introduire le plenoptic modeling qui est une fonction à 5 dimensions :

$$(V_x, V_y, V_z, \theta, \varphi)$$



(a) *(b)*

***Figure II.49** : (a) caméra panoramique, (b) principe de la modélisation plénoptique*

Ils enregistrent une scène statique en plaçant des caméras dans l'espace scène 3D, chacun sur un trépied permettant un mouvement vertical et de rotation. À chaque position, une image projetée cylindriquement est composée à partir des images capturées pendant une rotation panoramique. Cette projection cylindrique est obtenue par la reconstruction de la fonction plénoptique qui nécessite d'estimer le flot lumineux à chaque position d'une caméra. Pour cela, à partir de deux (ou plus) images de référence panoramiques projetées cylindriquement, le champ lumineux est calculé par disparité stéréoscopique. La projection cylindrique à un nouveau point de vue peut être obtenue par un placage de type cylindre → cylindre, puis peut être reprojetée sur un plan permettant d'obtenir une image.

(a)

(b)

***Figure II.50 :** (a) vue panoramique de la scène, (b) nouvelles vues synthétisées*

5.1.4.3 Light Fields et Lumigraphes

Light Fields [12] et Lumigraphes [11] sont tous deux basés sur une paramétrisation à 4D de la fonction plénoptique, le temps ***t*** et la langueur d'onde **λ** sont ignorés et le flot lumineux est supposé constant le long du rayon, quand la scène peut être contenue dans une boîte englobante. Au lieu d'utiliser **(V_x, V_y, V_z, θ, φ)** pour caractériser un rayon lumineux, ces deux méthodes utilisent une *tranche de lumière* (*light-slab*) à 4D :***(u,v,s t)*** *(Figure II.51)* . Ceci est possible en supposant que le flot lumineux ne change pas le long du rayon, ce qui est le cas pour la lumière entourant la scène (et si l'environnement n'est pas un milieu participant).

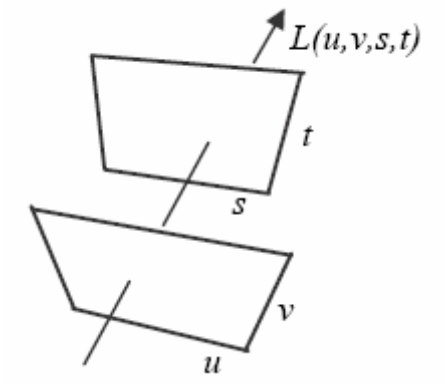


Figure II.51 *: le light-slab*

Les light-slabs son enregistrés par leurs intersections L(u,v,s,t) avec deux quadrilatères parallèles . Le premier est le plan caméra indexé par les coordonnées (u,v) et le deuxième est le plan focal indexé par les coordonnées (s,t).

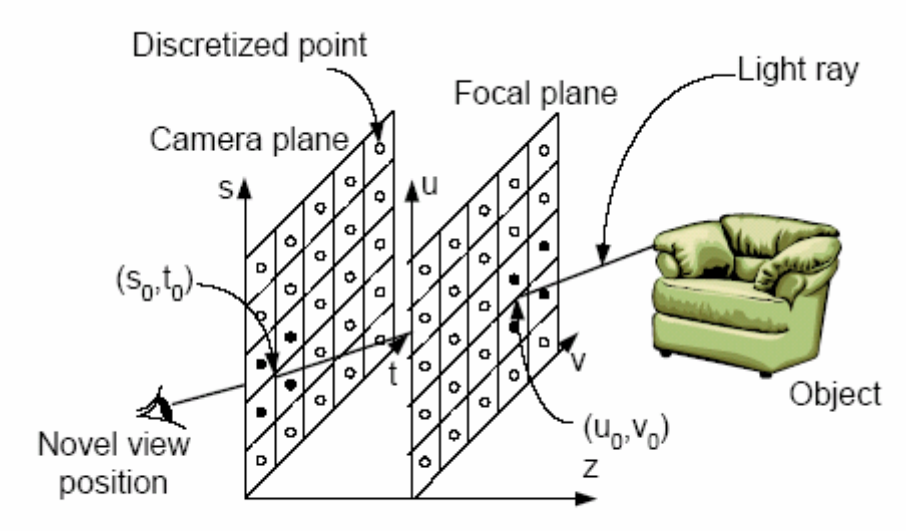
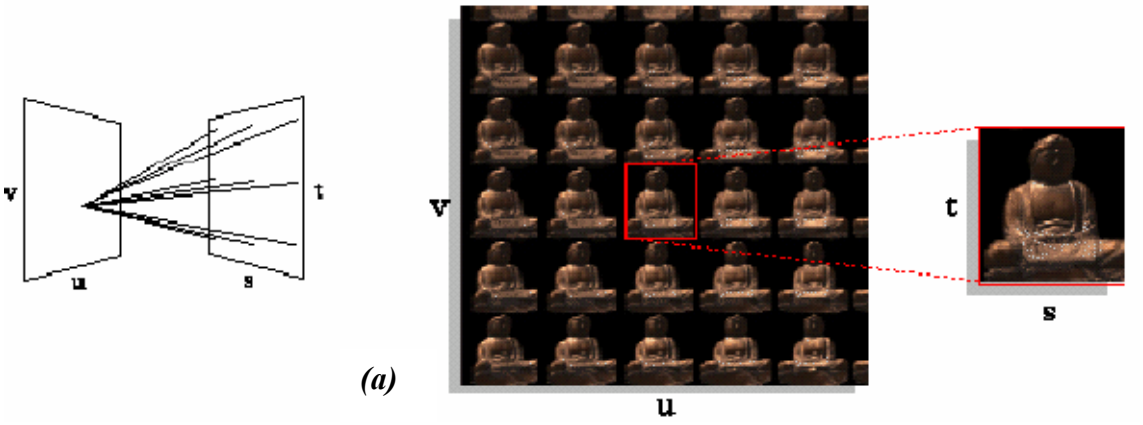


Figure II.52 *: Paramétrage de light field*

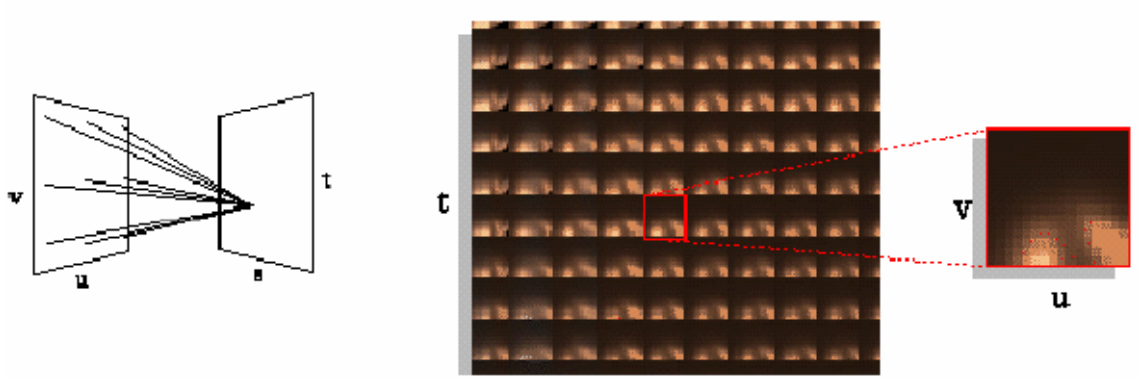
Levoy et Hanrahan [12] ont donc introduit une méthode de rendu de champ lumineux (*Light Field Rendering*). Ils proposent de paramétrer les rayons par deux plans dont le système de coordonnées est ***(u,v)*** pour le premier et ***(s,t)*** pour le second (*figure II.51*). Le champ lumineux est alors généré en assemblant une collection d'images. Pour une scène réelle, un

tableau 2D d'images est créé en plaçant le centre de projection de la caméra à chaque position ***(u,v)*** du premier plan. Dans leurs travaux, ils construisirent un dispositif photographique spécial contrôlé par ordinateur pour acquérir le champ lumineux d'une scène réelle. La caméra est déplacée aux positions d'une grille 3D dans l'espace et des photographies de la scène sont prises. Chacune de ces images est ensuite projetée (par placage de texture) sur le second plan ***(s,t)*** (*figure II.52*).

La structure peut être compressée avec un système préservant les données : une quantification vectorielle à taux fixé suivi d'un encodage entropique (gzip). Finalement, ils obtiennent un taux de compression de l'ordre de 120:1, en moyenne. La décompression est faite en appliquant ces deux étapes dans l'ordre inverse. Bien que l'algorithme consiste essentiellement en des interpolations linéaires de couleurs, donc très rapide, la décompression au vol est très coûteuse : environ 25% du temps CPU. Ceci empêche l'utilisation de plusieurs modèles simultanés dans une même scène, limitant la méthode. De plus, aucun travail n'a jusqu'ici permit de représenter un environnement complet à l'aide de *Light Fields*, du fait de sa complexité de mise en oeuvre.



(a) Tableau (u,v) d'images



(b) Tableau (s,t) d'images

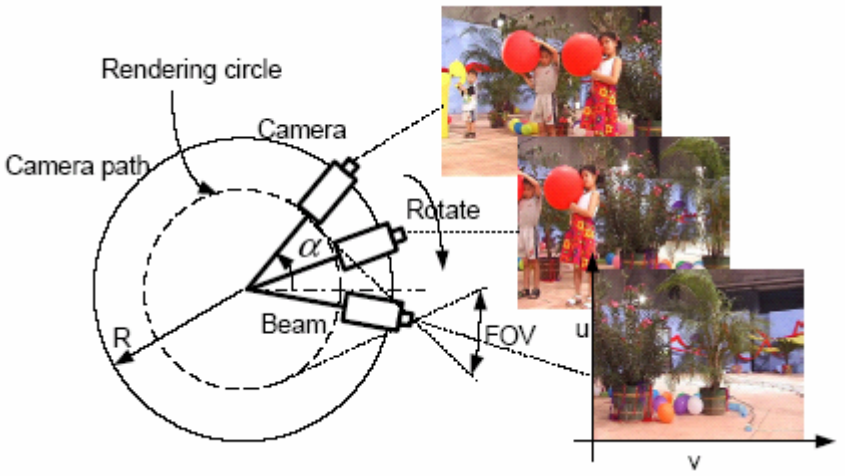
(c)

Figure II.53 : *(a)(b) Acquisition des échantillons d'images de référence pour la construction d'un light-slab ,(c) modèles 3D reconstruits*

Gortler *et al.* [11] ont développé leur système appelé *Lumigraphes* en même temps que les *Light Fields*. Dans leurs travaux, la fonction plénoptique est reconstruite sur la surface d'un cube englobant la scène, comme les *Light Fields*. La technique de paramétrisation d'un rayon de lumière est la même que pour les Light Fields. Dans leur méthode, les échantillons de la fonction plénoptique peuvent être acquis à l'aide d'une caméra tenue à la main. Le calibrage de la caméra, le temps de pose et une version approchée de la géométrie de la scène peuvent être obtenus à l'aide de techniques de Vision par Ordinateur. Etant donné qu'ils utilisent une caméra arbitrairement positionnée, les positions d'échantillonnage ne peuvent être ni spécifiées ni contrôlées ce qui ne garantit pas que ces échantillons de la fonction plénoptique soient régulièrement répartis. Ils proposent une technique de rééchantillonnage basée sur le modèle 3D approché de l'objet.

5.1.4.4 Mosaïques concentriques (Concentric mosaic)

Plus on contraint les positions possibles de la caméra, plus la fonction plénoptique devient facile à reconstruire. Shum *et al.* [53] suppose que les caméras et l'observateur (les points de vues) soient sur le même plan, ce qui réduit la fonction plénoptique en 3D et ils ont proposé les *mosaïques concentriques* qui permettent de reconstruire la fonction plénoptique pour un point de vue positionné sur un cercle. Leur méthode compose entre elles différentes photographies prises à différentes positions sur ce cercle. Les rayons sont caractérisés par trois paramètres : la position du pixel ***(u,v)*** et l'angle de rotation **α : (α ,u,v)**
Les nouvelles vues sont créées en combinant rapidement les rayons appropriés lors du rendu. Aucun effet de parallaxe n'est capturé par cette technique car les positions de la caméra sont contenues dans un plan et qu'une seule image est prise à chaque point de vue. Lors du rendu néanmoins, des distorsions verticales peuvent apparaître dans l'image obtenue. Différentes corrections sont proposées pour résoudre ce problème. Comparés aux *Light Fields* et *Lumigraphes*, les mosaïques concentriques sont bien moins gourmandes en espace mémoire.



(a)

(b)

***Figure II.54** : (a)Prise de vues de la mosaïque concentrique,(b) deux vues rendues [54]*

5.1.4.5 Panoramas (image mosaicing)

Quand le point de vue de capture est réduit à un point, la structure de capture de la fonction plénoptique est dénommée *Panorama.*

Si nous observons une scène statique avec une caméra en rotation pure autour de son centre optique, nous obtenons des images qui sont deux à deux en correspondance homographique et qui ayant une zone de recouvrement. L'assemblage de ces images forme une vue qui couvre la totalité de la surface visible d'une scène.

Pour reconstituer un panorama à partir d'un nombre d'images, deux cas sont possibles : si la transformation projective est connue alors les images peuvent être collées entre elles facilement par cette transformation. Sinon, une technique courante consiste à mettre, au moins, 4 points en correspondance sur la zone de recouvrement des deux images de chaque paire, de façon à pouvoir calculer la transformation projective associée et de l'appliquer pour coller les images entre elles.

Ainsi, Chen [18] a construit des images panoramiques cylindriques à 360^{o}; (à champ de vision vertical limité). Les images panoramiques *(figure II.55)* peuvent être créées, à l'aide d'appareils photos panoramiques spéciaux ou en combinant des photographies entre elles. A chaque point de capture, une rotation et un zoom peuvent être effectués en utilisant une déformation de type cylindre plan. La navigation en 3D se fait de façon discrète, d'un point à l'autre.

***Figure II.55** : images panoramiques*

5.1.5 Modélisation volumétrique de scène (volumetric scene medeling)

Les méthodes appelées volumiques ou *volumetric scene modelling* [31, 56] construisent le volume des objets en prenant des décisions sur la cohérence des éléments du volume *(voxels*), par exemple un voxel est dit cohérent si ses projections sur les images présentent des ressemblances de couleur. Ces méthodes permettent d'utiliser plusieurs images prises depuis des points de vues espacés, ce qui pose parfois problème à des méthodes fondées sur l'appariement entre images. Toutefois les méthodes volumiques demandent un calibrage très précis ce qui réduit leur champ d'application à des environnements d'expérimentation.

Certaines tentatives ont été faites pour réduire cette contrainte, comme le *space carving* décrit plus bas.

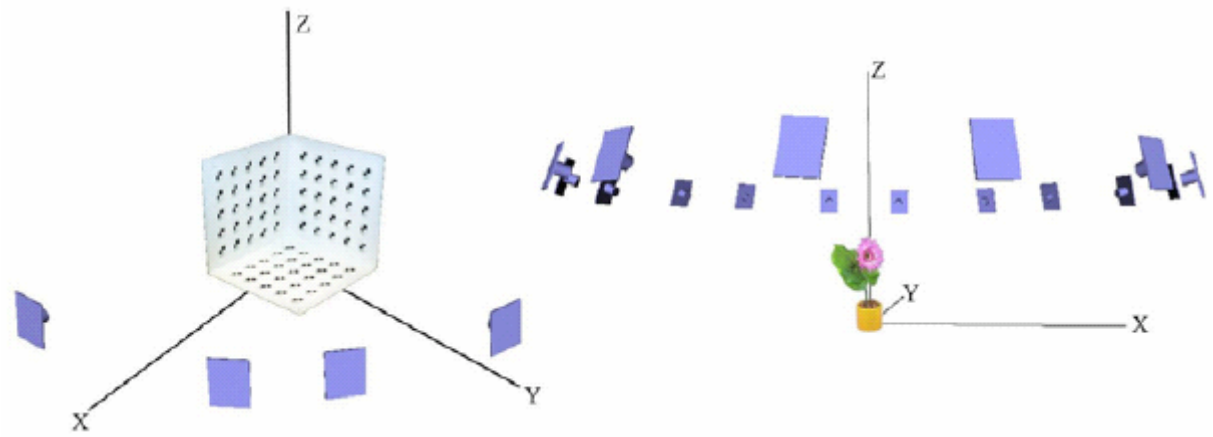
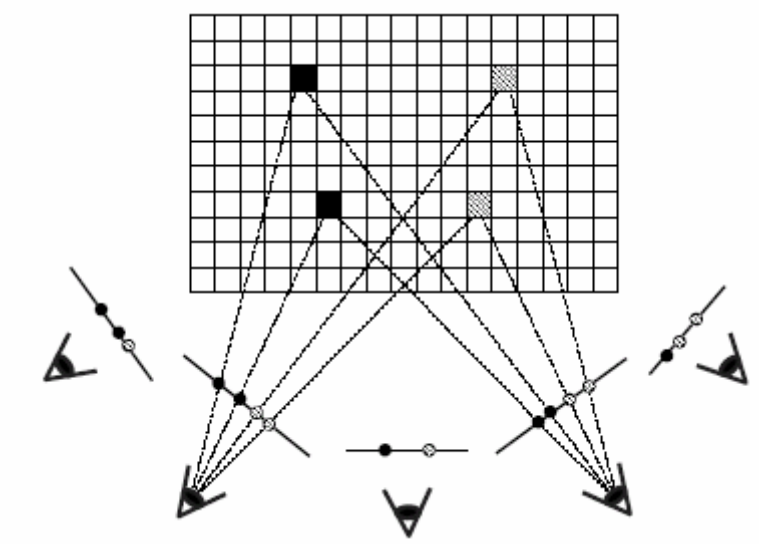


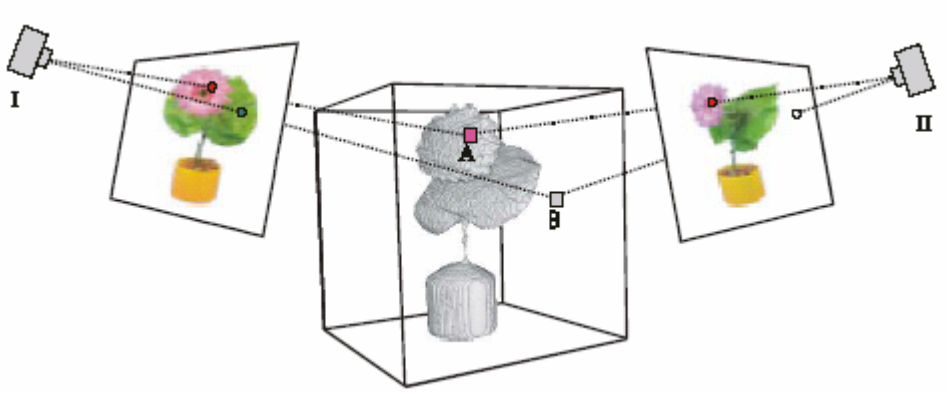
***Figure II.56** : Calibrage et prise de vues [57]*

5.1.5.1 Coloration de voxel (Voxel coloring)

Les algorithmes de coloration de voxels [56] utilisent la photocohérence *(photo-consistency en anglais)* de tous les voxels dans chaque image. Un voxel est dit cohérent si toutes ses projections dans les images où il est vu, présentent une ressemblance sur les couleurs ou sur les intensités. Un voxel est considéré appartenant à l'objet si la mesure de sa cohérence est supérieure à un certain seuil.



(a)



(b)

***Figure II.57** : (a) Voxel coloring : étant donnée un ensemble de vues (images) et une grille de voxels l'objectif est d'attribuer des valeurs de couleur aux voxels de telle sorte que ces derniers soient cohérents avec toutes les images. (b) exemple : Le voxel A est cohérent mais B ne l'est pas.*

Les algorithmes de coloration de voxels commencent par un volume de reconstruction initiale où tous les voxels sont opaques. La cohérence de chaque voxel est calculée, s'il n'appartient pas à un objet il devient transparent. Cette approche peut être assimilée intuitivement à de la sculpture. Afin de pouvoir traiter les occlusions, un ordonnancement des voxels doit être établi au préalable pour assurer que quand un voxel est traité tous ceux qui l'occultent ont été traités avant lui. Si cet ordonnancement existe, la reconstruction est assurée en traitant dans l'ordre les voxels. Toutefois l'utilisation de cet ordonnancement restreint les positions des caméras et réduit le champ d'application de ces algorithmes.

(a) (b)

***Figure II.58** : Reconstruction d'une fleur par voxel coloring utilisant 16 images : (a) vue réelle de la fleur, (b) trois vues du modèle 3D reconstruit [57].*

5.1.5.2 Space carving

Kutulakos et Seitz [58] ont proposé un autre algorithme, le Space carving, qui lève la restriction sur la position des caméras. Pour cela un plan traverse la scène dans les trois directions de l'espace, et à chaque itération, seuls les voxels appartenant au plan sont testés sur les images prises par des points de vue faisant face au plan. Kutulakos et Seitz ont montré que cette technique garantit une reconstruction correcte, en permettant une disposition quelconque des caméras. Cet algorithme a été étendu par la suite pour permettre d'utiliser une calibration approximative. La notion de photocohérence est relâchée en calculant la

ressemblance sur des voisinages de taille variable : ceci permet d'avoir des résultats de reconstruction plus ou moins précis selon la taille des voisinages.

5.2 Techniques basées images / basées géométrie

Ces techniques sont basées sur les images de profondeur. Une image de profondeur (*depth image*) est une image contenant couleurs et profondeurs en chaque pixel.
Etant donnée une image (ou plus) obtenue à partir d'une caméra placée dans une scène et sa carte de profondeur (calculer par une méthode active ou passive), la problématique est comment utiliser uniquement ces deux informations pour reconstruire un nouveau point de vue (une nouvelle image) pour une position de caméra différente ? Ici, nous connaissons la profondeur des pixels, c'est-à-dire que nous disposons d'une représentation surfacique discrète de la scène et nous connaissons aussi les caractéristiques de la caméra (position, direction, matrice de transformation).

5.2.1 Transfert d'images classique

Ces techniques sont dites de "prédiction algébrique de pixels" et ont été spécifiquement développées pour synthétiser de nouvelles images d'une scène uniquement à partir de photographies [5]. Pour cela, la "géométrie" de l'image est décrite sous forme de cartes de correspondances liant plusieurs vues de référence de cette scène. Une carte de correspondance est un tableau 2D appariant deux pixels issus de deux images de référence de la scène. Les pixels appariés correspondent approximativement au même point 3D dans l'espace de la scène. Grâce à cet appariement et à la connaissance des caractéristiques des caméras, ont peut obtenir la profondeur des points. Cet appariement peut être créé de diverses façons, comme par exemple en établissant des correspondances entre deux images d'un *référentiel* positionné dans la scène, et extrait par des techniques de traitement d'images. La géométrie est donc dite *implicite* car il faut l'extraire des images.

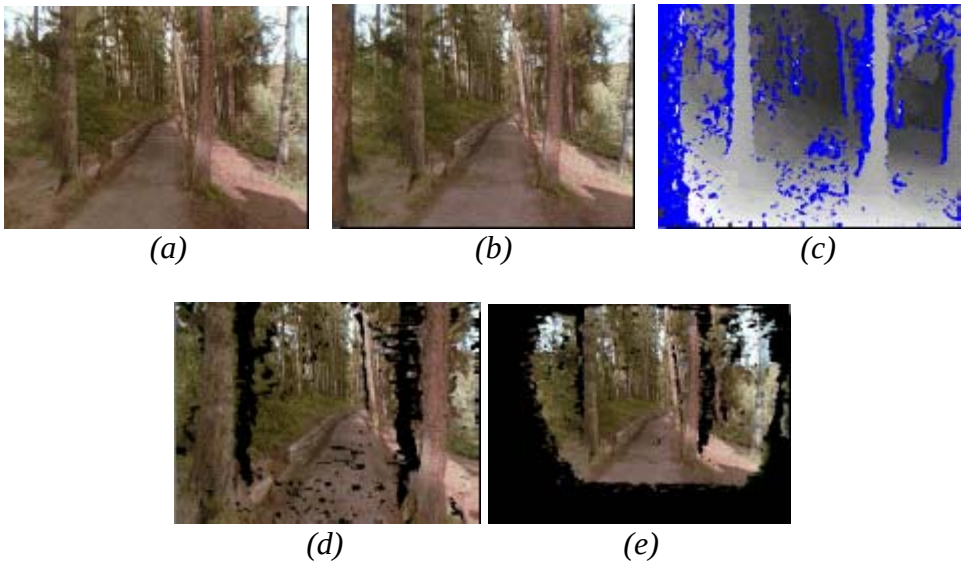


***Figure II.59**: Transfert d'images : (a)(b) images d'une paire stéréo,(c) carte de profondeur, (d)(e) deux nouvelles vues synthétisées.*

Ces méthodes sont également appelées "techniques de transfert" de pixels car elles consistent à "transférer" dans la nouvelle image chaque pixel de l'image de référence. Ces techniques

visent à prédire la position dans l'image source d'un pixel donné de l'image à construire. En fonction du type de caméra, de la position des points de vue (alignés ou non, par exemple), les équations algébriques de transfert sont plus ou moins complexes.

5.2.2 Déformation 3D d'images (3D image warping)

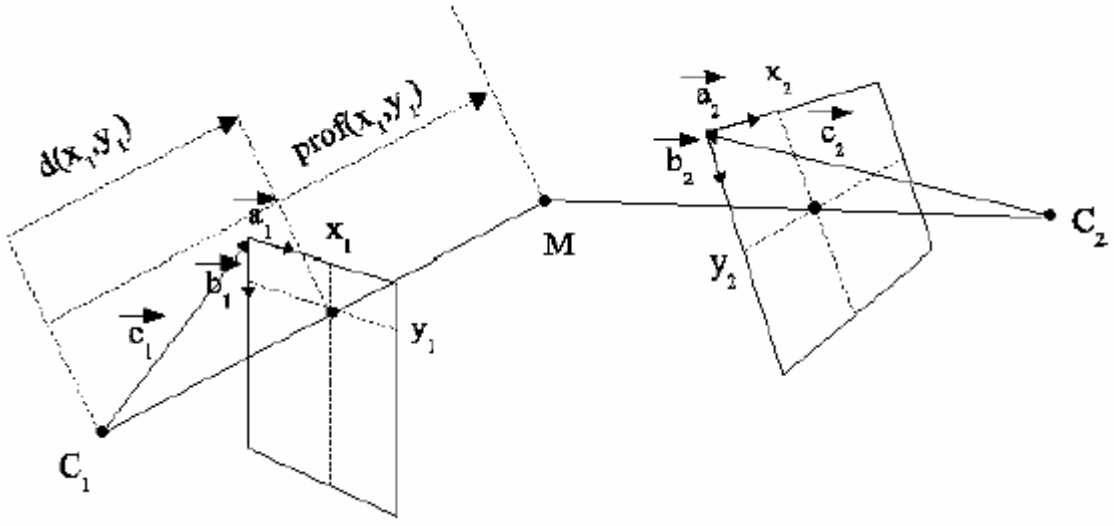


***Figure II.60** : 3D image warping*

Dans [55], McMillan formalise les équations de reprojection. Il dérive de la profondeur classique la profondeur projective *(projective depth)* ou encore disparité généralisée *(generalized disparity)*, pour laquelle l'expression des équations de reprojection est plus simple. La profondeur projective d'un pixel correspond au rapport de la distance focale de la caméra par la profondeur du point *(figure II.60)*

La disparité généralisé (δ)= rapport de la profondeur du pixel avec la distance pixel-centre de projection : $\delta(x_1, y_1) = prof(x_1, y_1)/d(x_1, y_1)$
Ainsi, si l'on souhaite calculer le point P2(x2,y2) de I2 à partir du point P1(x1,y1) , de disparité généralisée $\delta(x_1, y_1)$ on a :

$$x_2 = \frac{\overrightarrow{a_1}.(\overrightarrow{b_2} \times \overrightarrow{c_2})x_1 + \overrightarrow{b_1}.(\overrightarrow{b_2} \times \overrightarrow{c_2})y_1 + \overrightarrow{c_1}.(\overrightarrow{b_2} \times \overrightarrow{c_2}) + \overrightarrow{C_2C_1}.(\overrightarrow{b_2} \times \overrightarrow{c_2})\delta(x_1, y_1)}{\overrightarrow{a_1}.(\overrightarrow{a_2} \times \overrightarrow{b_2})x_1 + \overrightarrow{b_1}.(\overrightarrow{a_2} \times \overrightarrow{b_2})y_1 + \overrightarrow{c_1}.(\overrightarrow{a_2} \times \overrightarrow{b_2}) + \overrightarrow{C_2C_1}.(\overrightarrow{a_2} \times \overrightarrow{b_2})\delta(x_1, y_1)}$$

$$y_2 = \frac{\overrightarrow{a_1}.(\overrightarrow{c_2} \times \overrightarrow{a_2})x_1 + \overrightarrow{b_1}.(\overrightarrow{c_2} \times \overrightarrow{a_2})y_1 + \overrightarrow{c_1}.(\overrightarrow{c_2} \times \overrightarrow{a_2}) + \overrightarrow{C_2C_1}.(\overrightarrow{c_2} \times \overrightarrow{a_2})\delta(x_1, y_1)}{\overrightarrow{a_1}.(\overrightarrow{a_2} \times \overrightarrow{b_2})x_1 + \overrightarrow{b_1}.(\overrightarrow{a_2} \times \overrightarrow{b_2})y_1 + \overrightarrow{c_1}.(\overrightarrow{a_2} \times \overrightarrow{b_2}) + \overrightarrow{C_2C_1}.(\overrightarrow{a_2} \times \overrightarrow{b_2})\delta(x_1, y_1)}$$

De plus, McMillan proposa une méthode permettant de gérer correctement la visibilité lors du rendu, sans utiliser de tampon de profondeur, ainsi qu'une technique de reconstruction d'image incrémentale permettant de remplir l'image de destination par balayage de lignes direct. Cette technique est appelée *3D Image Warping*, et a été utilisée dans de très nombreux travaux [1,54].
Parmi les utilisations du 3D Warping, les *sprites with depth* de Shade [14], introduits en même temps que les LDI (section suivante). La profondeur du sprite est utilisée pour le

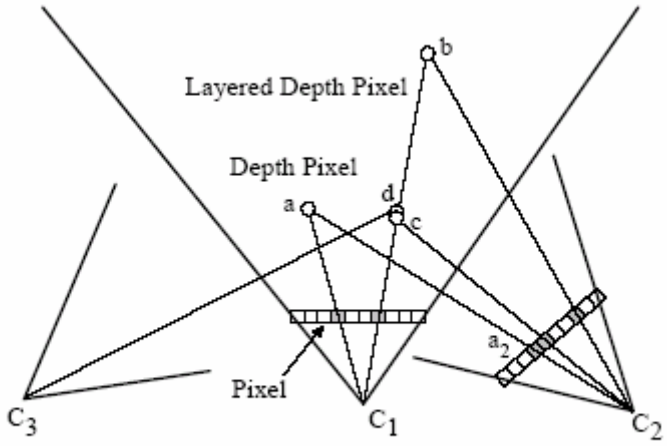
déformer légèrement en fonction du point de vue, augmentant ainsi leur temps de vie. Bien sûr, la déformation ne peut dépasser un certain seuil à cause des problèmes de dé-occlusion (trous) qui se produisent lors de la déformation.

(a) (b) (c)

FigureII.61 : *(a) image avec profondeurs : vue d'un modèle 3D d'une cathédrale, (b) (c) deux nouvelles vues obtenues par 3D image warping*

5.2.3 Images à plans de profondeurs (LDI : Layered Depth Images)

A cause des problèmes de dé-occlusion, une seule image ne contient pas assez d'information pour générer un nouveau point de vue complet. Certaines parties de la scène peuvent être cachées dans l'image de référence et être visibles à partir d'un autre point de vue. Les LDI ("*Layered Depth Images*", Images à Plans de Profondeurs) [14] ont été spécifiquement conçues pour résoudre ce problème.



FigureII.62 : *Layered Depth Images*

Une LDI est une vue de la scène à partir du point de vue d'une seule caméra, mais avec plusieurs pixels stockés le long de chaque rayon optique, c'est à dire que chaque rayon intersecte plusieurs couches de la scène. Le principe des couches (Layers) est un concept très puissant en modélisation et en rendu. Chang *et al*. [17] ont introduit les *LDI-tree* (LDI hiérarchiques) pour résoudre les problèmes d'échantillonnage inhérents aux LDI.

Nom	Données par pixel
Image	(R,V,B)
Image de profondeurs	(R,V,B,P)
Image à plans de profondeurs (LDI)	Liste de (R,V,B,P)

***Tableau II.02** : Types d'images et leurs composants : R=Rouge, V=Vert, B=Bleu, P=Profondeur*

Shade *et al*. ont donc proposé les LDI pour résoudre les problèmes d'occlusion en stockant plusieurs points de la scène par pixel, tout en gardant la simplicité de la déformation d'image de McMillan [55].
Une LDI peut être créée de plusieurs façons. Si on dispose d'un ensemble d'images de profondeurs de la scène, on peut les rétro-projeter sur la caméra de référence. On peut également effectuer un lancer de rayons et stocker les multiples intersections entre la scène et chaque rayon. Une fois la LDI construite, on dispose d'une liste de surfels par pixels. Le rendu s'effectue à l'aide de l'algorithme de déformation 3D d'images de McMillan, adapté de façon à ce que les trous soient comblés avec les informations stockées dans les couches de profondeurs des pixels. Les pixels étant dessinés d'arrière en avant, il est possible d'utiliser l'écrasement de point (*splatting*) directement afin d'améliorer la qualité de l'image en remplissant les trous qui peuvent néanmoins apparaître si la LDI n'est pas assez bien échantillonnée. Ceci permet un rééchantillonnage rapide des pixels de l'image. Dans leurs travaux, ils utilisent un *splat* de taille variable dépendant de la profondeur et de la normale à la surface. Une table d'indexation quantifiée peut être utilisée pour accélérer ce calcul et donc l'algorithme.

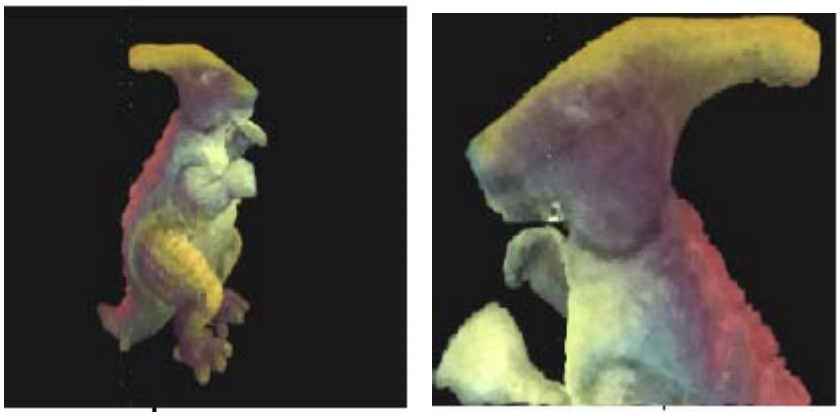


***Figure II.63**: Un modèle de dinosaure reconstruit par LDI à partir de 21 images*

Cet algorithme comporte quand même quelques problèmes. Premièrement, les pixels subissent deux rééchantillonnages avant d'arriver à l'écran, ce qui dégrade la qualité de l'image. Deuxièmement, de l'information peut être perdue si une surface est mieux échantillonnée dans une image de base qu'à partir du point de vue de la LDI. Les performances sont moyennes : 8 à 10 FPS pour une résolution de 300x300 pixels sur un Pentium 300Mhz.

5.3 Techniques hybrides

5.3.1 Placage de texture

Le placage de texture (*texture mapping*) est la plus simple et la plus ancienne des techniques à base d'images. Le placage de texture permet d'ajouter du réalisme à une scène pour une faible augmentation de la complexité du programme et du temps de rendu. Dans sa forme la plus simple, le placage de texture dépose et déforme une image (une texture) sur la surface d'un objet de la scène. L'algorithme se décompose en deux phases : la transformation de l'image de l'espace texture à l'espace écran et le filtrage pour traiter les problèmes d'aliassage.
Bien que la technique de placage de texture a une puissance éprouvée et actuellement intégrée dans toutes les cartes graphiques. Elle présente plusieurs limitations. En effet, le placage de texture ajoute des détails à une surface en modifiant la couleur, mais cette couleur reste la même quelque soient les conditions d'éclairement, ce qui révèle que le relief représenté par l'image n'est qu'une illusion. En plus, le placage de texture ne fonctionne bien que pour des surfaces planes ou légèrement courbées et les objets réalistes très détaillés nécessitent beaucoup de textures, qui peuvent de plus nécessiter d'être traitées en plusieurs passes, ralentissant d'autant les performances. Des phénomènes tels que la fumée, le feu ou l'eau sont difficiles à traiter en employant cette méthode.
Plusieurs variantes et extensions de cette technique ont étés proposé pour remédier à ces limitations et augmenter le réalisme des résultas comme par exemple, le placage de bosselures (*BumpMapping*), l'ombrage par carte d'horizon *(Horizon Mapping)* et le Déplacement de Surface (*Displacement Mapping*). Mais l'extension la plus remarquable est celle proposée par M. Oliveira [15] : le placage de texture en relief (*Relief texture mapping*)

5.3.2 Placage de texture en relief

Le placage de texture en relief est une extension de placage de texture conventionnel qui permet la représentation 3D de détails sur les surfaces et génère des effets de parallaxe. Cette technique produit des vues correctes des surfaces par la déformation de texture augmentées par des informations de profondeur en chaque pixel.
L'opération de placage de texture en relief s'opère en deux passes. Durant la première passe de pré-déformation (*pre-warping*), la texture est déformée en fonction de la direction d'observation selon les équations modifiées de McMillan. Cette phase est lente et effectuée par le CPU. La seconde passe est très rapide et ne consiste qu'à plaquer la texture déformée de manière ordinaire. Cette passe est généralement exécutée par la carte graphique.

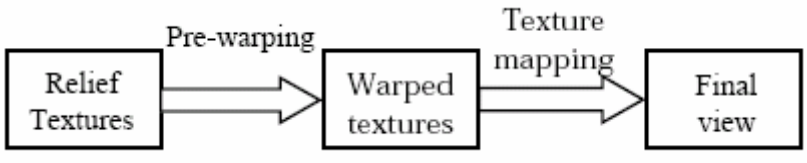


***Figure II.64** : Les étapes du placage de texture en relief*

Il s'avère que les textures en relief offrent des possibilités intéressantes, elles d'augmentent de manière significative le réalisme visuel de la scène rendue sans aucune charge supplémentaire sur le système de rendu. En effet, l'étape de prétraitement remplace les détails géométriques

sur le modèle polygonal de la scène par les détails de la texture. Ces détails de texture ressemblent à ceux produits par la géométrie sur le modèle polygonal.

(a) *(b)* *(c)*

***Figure II.65**: (a) une texture et sa carte de profondeur : texture en relief , (b) placage de texture classique, (c) Placage de texture en relief à partir du même point de vue que de (b)*

La texture en relief peut être utilisée aussi pour la reconstruction et la visualisation d'objets 3D à partir d'un ensemble de vues entourant l'objet considéré.

(a) ***(b)***

***Figure II.66 :** (a) Objet représenté par six textures en relief, (b) vue de l'objet reconstruit par placage de texture en relief*

5.3.3 Façade : une approche hybride entre image et géométrie

La méthode que propose Debevec dans sa thèse [10] est intéressante de part sa simplicité pour l'utilisateur et aussi pour les résultats obtenus qui semblent assez précis, et qui sont en tout cas visuellement réussis. C'est un système de reconstruction de bâtiments qui permet à l'utilisateur de modéliser un environnement architectural afin de le visualiser de différents points de vues.

Façade utilise des photographies, prises avec un appareil calibré. En effet, la calibration interne de l'appareil est nécessaire à la reconstruction. Mais Debevec souligne que pour les scènes architecturales, comportant des droites parallèles et orthogonales, la calibration peut être automatisée.

***FigureII.67** : des vues du Campanile à reconstruire*

La structure de données de base du système est le *bloc (voir figure II.70 (a)).* Ces blocs sont des primitives 3D contenant un certain nombre de paramètres tels que sa taille, sa hauteur …etc. L'utilisateur instancie de tels blocs pour construire un modèle hiérarchique de la scène, les relations entre les blocs permettant de déterminer les coordonnées de leurs sommets dans les repères relatifs aux autres et donc finalement au repère du sol, qui est le repère principal (*voir figure II.69)*. Il faut ensuite mettre en relation des segments de ces blocs avec leurs projections dans les images. Ces relations permettent de calculer les paramètres des blocs qui sont libres et de déterminer les positions des caméras. Il est aussi possible de formuler des contraintes entre les blocs. Ce peut être des contraintes de positionnement relatif entre blocs, des contraintes de taille.

Dans l'exemple de la *figure II.69,* tous les blocs ont même longueur et largeur, et le haut du bloc "first storey" coïncide avec le bas du bloc "roof". Ces contraintes réduisent énormément le nombre de correspondance à établir et le nombre de paramètres à résoudre (de 2896 à 240 paramètres pour le modèle du Campanile ce qui améliore la robustesse et l'efficacité du système.

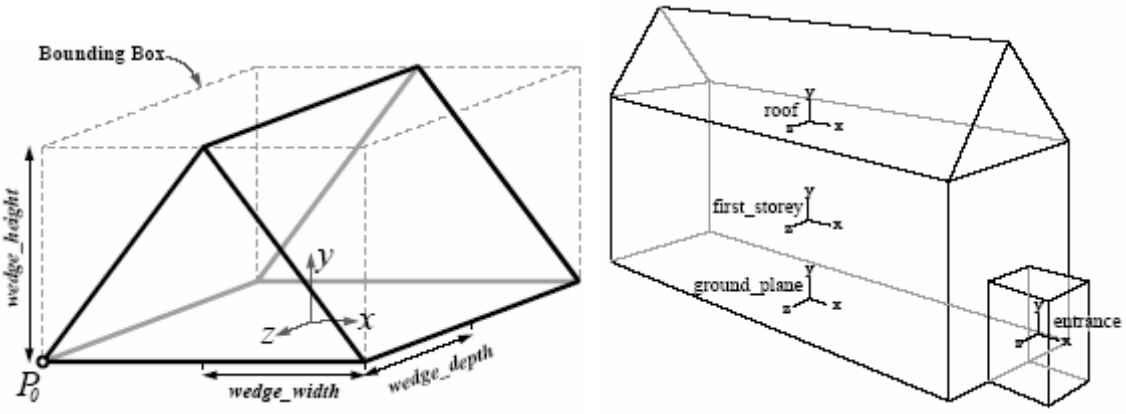
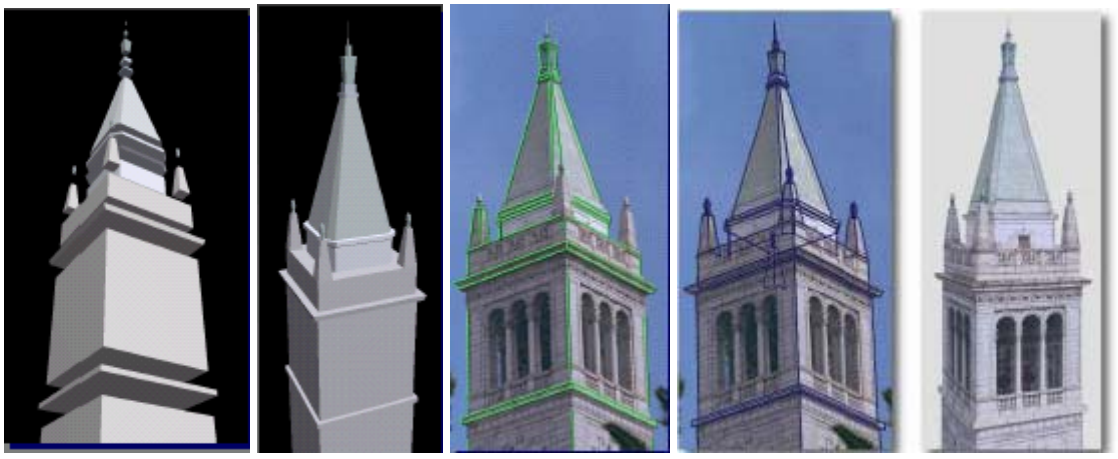


***Figure II.68** : éléments dans façade : les blocs*

Une fois les blocs instanciés, les correspondances et les contraintes spécifiés, le calcul des paramètres libres du modèle et des positions des caméras est lancé.

Après la reconstruction du modèle, les textures sont extraites des photographies. Un algorithme stéréoscopique permet alors de retrouver des détails plus fins sur le modèle (*model-based stereo*). Le principe est d'utiliser le modèle reconstruit pour projeter les images éloignées dans des points de vue plus proche afin d'éliminer les déformations perspectives trop importantes et de pouvoir utiliser un algorithme de correspondance stéréoscopique efficacement. Enfin, à partir des points de vue originaux et du nouveau point de vue de rendu, le système peut choisir la combinaison des images à utiliser pour texturer le modèle de manière à avoir un rendu réaliste (*view-dependent texture-mapping*). Cette méthode produit des résultats visuellement intéressants [10]. Certaines des techniques exposées par Debevec ont inspiré le logiciel CANOMA. Cette technique a été appliquée aussi pour la modélisation des effets spéciaux dans le film « *The Matrix* ».

La démarche de reconstruction de façade est simple et incrémentale pour l'utilisateur, mais serait peut-être plus à classer comme une méthode de construction à partir de photographies que de *re*construction, du fait de l'utilisation d'un modeleur. Enfin, l'utilisation de blocs, même s'ils facilitent la modélisation, pose le problème du manque de généralité de la méthode et de la nécessité d'ajouter des primitives pour des modèles plus complexes ou jamais rencontrés.



(a) (b) (c) (d) (e)

***Figure II.69** : (a) modèles de bloc, (b) Modèle recouvré, (c) Marquage des bords et des conteurs sur la vue réelle, (d) projection des bords du modèle recouvré sur la vue réelle, (e) le modèle du campanile synthétisé.*

6. Conclusion

Nous avons vu dans ce chapitre, un aperçu sur les principes de base de la synthèse de nouvelles vues à partir d'images réelles. Nous avons vu également que la synthèse à partir d'images est un domaine de la vision par ordinateur. L'un des objectifs de cette dernière, est reconstruction de la structure tridimensionnelle (3D) de l'espace à partir d'une ou plusieurs images.

Deux méthodes sont qualifiées pour résoudre ce problème, la première est dite active, qui est facile à réaliser et ne nécessite pas beaucoup de calcul, mais en revanche elle est coûteuse (matériels sophistiqués).

Par contre la deuxième méthode (stéréovision) nommée passive nécessite des caméras pour capter l'environnement. Elle s'appuie sur la connaissance de la géométrie du capteur, pour réduire les difficultés liées à la compréhension et l'interprétation de l'environnement, beaucoup de données et de calculs utilisés pour accomplir le processus du traitement.

Du fait que nous employons la stéréovision dans notre travail, nous avons prêté plus d'attentions à cette partie et nous avons détaillé les trois étapes nécessaires pour arriver à une reconstruction 3D :
- La première étape consiste à bien comprendre la formations de l'image par le modèle dit de « trou d'épingle » et à bien maîtriser les paramètres des caméras de prise de vue (calibrage).

- La seconde étape consiste à mettre en correspondance des indices visuels homologues extraits des deux images. Deux primitives images, une dans chaque image, mises en correspondance représentent les deux projections de la même primitive réelle dans la scène. Cette étape représente la phase cruciale et délicate de tout système de vision stéréoscopique.

- La troisième étape est la reconstruction tridimensionnelle de la scène : connaissant les paramètres du système optique de prise de vue obtenus lors d'une étape de calibration, on calcule pour chaque paire de primitives images homologues les coordonnées réelles 3D de cette primitive dans l'espace de la scène. La connaissance de la disparité entre les deux images stéréoscopique permet, par triangulation, de calculer la profondeur.

Et finalement, nous avons passé en revue les méthodes d'IBMR les plus connues dans la littérature illustrées par des exemples des résultats les plus significatifs. Ces méthodes sont diverses et dépendent souvent des moyens mis en oeuvre.

CHAPITRE III

LES TECHNIQUES D'IBMR POUR LES APPLICATIONS EN RÉALITÉ AUGMENTÉE

1. Introduction

La réalité augmentée consiste à augmenter la perception visuelle du monde réel par l'insertion réaliste d'objets visuels synthétiques. L'approche traditionnelle consiste générer des images d'un objet ou d'une scène 3D à l'aide d'un algorithme de rendu appliqué à un modèle tridimensionnel construit à l'aide d'un logiciel de modélisation. Ces méthodes utilisent les outils de l'infographie pour établir des modèles complexes des objets virtuels ajoutés dans la scène réelle. Un matériel graphique puissant est nécessaire pour accomplir les tâches de calcul intensif pour le rendu en temps réel. Les images de synthèse ainsi générées n'étonnent plus grand monde, pourtant les chercheurs ne cessent d'augmenter la qualité graphique des images tout en réduisant leur temps de calcul, ouvrant ainsi la voie à de nouvelles applications, toujours plus complexe.

Le coût en temps de calcul des étapes de modélisation et de rendu, ainsi que le faible réalisme des images produites ont encouragé le développement de nouvelles techniques basées sur des images réelles de la scène que l'on cherche à représenter. Le but de ces techniques est d'améliorer la modélisation d'environnements en 3D, tant au niveau de la précision et de la rapidité de conception, qu'au niveau du réalisme. En effet, utiliser des images réelles pour créer des images de synthèse offre un double avantage : l'élimination du difficile problème de modélisation géométrique et photométrique complète du monde réel et l'accélération de l'étape de rendu. En fait, les vues disponibles de la scène contiennent des informations géométriques et des informations de texture et couleurs sous une forme déjà rendue car les objets sont éclairés par une source de lumière réelle.

2. La réalité augmentée

La Réalité Augmentée (RA) a pour but d'améliorer notre perception du monde réel par ajout d'éléments qui ne sont pas a priori perceptibles par l'oeil humain.
Dans son état de l'art de la Réalité Augmentée, Azuma [28,29] définit la RA comme un système capable de combiner des images réelles et virtuelles, en 3D et en temps réel. La composition doit être effectuée en 3D, c'est-à-dire que nous devons disposer d'objets virtuels modélisés en trois dimensions, et positionnés dans un repère 3D associé à la scène. La définition exclut donc les compositions 2D, où des images quelconques (dessins, images de synthèse calculées selon une projection quelconque ...) sont simplement collées par-dessus des images réelles. La composition doit aussi être interactive, en temps réel : la définition d'Azuma concerne donc principalement les applications d'immersion où l'utilisateur perçoit des objets virtuels en même temps que l'environnement réel dans lequel il évolue.

***Figure III.1** : À gauche scène réelle, à droite scène augmentée*

Cependant, le terme de Réalité Augmentée est aussi couramment utilisé pour désigner la combinaison d'images réelles et virtuelles en 3D, mais sans la contrainte temps réel [63]. On parle aussi de post-production, c'est-à-dire que l'insertion des objets virtuels se fait généralement dans une étape postérieure à l'acquisition de la séquence vidéo. L'opérateur peut donc passer autant de temps qu'il le souhaite pour traiter chaque image de la séquence, ce qui conduit généralement à un résultat plus précis et plus réaliste qu'en temps réel.

2.1 Intérêts

Pourquoi doit on combiner le monde réel avec le monde virtuel ? La combinaison de la réalité augmentée apporte de l'intérêt. Elle complète la capacité de percevoir et d'intervenir au monde réel. Les objets virtuels affiche des informations que l'utilisateur ne peut pas sentir directement par ses sens. Les informations transmises par des objets virtuels aide l'utilisateur à réaliser des tâches du monde réel. La *figure III.2* présente deux exemples de la réalité augmentée.

(a) *(b)*

***Figure III.2** : (a) Annotations sur des voitures de course dans une diffusion en directe, (b) un simple scénario d'une table ronde dans le domaine de la planification urbaine*

2.2 Réalité augmentée contre réalité virtuelle

La réalité augmentée est une variation de la réalité virtuelle. Dans la réalité virtuelle, tout le monde réel est synthétisé par l'ordinateur pour créer un environnement virtuel. L'utilisateur est alors immergé complètement dans cet environnement et il est isolé du monde réel. Malheureusement, l'obtention d'un monde virtuel réaliste nécessite de disposer des modèles très précis de l'environnement et se révèle donc très coûteuse, surtout dans le cas d'environnements complexes du fait que les informations du monde réel sont trop vastes et complexes donc très difficiles à régénérer sur ordinateur.

Au contraire, la réalité augmentée fournit une vue composite à l'utilisateur. C'est une combinaison de la scène réelle vue directement par l'utilisateur et de la scène virtuelle produite par ordinateur qui fournit des informations complémentaires. L'utilisateur ne peut pas distinguer les objets virtuels de ceux du monde réel. Pour lui il existe seulement un monde unique où les frontières entre le réel et le virtuel sont effacées.

En général, RA hérite plusieurs caractéristiques de RV donc elles possèdent plusieurs points communs. L'interactivité en temps réel est un exemple, les deux systèmes ont une

architecture composée d'un générateur, des dispositifs d'affichage, d'un système de traçage. Ils ont aussi des caractéristiques comme interactif en temps réel.

2.3 Technologies d'affichage

Les technologies d'affichage joue un rôle très important dans un système de RA principalement dans les applications d'immersion où l'utilisateur perçoit des objets virtuels en même temps que l'environnement réel dans lequel il évolue. Les trois configurations d'affichage les plus populaires actuellement pour la réalité augmentée sont les moniteurs d'affichage et deux sortes de casques HMD (*Head Mounted Display*) : les HMD optiques et les HMD vidéos [59].

2.3.1 Affichage à base de moniteurs

L'approche la plus simple est l'affichage sur moniteur, comme représenté sur la *figure III.3* La caméra vidéo capture continuellement des d'images du monde réel et envoie chacune d'elles dans le système d'augmentation. Les objets virtuels sont alors fusionnés avec les images capturées. Le résultat final de cette fusion est afficher finalement sur un moniteur de bureau standard. L'avantage de cette technologie d'affichage est sa simplicité, tout ce qui est exigé est un PC de bureau et une caméra vidéo USB. En plus, en traitant chaque image individuellement, le système d'augmentation peut employer des approches de la vision par ordinateur pour extraire les informations de pose (position et orientation) sur l'utilisateur pour l'enregistrement. Cependant cette simplicité limite la liberté de l'immersion. Clairement, le visionnement du monde réel par un petit moniteur de bureau limite le réalisme et la mobilité dans le monde augmenté. En plus, puisque chaque image de la séquence vidéo doit être traitée par le système d'augmentation, il y a un retard potentiel entre l'instant où l'image est capturée à l'instant où l'utilisateur voit réellement l'image finale augmentée.

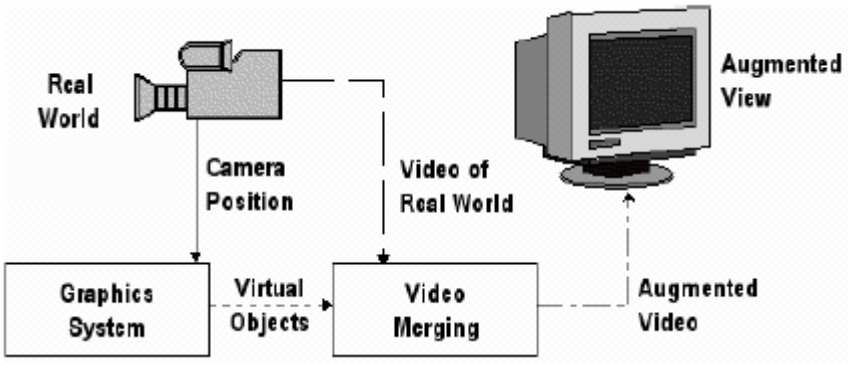


***Figure III.3** : Affichage à base de moniteurs*

2.3.2 HMD optiques

Les HMD optiques disposent d'un système optique qui est partiellement transparent, c'est-à-dire que la lumière du monde réel le traverse, et partiellement réfléchissant, ce qui permet de visualiser les images virtuelles projetées sur le système optique, en même temps que le monde réel *(figure III.4)*. Le combinateur optique réduit souvent l'intensité de la lumière de l'environnement réel, il ne permet qu'à une partie de la lumière de le traverser et l'autre partie est réfléchie. Quelques combinateurs permettent de sélectionner les lumières à réfléchir selon la longueur d'onde. L'utilisateur peut donc régler la quantité de la lumière réfléchie pour obtenir des vues combinées d'une qualité optimale.

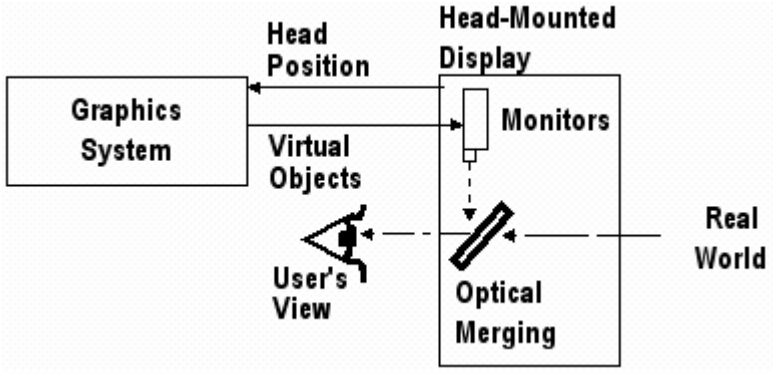


Figure III.4 : *Schéma d'un HMD optiques*

2.3.3 HMD Vidéos

Les HMD vidéos placent deux écrans opaques devant les yeux de l'utilisateur, qui ne perçoit donc plus directement le monde réel. La scène réelle est en fait filmée par deux caméras fixées sur le HMD, et le film est projeté en même temps que les images virtuelles sur les écrans du HMD *(figure III.5)*. Ce système offre donc la possibilité de traiter les images avant de les projeter, ce qui constitue une source d'information extrêmement riche pour la composition.

Il existe plusieurs techniques de composition vidéo. La technique chromatique est la plus simple. Une couleur spéciale est choisie pour le fond des images graphiques de l'ordinateur, par exemple le bleu. Aucun objet virtuel n'a la même couleur avec le fond.
Ensuite toutes les zones bleues sont remplacées par les parties correspondantes de la vidéo du monde réel. Donc l'utilisateur a l'impression que les objets virtuels superposent le monde réel. Une autre technique plus compliquée utilise des informations de profondeur. Chaque pixel de la scène réelle a une profondeur. En comparant la profondeur des pixels, la technique permet d'insérer des objets virtuels dans un environnement réel.

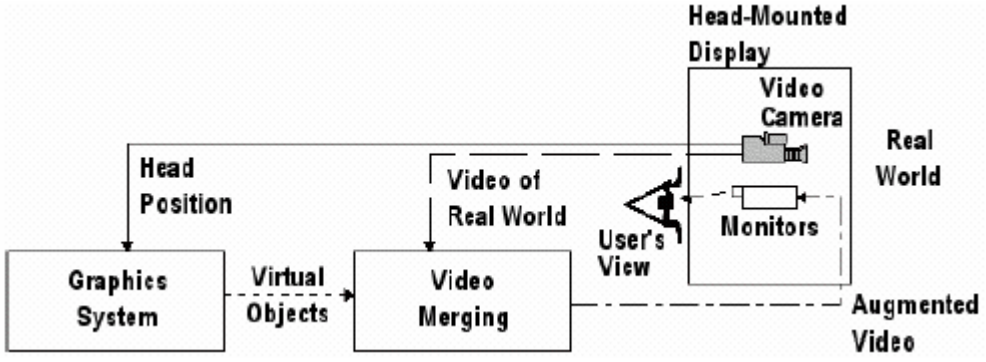


Figure III.5 : *Schéma d'un HMD vidéo*

3. Rendu et modélisation à base d'images et les applications en réalité augmentée

Les objets 3D à base d'images ont actuellement attiré beaucoup d'attentions dans le domaine de la réalité augmentée.
Les intérêts de l'utilisation des objets 3D à partir de photographies dans les applications de la réalité augmentée sont multiples et dépendent de l'application concernée. Mais le point commun est tout de même le besoin d'améliorer la modélisation d'environnements réels augmentés, tant au niveau de la précision et de la rapidité de conception, qu'au niveau du réalisme. Bien entendu, ces préoccupations sont relatives aux besoins et outils que dispose chaque application. Les recherches dans la synthèse à partir d'images sont menées depuis longtemps dans le domaine de la photogrammétrie et de la vision par ordinateur. Récemment, les recherches sont concentrées sur la reconstruction d'objets 3D photo réalistes à partir d'images réelles employant différentes techniques matérielles et logicielles de la vision par ordinateur et de l'infographie. Dans ce qui suit nous présentons quelques applications de la réalité augmentée où des techniques de la synthèse de nouvelles vues à partir d'images réelles sont employées avec un succès prouvé.

3.1 Applications

Les applications potentielles de la RA sont multiples : nous trouvons en particulier des applications en temps réel dans les domaines de la médecine, de la maintenance d'objets manufacturés, de l'industrie, du design intérieur ou du jeu. Les applications de post-production sont principalement les effets spéciaux pour le cinéma ou la publicité. Nous développons à présent chacun de ces aspects.

3.1.1 Médecine

La réalité augmentée est née de la nécessité d'exploiter de façon optimale les données virtuelles issues des simulations. Son application au monde médical est complexe et le plus souvent limitée à des régions où les repères osseux fixes sont nombreux (neurochirurgie, orthopédie). Le but est de visualiser des structures anatomiques et pathologiques qui ne peuvent pas être vues directement, en superposant des modèles virtuels en 3D sur la vue réelle du patient. Le patient devient ainsi transparent.
La RA peut-être utilisée par les médecins pour visualiser des données 3D extraites chez un patient par-dessus le corps du patient (images ultrasonores, tomographie 3D, images à résonance magnétique etc.). À l'aide d'un HMD, le médecin peut par exemple observer les organes à l'intérieur du ventre d'un patient. (figure III.6).

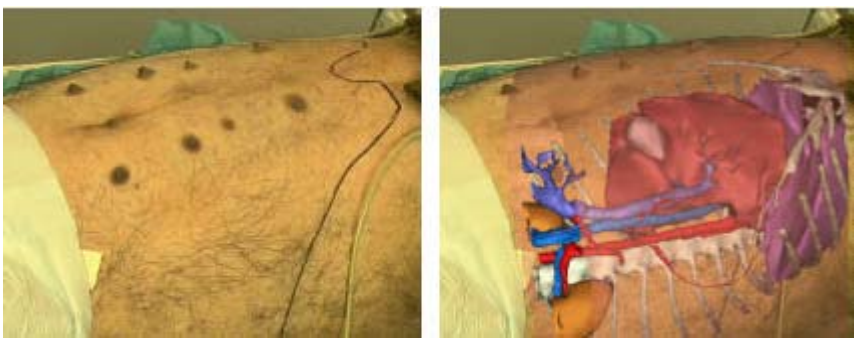


***Figure III.6** : affichage des organes intérieurs d'un patient*

Peuchot propose un système d'assistance à la chirurgie des scolioses [61]. L'objectif est de visualiser les déplacements de la vertèbre sous l'action des forces chirurgicales. Pour cela, une image 3D de la vertèbre est générée et superposée sur la partie visible de la vertèbre incriminée, qui est ainsi localisée directement dans le champ de vision du chirurgien. Dans le même ordre d'idées, des pointeurs virtuels peuvent désigner certains éléments d'anatomie pour aider les étudiants en chirurgie à les visualiser [62].

Les applications précédentes souffrent d'un défaut majeur : le manque de réalisme visuel. En effet, l'utilisateur différencie généralement assez facilement une image de simulation d'une image réelle. Ceci est principalement dû au fait que les techniques d'acquisition et de rendu de textures sont limités à un simple plaquage de couleur, ne tenant ainsi pas compte des effets de lumière propres au contexte, comme par exemple, la lumière froide, les reflets spéculaires, la texture non stationnaire, les inter-réflexions,...etc. dans les techniques de chirurgie mini-invasive. Alors qu'il semble très facile d'acquérir la texture, ainsi que la réflectance, d'un objet rigide de la vie courante (un vase par exemple), le problème est infiniment plus délicat dans le cas d'un organe déformable, entre autre en raison de son accessibilité réduite l'extraction d'une information cohérente de texture et de réflectance à partir d'images est nécessaire.

Ces limitations ont orienté la recherche vers les techniques d'IBMR pour reconstruire des modèles anatomiques 3D de différents organes humains à partir d'images. Par exemple dans [22], les images sont utilisées pour reconstruire la forme et la structure de l'estomac. Le modèle reconstruit est utilisé pour détecter automatiquement les anomalies.

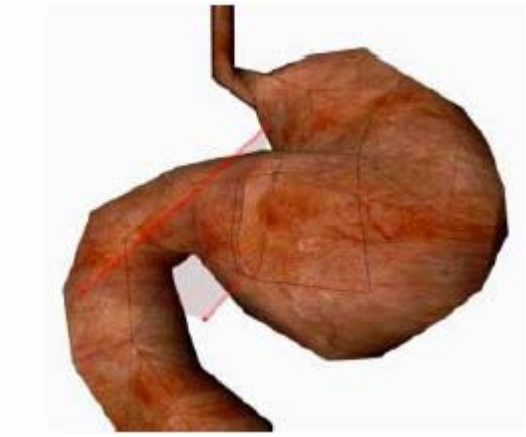


***Figure III.7** : estomac reconstruit à partir d'images*

3.1.2 Design intérieur

Un autre domaine d'application de la RA est le design intérieur : le designer dispose d'une base de données de modèles de meubles ou d'éléments décoratifs, qu'il peut positionner et visualiser dans la pièce physique à meubler. Ceci peut se faire par le biais d'une interface graphique *figure III.8* ou d'un HMD [30] : le designer peut alors se déplacer dans la pièce et visualiser en temps réel les différents éléments ajoutés.

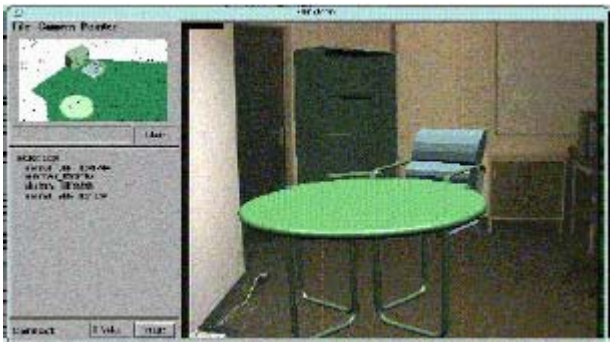


***Figure III.8** : Interface permettant d'insérer des meubles virtuels dans une photographie[30]*

L'apport des techniques d'IBMR consiste seulement à remplacer les objets virtuels synthétiques par des modèles 3D d'objets réels reconstruit par ces techniques. Il existe de nombreux travaux ayant traité la reconstruction 3D à base d'images.

Matusik et al [38]. Ont construit une approche pour acquérir et reconstruire une grande variété d'objets d'une qualité visuelle excellente. Le système d'acquisition est composé d'une table tournante, deux écrans plasma, un ensemble de cameras et des sources de lumière directionnelle et tournante (figure III.9).

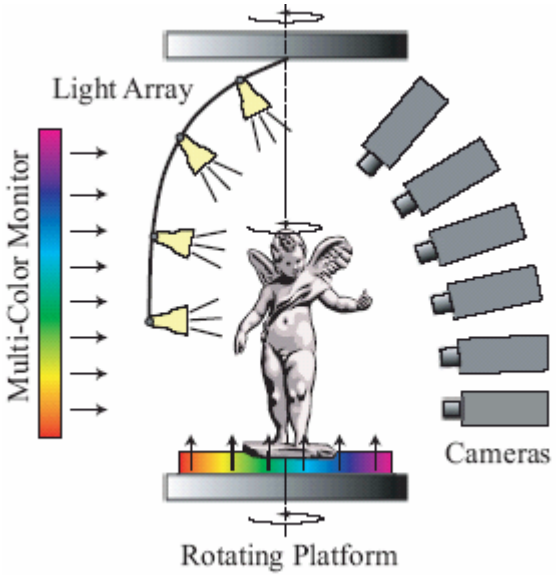


***Figure III.9** : Système d'acquisition combinant les techniques passives et actives*

L'approche de Matusik et al.[38] combine les méthodes actives (lumière structurée) et les méthodes passives (silhouettes, lightfield) pour reconstruire les modèles des objets scannés.

(a)

(b)

FigureIII.10 *: (a) trois objets reconstruits, (b) objets reconstruits incrustés dans un environnement réel (L'appareil photo et la plante ne font pas partie des scènes originales)*

3.1.3 Effets spéciaux

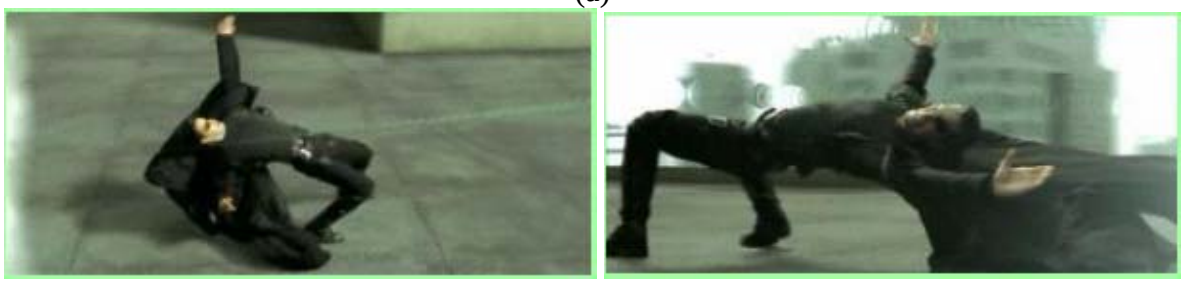
Les effets spéciaux pour le cinéma, la publicité ou les clips vidéo intègrent de plus en plus souvent des images de synthèse qui viennent s'intégrer au film. Le but recherché est de créer un maximum d'émotions chez le spectateur en le plongeant directement dans une action irréelle ou difficile à mettre en oeuvre par des effets spéciaux physiques ou mécaniques. On peut notamment citer quelques réalisations célèbres, comme le film Titanic où les éléments réels et virtuels se côtoient sans même que le spectateur ne s'en rende compte, ou encore l'édition spéciale de la trilogie Star Wars, qui fait apparaître des créatures virtuelles dans un décor réel (figure III.11). Pour ce film, les compositions ont été réalisées manuellement. Thalmann et al. [64] décrivent les différentes étapes à prendre en compte pour l'incrustation semi-automatique d'acteurs virtuels animés dans un environnement réel : extraction des paramètres de la caméra, création et animation des acteurs et rendu des images finales. En particulier, le rendu tient compte des objets réels cachés par les acteurs virtuels et vice versa, des collisions entre les acteurs virtuels et l'environnement réel et des ombres des acteurs virtuels sur le monde réel.

Figure III.11 : *Certains effets spéciaux de la trilogie Star Wars utilisent des images de synthèse superposées aux images réelles*

Dans ces premiers travaux les acteurs ou les objets synthétiques incrustés dans une scène ont un aspect artificiel, ce qui les rend facilement détectables par l'œil. Ce type d'effets spéciaux générés par des méthodes géométriques n'étonne plus grand monde. Pourtant les chercheurs ne cessent d'augmenter la qualité et le réalisme graphique des images.
Ces dernières années, un intérêt croissant est donné aux méthodes de reconstruction 3D à partir d'image réelles dans le domaine de la cinématographie. Cet intérêt est motivé par la rapidité et la qualité visuelle des résultats qu'offert ces méthodes.
Un exemple réussi de l'utilisation des modèles 3D à base d'images pour générer des effets spéciaux est celui de l'utilisation de la méthode *façade* de Paul Debevec [10,20] pour réaliser

un effet spécial révolutionnaire surnommé "Bullet Time" dans le film « The Matrix ». L'effet était de pouvoir tourner la caméra autour des acteurs figés. Cet effet est réalisé en deux étapes : Premièrement, la reconstruction de la scène où se déroule l'action. Cette scène est photographié depuis plusieurs angles de vues pour pouvoir ensuite être reproduite en 3D (architecture et textures) utilisant la technique façade.
Deuxièmement, la reconstruction d'un modèle 3D 'figé' de l'acteur. Pour cela, 120 appareils photos sont installés dans un studio comme indiqué dans la figure III.12, cachés par un tissu vert, seul l'objectif dépasse, il sera ensuite effacé de l'image. Pour cette séquence le nombre d'images prises par seconde varie entre 500 et 1000. Une fois la prise finie les images sont ensuite utilisées pour former un tout homogène : le modèle 3D de l'acteur utilisant toujours la technique de Debevec. L'assemblage final scène / acteur est réalisé par le logiciel « Cineon ».

(a)



(b)

***Figure III.12** : (a) montage du système de prise de vues, (b) deux vues de la scène reconstruite*

"The Matrix" a gagné plusieurs titres pour meilleur effet visuel de l'année 1999 et la technique du « bullet time » est appliquée ensuite dans plusieurs autres films (mission impossible II, Lost in space...).

3.1.4 Musé virtuel

Dans les musées virtuels, les établissements peuvent permettre à des visiteurs d'agir sur les copies virtuelles des objets qui sont physiquement trop sensibles pour toucher, ou égalisent probablement pour conserver sur l'affichage. Les outils de modélisation 3D classiques sont incapables de modéliser la forme des objets de l'acquis culturel d'intérêt. C'est dû à la complexité de forme de la plupart des objets façonnés (par exemple sculptures) et également de l'exactitude élevée demandée. Le modèle 3D dans beaucoup de cas devrait non seulement regarder visuellement semblable au vrai objet, mais devrait également être très précis, d'un point de vue géométrique. C'est nécessaire pour beaucoup d'applications telles que la

construction des catalogues 3D, la reproduction automatique des copies, l'utilisation des modèles 3D dans le contexte des plans de restauration, etc...

Figure III.13 : *annotations sur des anciens objets réels*

L'application des méthodes IBMR dans ce domaine est prouvée par de nombreux projets de recherche. Parmi ces projets de conservation de patrimoine culturel nous citons le projet Michelangelo de l'université de Standford [26]. Ce projet combine les technologies de la 3D scanning et les algorithmes de la vision par ordinateur développés par le professeur Marc Levoy et ses étudiants depuis 1992 [26,12].
La reconstruction des modèles 3D des sculptures et des architectures de Michelangelo a été effectuée durant l'année académique 1998/1999 par un groupe de 30 personnes de l'université de Standford et de l'université de Washington en Italie.

(a) *(b)*

Figure III.14 *: (a) la statue de David. (b) Modèles 3D des objets reconstruits par le projet Michelangelo*

Les modèles reconstruits dans ce projet sont d'une grande précision. Des exemples de ces modèles sont disponibles sur le site du projet [26]. Par cette technique de reconstruction les anciens objets (réels) de la figure III.13 deviennent eux aussi virtuels.

4. Conclusion

Dans ce chapitre, nous avons présenté quelques notions et applications du concept de la réalité augmentée. Cette dernière a pour but d'améliorer notre perception du monde réel par ajout d'objets qui ne sont pas a priori perceptibles par l'oeil humain. La réalité augmentée est un domaine très vaste et une large bibliographie traite ce sujet vaste.

L'approche traditionnelle consiste générer des images d'un objet ou d'une scène 3D à l'aide d'un algorithme de rendu appliqué à un modèle tridimensionnel construit à l'aide d'un logiciel de modélisation. Les images de synthèse ainsi générées n'étonnent plus grand monde, pourtant les chercheurs ne cessent d'augmenter la qualité graphique des images tout en réduisant leur temps de calcul.

Le coût (en temps de calcul) des étapes de modélisation et de rendu, ainsi que le faible réalisme des images produites ont encouragé l'utilisation des techniques d'IBMR dans le but d'améliorer la qualité de la combinaison réel/virtuel, tant au niveau de précision et de rapidité de conception, qu'au niveau du réalisme. En effet, utiliser des images réelles pour créer des images de synthèse offre un double avantage : l'élimination du difficile problème de modélisation géométrique et photométrique complète du monde réel et l'accélération de l'étape de rendu.

Les objets 3D à base d'images ont actuellement attiré beaucoup d'attentions dans le domaine de la réalité augmentée. Les exemples d'applications que nous avons présenté montre un succès prouvé des techniques de l'IBMR dans les applications de la réalité augmentée.

CHAPITRE IV

UNE NOUVELLE APPROCHE DE RECONSTRUCTION D'OBJETS 3D PAR LA COMBINAISON ENVELOPPE VISUELLE / STEREOVISION

1. Introduction

Ces dernières années, la reconstruction d'objets 3D à partir d'images [13][15][26] a vue naître plusieurs méthodes et techniques qui utilisent uniquement les informations géométriques et photométriques présentes dans les images. Ces techniques ont trouvé des applications passionnantes dans les domaines de la réalité augmentée, la cinématographie, les jeux…etc.
La reconstruction de la géométrie 3D est la clé des méthodes de la reconstruction des objets 3D à partir d'images. Une fois que la structure de la scène est récupérée, nous pouvons examiner la scène depuis des points de vue arbitraires.

Dans ce qui suit, on se place dans deux contextes, le premier est celui de la reconstruction de forme à partir de silhouette (enveloppe visuelle) et le deuxième est ce lui de la reconstruction stéréo.
L'enveloppe visuelle est la forme maximale résultante de l'intersection des cônes des silhouettes de l'objet. Cette forme est cohérente avec toutes les silhouettes de l'objet. L'enveloppe visuelle a été très étudiée, de manière implicite et explicite, dans les communautés de la vision par ordinateur et de l'image de synthèse.
La stéréovision est un domaine de la vision par ordinateur qui a comme objectif l'extraction de la structure tridimensionnelle (perdue pendant le processus de la formation de l'image par projection) d'une scène à partir d'images.

Les deux méthodes citées précédemment ont leurs inconvénients inhérents : les méthodes de la stéréovision sont instables et non fiables en particulier pendant l'étape de la mise en correspondance des primitives stéréo notamment pour les surfaces non texturées et les régions occluses. L'enveloppe visuelle ne peut pas récupérer les régions concaves même si un grand nombre de vues est utilisé. Cependant, les deux méthodes sont tout à fait complémentaires en nature. Comme Simon Baker et al. précisent [65], la technique de l'enveloppe visuelle limite au minimum l'espace englobant l'objet, ce qui aide les algorithmes stéréo à éviter des calculs inutiles pour des endroits en dehors du volume de l'objet (ceci peut potentiellement réduire la possibilité des mises en correspondance incorrectes). Les méthodes de la stéréovision raffinent le modèle reconstruit de l'objet par la détection des points et des régions concaves sur la surface de l'objet. Par conséquent, ces deux méthodes peuvent être combinées pour surmonter leurs inconvénients et pour améliorer la qualité de la reconstruction.

L'idée de la combinaison des méthodes de la stéréovision et de l'enveloppe visuelle est déjà abordée.

Carlos H et al.[51] a utilisé le principe général de la combinaison des deux méthodes c.à.d reconstruction de l'enveloppe visuelle puis utilisation des informations de profondeurs calculées par stéréovision pour creuser les détailles de l'objet sur l'enveloppe visuelle. Sa méthode manque d'optimisation et ne tire pas profit de toutes les informations de l'enveloppe visuelle construite.

Li, et al. [66] ont proposé une méthode pour récupérer le modèle polyèdrale de l'enveloppe visuelle en calculant les intersections de la géométrie constructive du solide (CSG) dans la carte graphique. Le modèle acquis a été employé pour limiter la recherche par pixel sur la droite épipolaire pour la reconstruction stéréo multi-vue. Cependant, cette approche exige de grandes ressources de calcul pour traiter des scènes simples.

Li plus tard explore une idée complémentaire et présente une technique multi-passe [67] pour le rendu en temps réel de l'enveloppe visuelle. Dans cette approche, Li a évité un enveloppe visuelle polyèdrale pour une représentation dépendante de point de vue d'une carte de profondeur, qui était alors projectivement texturé utilisant un ensemble approprié d'images d'entrée pour la synthèse d'image. Cependant, les résultats démontrent que l'approximation de l'enveloppe visuelle à la géométrie de l'objet n'est pas suffisante pour le rendu. Les objets reconstruits sont flous et des artifices étaient présents dans les images particulièrement sur les éléments concaves de la surface de l'objet puisque l'enveloppe visuelle n'a pas exactement représenté la géométrie de l'objet.

Yang, et al. [68] ont présenté une approche multi-passe pour la reconstruction stéréo qui a été mise en application à l'aide d'un matériel graphique programmable. Les résultats sont supérieurs à ceux de l'enveloppe visuelle, mais ils restent des artifices observables dans les images, particulièrement le long des contours de l'objet, qui peuvent êtres enlevés en considérant l'enveloppe visuelle.

Dans ce travail nous traitons l'idée de la combinaison de l'enveloppe visuelle/stéréovision d'une façon différente des approches mentionnées. Premièrement nous proposons une optimisation de l'algorithme de mise en correspondance en utilisant les informations issues de la géométrie de l'enveloppe visuelle et deuxièmement nous procédons différemment dans la combinaison des informations issues des deux méthodes pour la reconstruction de l'objet.

2. L'enveloppe visuelle

Supposons que l'on dispose de plusieurs silhouettes d'un même objet correspondant aux points de vue de différentes caméras.
Comme on vue dans le chapitre II, l'enveloppe visuelle est l'intersection des cônes des silhouettes d'entrée. Le volume qui résulte de l'intersection est une limite approximative de la forme de l'objet. Plus de silhouettes utilisées, plus l'enveloppe visuelle convergera vers un volume qui est plus serré et plus proche de la forme de l'objet réel, mais ce volume ne convergera pas nécessairement à la vraie géométrie de cet objet. Ce là est dû à la présence potentielle des régions concaves sur l'objet qui sont difficiles, sinon impossible, à détecter en utilisant seulement des silhouettes. En dépit de cette imperfection, l'enveloppe visuelle est toujours une bonne première approximation de la géométrie réelle de l'objet.

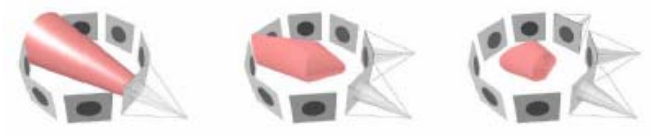


Figure IV.1 *: Principe de la reconstruction de l'enveloppe visuelle*

L'enveloppe visuelle a été très étudiée, de manière implicite et explicite, dans les communautés de la vision par ordinateur et de l'image de synthèse. En particulier, il a récemment été montré [69] que l'enveloppe visuelle d'un objet de surface courbe est un polyèdre topologique que l'on peut déterminer avec une calibration faible. Cependant la solution fournie par cet article est peu adaptée à la plupart des situations réelles. Il existe

beaucoup d'autres algorithmes fournissant des approximations de l'enveloppe visuelle dans les deux communautés.
Certains s'intéressent au volume délimité par l'enveloppe visuelle et se basent sur des discrétisations de l'espace (approches volumétriques). D'autres visent à reconstruire la surface de l'enveloppe visuelle en fournissant des points isolés ou un maillage (approches surfaciques).

Les approches volumétriques : se basent sur une discrétisation de l'espace en cellules élémentaires, les voxels, qui sont sculptés au regard de leur projection sur les images et de l'appartenance ou non de celle-ci aux silhouettes de l'objet. Une première approche fut proposée par Martin et Aggarwal [70], qui utilisaient de cellules parallélépipédiques alignées sur les axes. Plus tard une représentation adaptative de l'enveloppe visuelle fut proposée [71] sous forme d'octree. Au cours des années 90, des approches efficaces [72] ont été présentées pour calculer des représentations voxeliques. Les approches mentionnées sont purement géométriques et ne considèrent pas l'information photométrique. Des méthodes récentes [58] l'utilisent en revanche pour sculpter des voxels selon la *cohérence photométrique* de leur projection sur les différentes images. Consulter [31] pour un tour d'horizon des approches volumétriques. Toutes les approches mentionnées se basent sur une grille régulière de voxels et peuvent traiter des objets de géométrie complexe. Cependant ces approches sont coûteuses en ressources et imprécises puisque la plupart des voxels ne rendent pas compte de la surface de l'enveloppe visuelle, qui est l'information utile.

Les approches surfaciques utilisent une stratégie différente. Des éléments de la surface de l'enveloppe visuelle, tels des points ou des facettes, sont estimés par intersection des surfaces des cônes de vue associés aux contours occultants. Boyer et al.[73] s'intéressent à des points isolés reconstruits en utilisant des approximations locales du second ordre de la surface. Plus récemment, des approches reconstruisent des fragments de surface, ou des bandes de l'enveloppe visuelle.

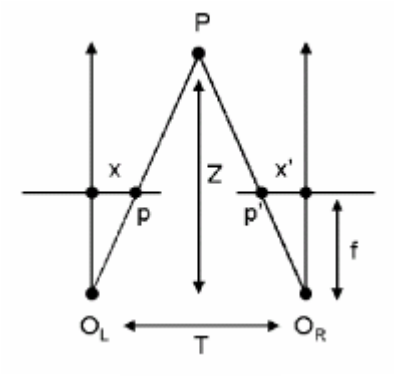
Les approches surfaciques peuvent être précises comparées aux approches volumétriques, cependant les modèles produits sont souvent incomplets ou erronés, en particulier si l'on considère des objets complexes. Ces anomalies découlent de la sensibilité aux instabilités numériques de ces algorithmes et des calculs impliquant les cônes de vue, dont le lieu d'intersection est mal défini. Matusik et al. [74] ont montré que l'on peut obtenir de nouvelles images d'un objet grâce à son enveloppe visuelle en n'effectuant que des calculs 2D. Ce résultat intéressant découle de la structure projective de l'enveloppe visuelle. Cependant la méthode proposée ne fournit pas de modèle géométrique explicite comme le requièrent de nombreuses applications.

3. La stéréovision

L'extraction de la structure tridimensionnelle d'une scène à partir d'images stéréo est un problème qui a été étudié par la communauté de la vision par ordinateur pendant des décennies [7]. Les premiers travaux étés concentrés sur les principes fondamentaux de la correspondance d'images et de la géométrie stéréo. Comme en a vue dans le chapitre II, la recherche dans le domaine de la stéréovision a mûri sensiblement au long des années et beaucoup d'algorithmes de mise en correspondance ont étés développés, permettant au stéréo d'être appliqué à plusieurs domaines.

Une image prise par une seule caméra est considérée comme une représentation bidimensionnelle d'un espace tridimensionnelle. Il y a donc perte d'information durant le processus de

la formation d'une image. En particulier, la troisième dimension. La récupération de cette dernière est le but de la stéréovision.
Tout l'art de la récupération de la dimension perdue par stéréovision consiste à utiliser les informations présentes dans deux images ou plus de la scène prises depuis des points de vue différents.



$$Z = f\frac{T}{d},$$

Figure IV.2 *: principe de la stéréovision*

Du fait de l'écartement des caméras qui prennent les images de cette scène, les deux images d'une paire stéréo ne sont pas sans relations. Ainsi un observateur de la scène verra ses images décalées d'une image de la paire sur l'autre, d'un certain nombre de pixels. Ce décalage est appelé *disparité*. Si on est capable d'attribuer une disparité à chaque pixel d'une image, on attribuera par extension une disparité à tous les points d'un objet sur l'autre image. On sera alors capable de replacer tous les points de cet objet dans l'espace, donc de reconstruire l'objet dans la scène à partir des deux images et de l'information de profondeur inférée.

4. L'approche proposée

4.1 Système de prise de vues

L'objet à numériser est placé sur un support fixe et une caméra tourne autour de lui sur un trajectoire circulaire(dont la géométrie est connue) et une image est prise à pas angulaire régulier (typiquement 5 ou 10 degrés). Afin de pouvoir facilement segmenter l'objet du fond, un font bleu est utilisé. Nous disposant des informations de calibrage de la caméra utilisée : paramètres intrinsèques donnés par le fabriquant, et les paramètres extrinsèques sont issues de la géométrie de la trajectoire.

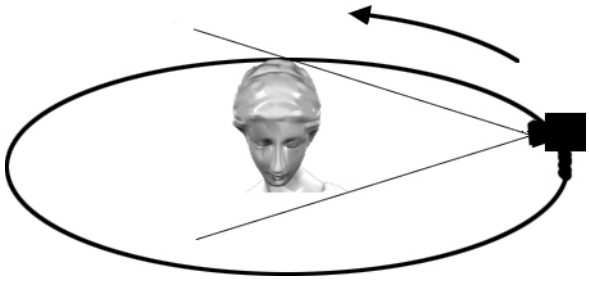


Figure IV.3 *: système de prise de vues*

4.2 Etapes du processus de reconstruction

La rectification est appliquée aux images stéréo au préalable pour aligner les lignes épipolaires, afin d'accélérer le processus de la mise en correspondance. Le processus de la reconstruction peut être décomposé en trois étapes :

(1) Calcul de l'enveloppe visuelle
(2) Mise en correspondance et reconstruction
- Restriction de la zone de recherche sur la ligne épipolaire pour chaque pixel à partir de l'enveloppe visuelle,
- mise en correspondance et calcul de la profondeur et creusage dans l'enveloppe visuelle selon cette profondeur dans la direction de l'image de gauche.

(3) placage de texture sur le modèle.

4.2.1 Calcul de l'enveloppe visuelle

Soit ***n*** images calibrées et rectifiées d'un objet 3D, et soit ***Si*** la silhouette correspondante à l'image ***i***, ***Pi*** la matrice de projection de la caméra correspondante et ***p*** un point 3D quelconque. Nous pouvons définir le cône ***Ci*** généré par la silhouette ***Si*** comme l'ensemble de droites ***liv*** telles que :

$$C_i = \{l_{iv} = P_i^{-1} p, p \in S_i\}$$

L'enveloppe visuelle ***E*** est l'intersection des cônes définie par l'ensemble des silhouettes ***Si, E*** s'écrit :

$$E = \bigcap_{i=1,\ldots,n} C_i$$

Il y a deux méthodes différentes pour la reconstruction de l'enveloppe visuelle: volumétrique [70], [72] et polyèdrale (surfacique) [9]. Puisque la reconstruction volumétrique exige de grandes ressources en mémoire et fournit une précision limitée avec un temps important de rendu. Nous choisissons la méthode polyèdrale qui emploie une représentation plus élégante, qui exige moins de mémoire, et convient au rendu rapide. L'algorithme utilisé pour calculer l'enveloppe visuelle d'un objet est celui proposé par Matusik, Buehler et al. [74]. L'algorithme procède à calculer l'intersection des cônes par l'intersection de polyèdres.

4.2.2 Restriction du champ de recherche d'un correspondant d'un point

Après le calcul de l'enveloppe visuelle nous disposant d'une géométrie approximative de la forme réelle de l'objet. Cette connaissance de la géométrie nous permet de restreindre le champ de recherche des correspondances sur les lignes épipolaires comme suit :

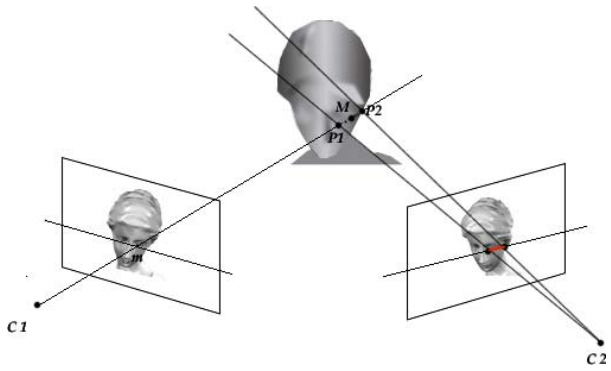


***Figure IV.4** : Restriction du champ de recherche*
Segment en rouge sur l'image de droite

Supposons que nous voulons calculer le correspondant du point ***m*** (Figure 04) sur l'image de gauche,

Soit ***L*** la droite passant par le centre de projection ***C1*** et le point ***m*** sur l'image gauche. L'intersection de l'enveloppe visuelle et la droite ***L*** définie un segment ***[p1p2]*** qui contient obligatoirement le point ***M*** de la surface de l'objet dans l'espace 3D (***M*** a la projection ***m*** sur l'image de gauche).
La projection du segment ***[p1p2]*** sur l'image de droite définie le champ de recherche du point correspondant au point ***m*** sur l'image de droite.
Cette restriction du champ du champ de recherche à deux avantages :

- Accélération de la mise en correspondance (malgré le temps du calcul de la restriction du champ de recherche).
- Augmentation de la fiabilité et la qualité des résultats de la mise en correspondance (réduire la possibilité des mises en correspondance incorrectes).

4.2.3 Mise en correspondance et reconstruction

Dans cette étape nous essayons de mettre en correspondance les points en comparant de petits voisinages (une fenêtre carrée) autour de chaque Pixel. Ils existent plusieurs critères pour évaluer la différence entre les voisinages de Pixel, tels que la corrélation croisée normalisé (NCC : *normalized cross correlation*), la somme des différences carrées (SSD *sum of squared difference)*, la somme de la différence absolue (SAD : *sum of absolute difference*), etc... dans notre cas nous avons choisi la méthode SAD qui est la plus rapide et capable de donner des résultats satisfaisants.

Une fois que la mise en correspondance entre deux pixels est établie, on calcule la disparité entre eux. La disparité est définie comme la différence entre les deux positions des deux pixels mis en correspondance dans l'image gauche et droite. Nous pouvons ensuite calculer la profondeur et ainsi la position 3D du point mis en correspondance d'une manière simple en employant les informations de calibrage des caméras.

La position du point 3D reconstruit est utilisée pour creuser les détailles dans la surface de l'enveloppe visuelle selon la direction du rayon parton du centre de projection de l'image de gauche jusqu'au point 3D reconstruit. Ce processus est réalisé pour chaque point de l'image gauche mis en correspondance avec son homologue dans l'image de droite. Ainsi le modèle de l'objet est raffiné successivement en répétant le processus de mise en

correspondance et creusage pour chaque couple de points mis en correspondance et pour chaque couple de vues.

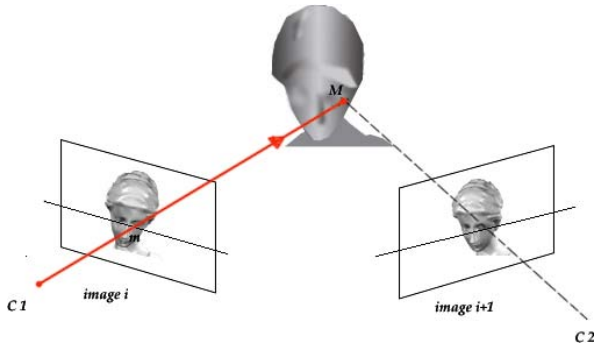


***Figure IV.5** : direction de creusage pour un point reconstruit.*

Le raffinement de la forme de l'objet à partir de son enveloppe visuelle au fur et à mesure que les informations stéréo (positions 3D des points mis en correspondance sur la surface de l'objet) sont disponibles, a deux avantages :

- La mise en correspondance et la reconstruction sont réalisée en une seule passe ce qui diminue le temps de calcul et évite l'utilisation de variables intermédiaires (tableaux volumineux).
- Le champ de recherche des correspondants sera plus restreint dans les prochains couples d'images stéréo.

4.2.4 Placage de texture

Un placage de texture est finalement effectué selon la méthode décrite dans [75].

Le pseudo code de l'algorithme général de la reconstruction de l'objet 3D par la combinaison des deux techniques est comme suit :

Debut
Construire l'enveloppe visuelle
Pour chaque couple ***(i,i+1)*** d'image, ***i = 1 : n*** **faire**
    Pour chaque pixel de l'image ***i*** **faire**
        (1) Calculer la zone de recherche de son correspondant dans l'image i+1
        (2) Rechercher le correspondant de ce pixel dans cette zone ;
        Si correspondant trouvé **alors**
            (3) calculer sa position 3D et creuser dans l'enveloppe visuel selon la direction entre le centre projection de l'image ***i*** et le point 3D construit ;
        finsi
    FinPour
FinPour
Placage de texture ;
Fin

5. Résultas

- ***Prise des images de teste***

A cause de l'indisponibilité du matériel nécessaire pour acquérir des image d'un objets réel (a cause notamment des problèmes de calibrage), Les images utilisées pour tester notre méthodes sont prise par le logiciel 3D StudioMax , c'est-à-dire le système de prise de vue est simulé sur ce logiciel, ce qui permet de prendre des images calibrées depuis déférents points de vues d'un modèle 3D de l'objet à reconstruire.

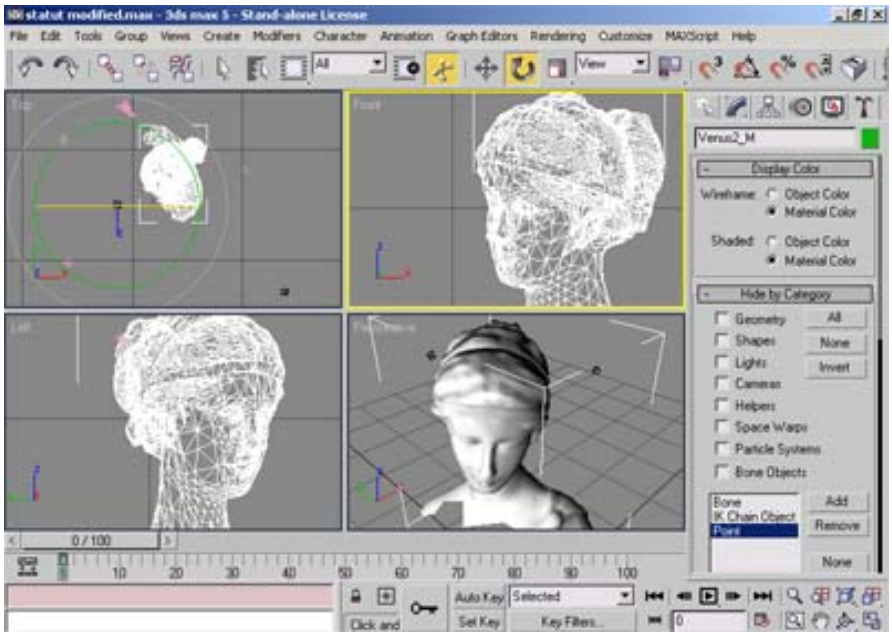


Figure IV.6: Prise des vues avec 3D StudioMax

- ***Préparation des silhouettes***

La préparation des silhouettes à partir des images a été effectuée manuellement afin de traiter plus rapidement les autres aspects du problème.

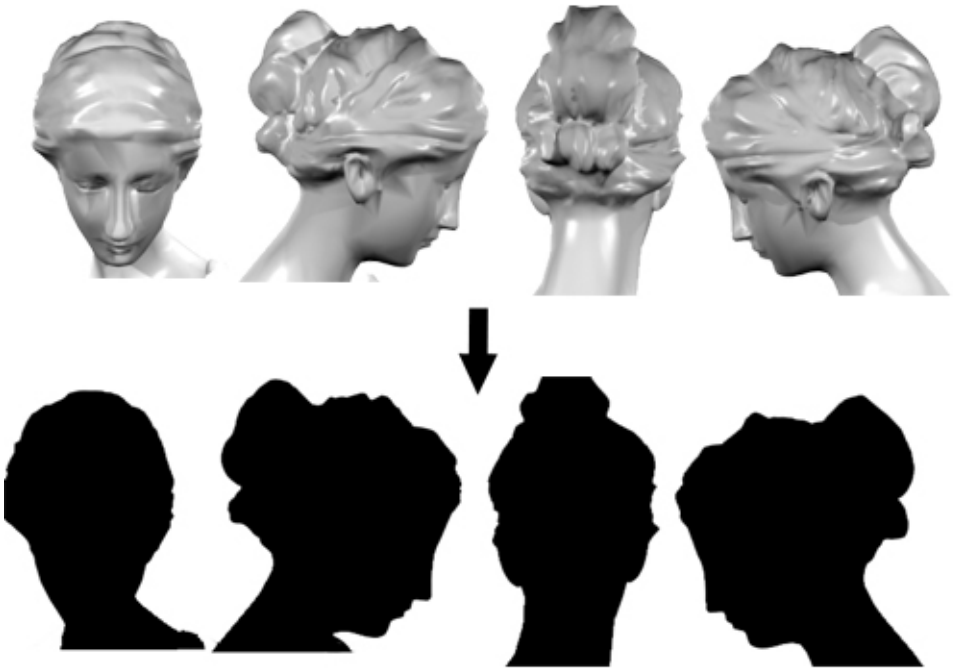


Figure IV.7 : Quatre vues de l'objet et les silhouettes correspondantes

- ***Résultas de l'algorithme de stéréovision***

La figure suivante montre le résultat de l'algorithme de mise en correspondance avec une visualisation 3D en Matlab.

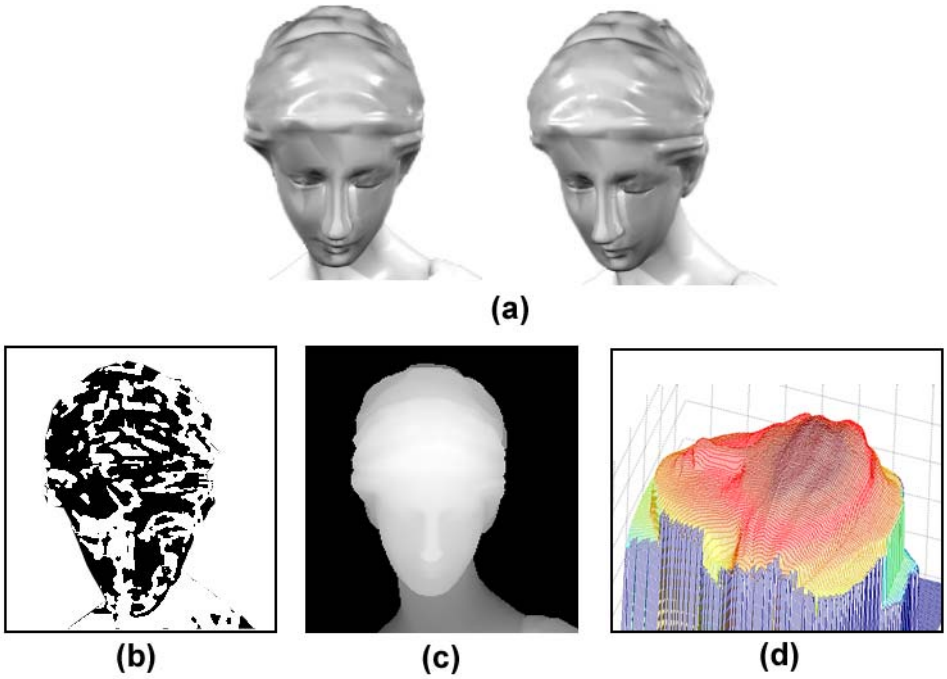


Figure IV.8 : (a) Deux images stéréo, (b) carte des disparités, (c) carte des profondeurs correspondante, (d) représentation en fil de fer des données de la carte des profondeurs sous Matlab.

- ***Reconstruction de l'enveloppe visuelle***

Enveloppe visuelle reconstruite à partir de huit vues calibrées.

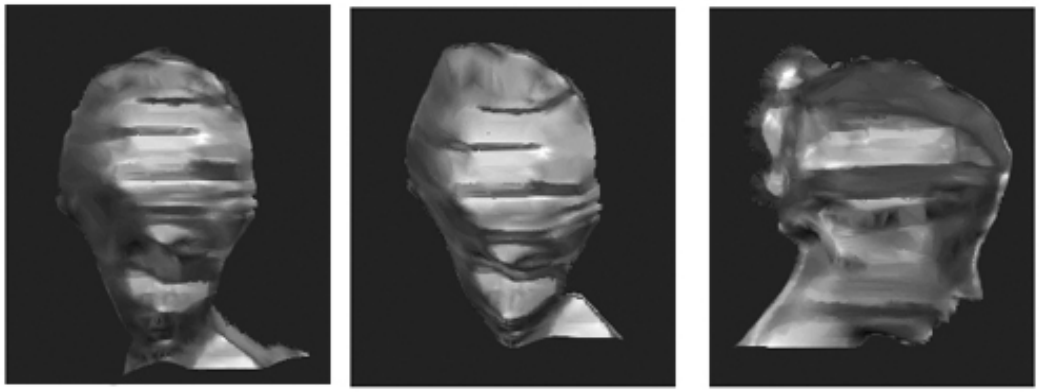


Figure IV.9 *: Enveloppe visuelle reconstruite*

- ***Résultat final de l'algorithme de reconstruction***

Figure IV.10 *: Résultat final de la reconstruction*

6. Conclusion

Nous avons présenté une méthode qui combine la reconstruction à partir d'images stéréo et la technique de l'enveloppe visuelle pour la reconstruction d'objets 3D. D'abord, nous avons employé les silhouettes des images prises depuis plusieurs points de vue pour construire une enveloppe visuelle comme première forme de l'objet. L'enveloppe visuelle est alors employée pour limiter la longueur de champs de recherche des correspondances sur la ligne épipolaire. La limitation de champs de recherche améliore la vitesse et la qualité de la reconstruction stéréo. L'information stéréo a compensé certains des inconvénients inhérents de la méthode de l'enveloppe visuelle, tel que l'incapacité de reconstruire les détails extérieurs et les régions concaves.

CONCLUSIONS ET PERSPECTIVES

Conclusion et perspectives

Nous avons présenté dans ce mémoire les principes de base de la synthèse de nouvelles vues à partir d'images réelles. La synthèse à partir d'images est l'intersection entre l'infographie et la vision par ordinateur. L'un des objectifs de cette dernière est la reconstruction de la structure tridimensionnelle (3D) de l'espace à partir d'une ou plusieurs images.

Deux méthodes sont qualifiées pour résoudre ce problème, la première est dite Active, elle est facile à réaliser et ne nécessitant pas beaucoup de calcul, mais en revanche elle est très coûteuse (matériels sophistiqués).
Par contre la deuxième méthode (stéréovision) nommée passive nécessite des caméras pour capter l'environnement. Elle s'appuie sur la connaissance de la géométrie du capteur, pour réduire les difficultés liées à la compréhension et l'interprétation de l'environnement, beaucoup de données et de calculs utilisés pour accomplir le processus du traitement.
La synthèse de nouvelles vues à partir d'images réelles est un domaine récent et sa bibliographie est un peut limitée, mais beaucoup de techniques d'IBMR ont étés développées ces dernière années. Ces techniques ont trouvé des applications passionnantes dans les domaines de la réalité augmentée, la réalité virtuelle, jeux, cinématographie...etc.

La réalité augmentée est un client par nature des méthodes d'IBMR. Les objets 3D à base d'images ont actuellement attiré beaucoup d'attentions dans le domaine de la réalité augmentée. Les exemples d'applications que nous avons présenté montre un succès prouvé des techniques de l'IBMR dans les applications de la réalité augmentée.

L'approche traditionnelle consiste générer des images d'un objet ou d'une scène 3D à l'aide d'un algorithme de rendu appliqué à un modèle tridimensionnel construit à l'aide d'un logiciel de modélisation. Les images de synthèse ainsi générées n'étonnent plus grand monde, pourtant les chercheurs ne cessent d'augmenter la qualité graphique des images tout en réduisant leur temps de calcul.

Le coût (en temps de calcul) des étapes de modélisation et de rendu, ainsi que le faible réalisme des images produites ont encouragé l'utilisation les techniques d'IBMR dans le but d'améliorer la modélisation d'environnements en 3D, tant au niveau de la précision et de la rapidité de conception, qu'au niveau du réalisme. En effet, utiliser des images réelles pour créer des images de synthèse offre un double avantage : l'élimination du difficile problème de modélisation géométrique et photométrique complète du monde réel et l'accélération de l'étape de rendu.

Dans ce contexte de travail nous avons proposé une approche qui combine deux méthodes d'IBMR pour la reconstruction d'objets 3D réalistes. La première est la reconstruction à partir d'images stéréo et la deuxième est une technique appelée 'enveloppe visuelle'. Ces deux méthodes sont complémentaires en nature. La technique de l'enveloppe visuelle est une première forme de l'objet qui limite au minimum l'espace englobant cet objet, ce qui aide les algorithmes stéréo à éviter des calculs inutiles pour des endroits en dehors du volume de l'objet. Les méthodes de la stéréovision raffinent le modèle reconstruit de l'objet par la détection des points et des régions concaves sur la surface de l'objet. Par conséquent, ces deux méthodes sont combinées pour surmonter leurs inconvénients et pour améliorer la qualité de la reconstruction. L'approche proposée a fait l'objet d'une communication internationale IEEE [76].

Parmi les perspectives immédiates de notre travail on peut bien sûr citer la validation de l'approche sur des scènes réelles. Nous avons déjà démontré la faisabilité d'une telle approche sur des images synthétiques, néanmoins, la validation sur des scènes réelles nous permettrait de pouvoir quantifier tant en précision qu'en qualité la reconstruction ainsi obtenue. Nous travaillons actuellement sur l'adaptation de l'approche proposée pour l'acquisition des scènes dynamiques en temps réel ainsi que l'amélioration de des résultats par le développement de quelques filtres pour améliorer la qualité visuelle du résultat.

RÉFERENCES BIBLIOGRAPHIQUES

Références bibliographies

[1] Darius Burschka, Dana Cobzas, Zach Dodds, Greg Hager, Martin Jagersand, Keith Yerex : *"Recent Methods for Image-based Modeling and Rendering"*. IEEE Virtual Reality 2003 tutorial 1.

[2] Alain CROUZIL : *"Perception du relief et du mouvement par analyse d'une séquence stéréoscopique d'images"*. Thèse PhD de l'Université Paul Sabatier DE TOULOUSE. Septembre 1997

[3] Radu Horaud. '*Vision 3-d projective, affine et euclidienne*'. Technical report, INRIA Rhônes-Alpes and GRAVIR-CNRS, ftp ://ftp.inrialpes.fr/pub/movi/cours/vision-3d.ps.gz, ftp ://ftp.inrialpes.fr/pub/movi/cours/vision-3d.pdf, Janvier 2000.

[4] Patrick Etyngier '*Introduction intuitive à la géométrie projective*'. Tutoiriel de cours Ecole Supérieure d'Informatique - Electronique – Automatique Université de Renne Juin 2003.

[5] Jérome Blanc :' *Synthèse de nouvelles vues d'une scène 3D à partir d'images existantes'*. Thèse PhD de l'Institut National Polytechnique de Grenoble Janvier 1998.

[6] D. Scharstein & R. Szeliski, '*A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms*', IJCV, vol. 47, no. 1, pp. 7-42, 2002.

[7] M. Brown, D. Burschka & G. Hager, '*Advances in Computational Stereo*'. PAMI, vol. 25, no. 8, pp. 993-1008, 2003.

[8] R.Hartley A.Zisserman. '*Multiple View Geometry in Computer Vision'*. Cambridge, 2001.

[9] Adrien Bartoli *'Reconstruction et alignement en vision 3D : points, droites, plans et caméras'***.** Thèse PhD de L'INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE Septembre 2003.

[10] Paul Ernest Debevec. '*Modeling and Rendering Architecture from Photographs*'. PhD thesis . University of California at Berkeley 1996

[11] Gortler, S., et al. '*The Lumigraph'*. *Proc. SIGGRAPH 96* (New Orleans, LA, August 4-9, 1996), pp. 43-54.

[12] Levoy, M., Hanrahan, P. '*Light Field Rendering'* . *Proc. SIGGRAPH 96* (New Orleans, LA, August 4-9, 1996), pp.31-42.

[13] Oliveira, M., Bishop, G. '*Image-Based Objects*'. Proceedings of 1999 ACM Symposium on Interactive 3D Graphics. pp.191-198.

[14] Jonathan Shade, Steven Gortler, Li wei He, and Richard Szeliski.. '*Layered Depth Images'*. *Proc. SIGGRAPH 98* (Orlando, FL, July 19-24, 1998), pp. 231-242.

[15] M. Oliveira, G. Bishop, D. McAllister. '*Relief texture mapping'*. In Siggraph 2000, Computer Graphics Proceedings, 359–368.

[16] Fabio Policarpo, Manuel M. Oliveira, Joào L.D. Comba.' *Real-Time Relief Mapping on Arbitrary Polygonal Surfaces'*. ACM Transactions on Graphics. Volume 24, Number 3, July 2005.

[17] **.**C. Chang, G. Bishop and A. Lastra, "*LDI tree: A hierarchical representation for image-based rendering*", Computer Graphics (SIGGRAPH'99), August 1999, pp. 291-298.

[18] .S. E. Chen, "*QuickTime VR – An Image-Based Approach to Virtual Environment Navigation"*, Computer Graphics (SIGGRAPH'95), August 1995, pp. 29-38.

[19] .S. E. Chen and L. Williams, "*View interpolation for image synthesis"*, Computer Graphics (SIGGRAPH'93), August 1993, pp. 279-288.

[20] .P. Debevec, Y.-Z. Yu and G. Borshukov, "*Efficient View-Dependent Image-Based Rendering with Projective Texture-Mapping"*, 9th Eurographics Rendering Workshop, Vienna, Austria, June 1998.

[21] .Adelson, E. H., and J. R. Bergen, *"The Plenoptic Function and the Elements of Early Vision,"*. Computational Models of Visual Processing, Chapter 1, Edited by Michael Landy and J. Anthony Movshon. The MIT Press, Cambridge, Mass. 1991.

[22] .Danail Stoyanov, Mohamed ElHelw, Benny P Lo, Adrian Chung. '*Photorealistic Visualization for Virtual and Augmented Reality in Minimally Invasive Surgery*'. Visual Information Processing Group, Department of Computing, Imperial College, London, UK 2003

[23] Leonard McMillan, Gary Bishop '*Plenoptic Modeling: An Image-Based Rendering System'*. Proceedings of SIGGRAPH 95 (Los Angeles, California, August 6-11, 1995)

[24] A. Criminisi and A. Zisserman. '*Shape from texture: homogeneity revisite '* University of Oxford BMVC 2000.

[25] Youichi Horry, Ken-ichi Anjyo, and Kiyoshi Arai. '*Tour into the picture: Using a spidery mesh interface to make animation from a single image'*. In Proceedings of the ACM SIGGRAPH Conference (SIGGRAPH-97), pages 225-232, Losa Angeles, CA, USA, august 1997. ACM Press.

[26] Marc Levoy, Kari Pulli, Brian Curless, Szymon Rusinkiewicz, David Koller, Lucas Pereira, Matt Ginzton, Sean Anderson, James Davis, Jeremy Ginsberg, Jonathan Shade, and Duane Fulk. *'The digital michelangelo project: 3d scanning of large statues'*. In Proceedings of the 27th annual conference on Computer graphics and interactive techniques, pages 131-144. ACM Press/Addison-Wesley Publishing Co., 2000. http://graphics.stanford.edu/projects/mich/

[27] .Steven M. Seitz and Charles R. Dyer. '*View morphing'*. In Proceedings of the ACM Conference on Computer Graphics, pages 21-30, New Orleans, LA, USA, August 4-9 1996. ACM.

[28] Ronald T. Azuma '*A Survey of Augmented Reality*'. In Presence: Teleoperators and Virtual Environments 6, 4 (August 1997), 355-385.

[29] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. '*Recent advances in augmented reality*'. IEEE Computer Graphics and Applications, 21(6) :34–47, November/December 2001.

[30] Gilles Simon .'*Vers un Système de Réalité Augmentée Autonome'*. Thèse Phd de l'université Henri Poincaré - Nancy 1 décembre 1999

[31] Greg Slabaugh, Bruce Culbertson, Tom Malzbender, Ron Schafer. '*A Survey of Methods for Volumetric Scene Reconstruction from Photographs'*. International Workshop on Volume Graphics, 2001, held in Stony Brook, New York - June 21- 22.

[32] O. Faugeras . '*Three-Dimensional Computer Vision, A Geometric Viewpoint* '. The MIT Press Cambridge, Massachusetts, 1993.

[33] Boubakeur Boufama .'*Reconstruction tridimensionnelle en vision par ordinateur : Cas des caméras non étalonnées'* . Thèse PhD de l'INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE . Décembre 1994

[34] David ROUSSEL . '*Reconstruction de courbes et de surfaces 3D en stéréo-acquisition'*. Thèse PhD de l'Université Paris XI. janvier 1999

[35] Adrien Bartoli. '*Reconstruction et alignement en vision 3D : points, droites, plans et caméras'*. Thèse PhD de L'INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE septembre 2003.

[36] Manuel M. Oliveira. '*Image-Based Modeling and Rendering Techniques: A Survey'*. Instituto de Informática, RITA · Volume IX · Número 2 · 2002

[37] A. Laurentini. '*The visual hull concept for silhouette based image understanding'*. IEEE Trans. on PAMI, (16(2)), 1994.

[38] Wojciech Matusik, Hanspeter Pfister, Addy Ngan, Paul Beardsley, Remo Ziegler,Leonard McMillan. '*Image-Based 3D Photography using Opacity Hulls '*. ACM SIGGRAPH 2002

[39] Leonid Levkovich-Maslyuk, Alexey Ignatenko, Alexander Zhirkov, Anton Konushin, In Kyu Park, Mahnjin Han, Yuri Bayakovski. '*Depth Image-based Representations and Compression For Static and Animation 3-D Objects'*. IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 14, NO. 7, JULY 2000.

[40] M. Laurent. '*Acquisition 3D au service de la l'ndustrie'*. Magazine électronique CiMax: Edition RobAut N°19 - Septembre-Octobre-Novembre 97

[41] .François Chaumette, Patrick Rives. '*Modélisation et calibration d'une caméra*'. In *AFCET RFIA*, 1990.

[42] .Fadi Dornaika, Christophe Garcia. '*Robust Camera Calibration using 2D to 3D Feature Correspondences'*. Technical Report 1997/3, GMD, 1997.

[43] .Olivier Faugeras, Quang-Tuan Luong, and S.J. Maybank. *'Camera Self-Calibration : Theory and Experiments'*. In Proc. European Conference on Computer Vision, pages 321–334, Santa- Margerita, Italy, 1992.

[44] Pierre-Alexandre Fortin. '*Vision stéréoscopique : appariements*'. Université Laval Canada - juillet 2004

[45] .G-Q. Wei, W. Brauer & G. Hirzinger, *'Intensity- and gradient-based stereo matching using hierarchical gaussian basis functions'*, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 11, pp. 1143-1160, 1998.

[46] .G. Van Meerbergen, M. Vergauwen, M. Pollefeys & L. Van Gool, *'A Hierarchical Symmetric Stereo Algorithm Using Dynamic Programming*', IJCV, vol. 47, no. 1, pp. 275-285, 2002.

[47] .S. Roy & I.J. Cox, *'A Maximum-Flow Formulation of the N-camera Stereo Correspondence Problem'*, ICCV, pp. 492-499, 1998.

[48] .S. Roy, 'Stereo Without Epipolar Lines : *'A Maximum-Flow Formulation*'. IJCV, vol. 34, no. 2, pp. 147-161, 1999.

[49] Boubakeur S. Boufama. *'Recovering the Three-Dimensional Structure Using Two–Dimensional Images : Essentials and Review'*. The International Conference on Complex Systems. Cisc'4 University Of Jijel September 2004.

[50] Renaud Keriven. '*A variational framework for shape from contours*'. Research Report CERMICS-2002-221b Ecole Nationale des Ponts et Chaussees, CERMICS, France. 2002.

[51] Carlos Hernández Esteban, Y. Yemez and F. Schmitt: '*3D Reconstruction of Real Objects from silhouettes and stereo'*. Pattern Recognition and Computer Vision Colloquium Prague, Czech republic may 2001

[52] .Thaddeus Beier and Shawn Neely. '*Feature based image metamorphosis*'. In Edwin E. Catmull, editor, Proceedings of the 19th Annual ACM Conference on Computer Graphics and Interactive Techniques, pages 35-42, Chicago, IL, July 1992. ACM Press.

[53] Heung-Yeung Shum and Li-Wei He. '*Rendering with concentric mosaics'*. In Proceedings of the 26th annual conference on Computer graphics and interactive techniques, pages 299-306. 1999.

[54] Cha Zhang, Tsuhan Chen. '*A Survey on Image-Based Rendering'*. Technical Report AMP 03-03 Advanced Multimedia Processing Lab Carnegie Mellon University Pittsburgh, PA 15213. June 2003

[55] Leonard McMillan. '*An Image-Based Approach to Three-Dimensional Computer Graphics'*. PhD thesis, April 1997.

[56] Steven M. Seitz, Charles R. Dyer. '*Photorealistic Scene Reconstruction by Voxel Coloring'*. Journal of Computer Vision, volume 35, number 2, 1999.

[57] Yasemin Kuzu, Olaf Sinram. ' *Phoptorealistic object reconstruction using voxel coloring and adjusted image orientations* '.Photogrammetry and Cartography Technical University of Berlin Germeny *2001*.

[58] K. N. Kutulakos and S. M. Seitz, "A Theory of Shape by Space Carving," *International Journal of Computer Vision*, Vol. 38, No. 3, July 2000, pp. 199-218.

[59] James R. Vallino. '*Interactive Augmented Reality'*. PhD Thesis, University of Rochester, Rochester, NY. November 1998.

[60] Shahzad Malik. '*Robust Registration of Virtual Objects for Real-Time Augmented Reality* ' Master thesis of Computer Science Carleton University Ottawa, Ontario, Canada - May 2002

[61] Peuchot (B.). '*Virtual Reality As An Operative Tool During Scoliosis Surgery'*. In : Proceedings of Imagina, Monte-Carlo, France, p. 262.

[62] Uenohara M. et Kanade T. '*Vision based object registration for real time image overlay'*. Journal of Computers in Biology and Medecine, 1996.

[63] Zisserman A., Fitzgibbon A. et Cross G. '*VHS to VRML: 3D graphical models from video sequences'*. In : Advanced Research Workshop on Confluence of Computer Vision and Computer Graphics, Ljubljana, Slovenia.

[64] Thalmann (N. M.) et Thalmann (D.).' *Animating Virtual Actors in Real Environments'*. ACMMS'97, vol. 5 (2), 1997, pp. 113-125.

[65] S.Baker, T.Sim, and T. Kanade. '*A characterization of inherent stereo ambiguities'*. In Proceedings of the 8th International Conference on ComputerVision, pages 428–435, Vancouver, British Columbia, July 2001.

[66] Li M, Magnor M, and Seidel H, "Hardware-Accelerated Visual Hull Reconstruction and Rendering." Graphics Interface'2003. (2003),

[67] Li M, Magnor M, and Seidel H, "Improved Hardware-Accelerated Visual Hull Rendering." Vision, Modeling, and Visualization 2003.

[68] Yang, J.C., Matthew, E., Buehler, C. and McMillan, L. 2002. "A Real-Time Distributed Light Field Camera," Proceedings Eurographics Workshop on Rendering 2002, 77-85.

[69] S. Lazebnik, E. Boyer, et J. Ponce. '*On How to Compute Exact Visual Hulls of Object Bounded by Smooth Surfaces'*. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Kauai, (USA), volume I, pages 156–161, December 2001.

[70] W.N. Martin et J.K. Aggarwal. '*Volumetric description of objects from multiple views'*. IEEE Transactions on PAMI, 5(2):150–158, 1983.

[71] C.H. Chien et J.K. Aggarwal. '*Volume/surface octress for the representation of three-dimensional objects'*. ComputerVision, Graphics and Image Processing, 36(1):100–113, 1986.

[72] Richard Szeliski. '*Rapid Octree Construction from Image Sequences'*. Computer Vision, Graphics and Image Processing, 58(1):23–32, 1993.

[73] E. Boyer et M.-O. Berger. '*3D surface reconstruction using occluding contours'*. International Journal of ComputerVision, 22(3):219–233, 1997.

[74] Matusik Wojciech, Buehler C, and McMillan L.' *Polyhedral Visual Hulls for Real-Time Rendering'*,12th Eurographics Workshop on Rendering (2001), 115-125.

[75] F. Schmitt and Y. Yemez. *3d color object reconstruction from 2d image sequences*. In IEEE Interational Conference on Image Processing, 1999. Kobe.

[76] Dib Abderrahim, Bouzenada Mourad, Batouche M^{ed} Chowki :*'Une nouvelle approche de reconstruction d'objets 3D par la combinaison enveloppe visuelle / stéréovision'*. 4ème Conférence Internationale JTEA 2006, 12-14 Mai 2006, Tunisie