

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khider Biskra
Faculté des Sciences Exactes, des sciences de la Nature et de la Vie
Département d'Informatique

Numéro d'ordre :.....
Série :.....



Mémoire

Présenté en vue de l'obtention du diplôme de Magister en Informatique

Option: Intelligence Artificielle et Systèmes Distribués

Titre :

UNE METHODE MULTI-VUE POUR LA RECONSTRUCTION 3D

Par :

M. Djaber ROUABHIA

Soutenu le : ... /.../2011

Devant le jury :

Pr N. DJEDI

Dr. M. C. BABAHENINI MCA

Dr. F. CHERIF

Dr. A. ZIDANI

Université de Biskra

Université de Biskra

Université de Biskra

Université de Batna

Président

Rapporteur

Examineur

Examineur



Sommaire

Sommaire

Liste des figures

ملخص

Résumé

Abstract

Introduction générale

1. **SYNTHESE D'IMAGES CLASSIQUE** 1
 2. **PROBLEMATIQUE** 3
 3. **DOMAINES D'APPLICATIONS DE LA RECONSTRUCTION 3D** 5
 4. **ORGANISATION DU MEMOIRE** 6
-

Synthèse d'images à partir d'images

1. **INTRODUCTION** 7
 2. **MODELE DE LA CAMERA ET FORMATION DE L'IMAGE** 7
 3. **LE CALIBRAGE** 9
 4. **SYNTHESE DE NOUVELLES VUES A PARTIR D'IMAGES REELLES** 10
 5. **METHODES DE RENDU/MODELISATION A BASE D'IMAGES (IBMR)** 11
 6. **TECHNIQUES DE RENDU PUREMENT A BASE D'IMAGES** 12
 7. **CONCLUSION** 17
-

Enveloppe visuelle, Appariement et Stéréovision

1. **INTRODUCTION** 18
 2. **L'ENVELOPPE VISUELLE** 18
 3. **LA MISE EN CORRESPONDANCE** 18
 4. **LA STEREOVISION** 26
 5. **CORRESPONDANCE STEREO** 31
 6. **CONCLUSION** 35
-

Classification multicritères des méthodes de reconstruction

1. **INTRODUCTION** 36
 2. **CLASSIFICATION MULTICRITERES DES METHODES DE RECONSTRUCTION 3D** 36
 3. **RECAPITULATIF** 50
 4. **CONCLUSION** 51
-

Une méthode multi-vue pour la reconstruction 3D

1.	MOTIVATIONS	52
2.	APPROCHE MULTI-VUE POUR LA RECONSTRUCTION 3D	53
3.	RESULTATS	ERREUR ! SIGNET NON DEFINI.
4.	CONCLUSION	66

Conclusion générale et perspectives

1.	PERSPECTIVES	68
----	--------------	----

Références

Liste des figures

Introduction générale

FIGURE 1 : DEUX METHODES POUR LA MODELISATION 3D.	2
FIGURE 2: EFFETS DU NOMBRE DE POLYGONES SUR LA PRECISION DU MODELE 3D.	2
FIGURE 3: RENDU DE LA SCENE 3D MODELISEE.....	2

Synthèse d'images à partir d'images

FIGURE I- 1: (A) LE TROU D'ÉPINGLE ET (B) SON MODELE GEOMETRIQUE.	7
FIGURE I- 2: FORMATION DE L'IMAGE.	8
FIGURE I- 3: UNE MIRE DE CALIBRAGE PHOTOGRAPHIEE SOUS PLUSIEURS ANGLES.	9
FIGURE I- 4: PRINCIPE DU RENDU A BASE D'IMAGES.....	10
FIGURE I- 5: EXEMPLES DE TRAITEMENTS GEOMETRIQUES SUR MODELE 3D CONSTRUIT A BASE D'IMAGES.	11
FIGURE I- 6: ETAPES DE "TOUR INTO THE PICTURE".....	12
FIGURE I- 7: CONTOURS D'UN OBJET.	13
FIGURE I- 8: SHAPE FROM SHADING.....	13
FIGURE I- 9: PRINCIPE DE LA RECONSTRUCTION DE L'ENVELOPPE VISUELLE : ENVELOPPE VISUELLE OBTENUE AVEC 3 VUES.	14
FIGURE I- 10: EXTRACTION DE LA FORME DU TISSU A PARTIR DE LA TEXTURE.....	14
FIGURE I- 11: SEQUENCE D'IMAGES SUBISSANT UN MORPHING.	15
FIGURE I- 12: EXEMPLE D'INTERPOLATION DE POINTS DE VUE.	15
FIGURE I- 13: IMAGES VERTICALES OBTENUES PAR DEFORMATION DE POINTS DE VUE.	16
FIGURE I- 14: LA FONCTION PLENOPTIQUE DECRIT TOUTES LES INFORMATIONS DE L'IMAGE VISIBLES A PARTIR D'UN POINT DE VUE.	16

Enveloppe visuelle, Appariement et Stéréovision

FIGURE II- 1: PRINCIPE DE LA RECONSTRUCTION DE L'ENVELOPPE VISUELLE.	18
FIGURE II- 2: CORRELATION AU HAUT, LES IMAGES GAUCHE ET DROITE AVEC LA FENETRE DE REFERENCE. AU BAS, UNE VUE RAPPROCHEE DES DEUX FENETRES PAREES.	21
FIGURE II- 3: EXEMPLE D'OCCULTATION.	23
FIGURE II- 4: EXEMPLE DE CHEMIN MINIMAL (EN BLANC) DANS L'ESPACE (X, D). LES NIVEAUX DE GRIS REPRESENTENT LES VALEURS DE DISSIMILARITE.	23
FIGURE II- 5: EXEMPLE DE RESULTAT OBTENU PAR LA METHODE DE PROGRAMMATION DYNAMIQUE. ON REMARQUE LES TRAIT HORIZONTALS IRRÉGULIERS ENTRE LES LIGNES.....	24
FIGURE II- 6: VOLUME ENGENDRE EN EMPILANT LES PLANS (X, C) POUR CHAQUE LIGNE DE L'IMAGE.	25
FIGURE II- 7: VOLUME 3D DEFINI EN RELIANT CHAQUE NŒUD A SES VOISINS.	25
FIGURE II- 8: VOLUME 3D AUQUEL ON JOUÛTE UNE SOURCE ET UN DRAIN.	26
FIGURE II- 9: INFINITE DE POINTS DONT LA PROJECTION EST LE MEME POINT SUR L'IMAGE.	26
FIGURE II- 10: ÉLIMINATION DE L'AMBIGUÛTE EN UTILISANT DEUX IMAGES PRISES DEPUIS DES POINTS DE VUE DIFFERENTS.....	27
FIGURE II- 11 : SYSTEME DE STEREOVISION.	27
FIGURE II- 12: SCHEMA SIMPLIFIE DU SYSTEME DE STEREOVISION.	28
FIGURE II- 13: LA GEOMETRIE EPIPOLAIRE.	29
FIGURE II- 14 : GEOMETRIE EPIPOLAIRE ET MATRICE ESSENTIELLE.	31
FIGURE II- 15: DIFFICULTES D'ÉTABLIR DES CORRESPONDANCES STEREO.	31
FIGURE II- 16: RECTIFICATION DE LA PAIRE STEREO.....	33
FIGURE II- 17: CONTRAINTE D'ORDRE P3 ET P1 ONT LE MEME ORDRE MAIS P1 ET P2 SONT INVERSES DANS L'IMAGE DE GAUCHE.	33
FIGURE II- 18: CONTRAINTE D'UNICITE.....	34
FIGURE II- 19: CONTRAINTE DE CONTINUITÉ FIGURALE.....	35

Classification multicritères des méthodes de reconstruction

FIGURE III - 1: L'IMAGE ORIGINALE DE LA MAISON (A GAUCHE) ET UNE IMAGE DU MODELE.	37
FIGURE III - 2: DEUX IMAGES REELLES ET UNE IMAGE DU MODELE 3D OBTENU.....	37
FIGURE III - 3: RESULTAT D'UNE RECONSTRUCTION METRIQUE.	37
FIGURE III - 4: EXEMPLE DE RECONSTRUCTION EUCLIDIENNE.....	38
FIGURE III - 5: RECONSTRUCTION SEMI-AUTOMATIQUE.....	39
FIGURE III - 6: RECONSTRUCTION DESCENDANTE.	39
FIGURE III - 7: RECONSTRUCTION DESCENDANTE.	40
FIGURE III - 8: EXACTITUDE DU MODELE GEOMETRIQUE.	41
FIGURE III - 9: EXEMPLES DE RECONSTRUCTIONS DESCRIPTIVES.	42
FIGURE III - 10: RESTITUTION DE L'APPARENCE VISUELLE D'UN EDIFICE.	42
FIGURE III - 11: RECONSTRUCTION INTEGRANT PLUSIEURS ASPECTS A LA FOIS.	42

FIGURE III - 12: MODELE RECONSTRUIT D'UNE VILLE.	43
FIGURE III - 13: RECONSTRUCTION D'UNE SCENE MOYENNE.	43
FIGURE III - 14: OBJETS DE PETITES TAILLES.	43
FIGURE III - 15: REGROUPEMENT DES ELEMENTS RECONSTRUITS.	44
FIGURE III - 16: VUE PANORAMIQUE.	44
FIGURE III - 17: RECONSTRUCTION 2D/3D.	45
FIGURE III - 18: PRINCIPE DE FONCTIONNEMENT D'UN PALPEUR MECANIQUE.	45
FIGURE III - 19: PROJECTION DU PLAN LASER LUMINEUX SUR UNE SCENE.....	46
FIGURE III - 20: PROJECTION DU PLAN LUMINEUX SUR UNE SCENE L'OBJET EST PLACE SUR UNE TABLE TOURNANTE CONTROLEE PAR ORDINATEUR ET ECLAIREE PAR DEUX RAIES LUMINEUSES.	46
FIGURE III - 21: SCHEMATISATION DE LA RECONSTRUCTION DU RELIEF A L'AIDE DE LUMIERE STRUCTUREE.....	47
FIGURE III - 22: LUMIERE STRUCTUREE ILLUMINANT UN VISAGE.....	47
FIGURE III - 23: RECONSTRUCTION PASSIVE.....	48
FIGURE III - 24: RESTITUTION A PARTIR D'UNE SEULE IMAGE.....	48
FIGURE III - 25: RESTITUTION A PARTIR DE DEUX IMAGES.....	49
FIGURE III - 26: RESTITUTION A PARTIR UN BLOC D'IMAGE.	50

Une méthode multi-vue pour la reconstruction 3D

FIGURE VI - 1: <i>SILHOUETTE A ANALYSEE</i>	53
FIGURE VI - 2: <i>UNE SPHERE QUI ENGLOBE LE VOLUME DE L'OBJET</i>	53
FIGURE VI - 3: <i>DEUX PROFONDEURS P_M ET P_N POUR UNE SILHOUETTE</i>	53
FIGURE VI - 4: <i>CALCUL PARALLELE DES EXTREMITES E_2 ET E_1 POUR LA PREMIERE SILHOUETTE</i>	54
FIGURE VI - 5: <i>CALCUL PARALLELE DES PROFONDEURS GAUCHE ET DROITE POUR UNE LIGNE</i>	54
FIGURE VI - 6: <i>CALCUL PARALLELE DES PROFONDEURS GAUCHES POUR PLUSIEURS LIGNES</i>	55
FIGURE VI - 7: <i>PROFONDEURS DE LA MEME LIGNE DE DEUX SILHOUETTES SUCCESSIVES</i>	55
FIGURE VI - 8: <i>PRESENCE DE BRUITS APRES LA MISE EN CORRESPONDANCE</i>	57
FIGURE VI - 9: <i>MINIMISATION DE LA ZONE DE RECHERCHE EN UTILISANT L'ENVELOPPE VISUELLE</i>	58
FIGURE VI - 10: <i>DELIMITATION VERTICALE (PAR LES VALEURS E_1 ET E_3) DE LA ZONE DE RECHERCHE</i>	58
FIGURE VI - 11: <i>CALCUL DE LA DISTANCE D SACHANT P ET A</i>	59
FIGURE VI - 12: <i>CALCUL DE LA VALEUR D_{MAX_M} POUR LA LIGNE M (AVEC $A=10^\circ$)</i>	59
FIGURE VI - 13: <i>RESTRICTIONS SUCCESSIVES DE LA ZONE DE RECHERCHE POUR UN POINT</i>	60
FIGURE VI - 14: <i>SYSTEME DE PRISE DE VUES</i>	61
FIGURE VI - 15: <i>LES 36 PHOTOGRAPHIES UTILISEES POUR LE TEST DE NOTRE APPROCHE</i>	63
FIGURE VI - 16: <i>LES 18 SILHOUETTES UTILISEES POUR LE TEST DE NOTRE APPROCHE</i>	63
FIGURE VI - 17: <i>DIFFERENTS POINTS DE VUE DE L'ENVELOPPE VISUELLE OBTENUE</i>	64
FIGURE VI - 18: <i>NUAGE DE POINTS REPRESENTANT L'OBJET</i>	65
FIGURE VI - 19: <i>RECONSTRUCTION PARTIELLE DE L'OBJET</i>	66

ملخص

تَشَكُّلُ الصُّورِ مَصْدَرًا لِلْمَعْلُومَاتِ لَا يُسْتَهَانُ بِهِ، سِوَاءَ تَعَلُّقِ الْأَمْرِ بِالتَّفْكِيرِ (باعتبارها كمساعدة في اتخاذ القرارات، أو في التصميم)، أو بالتواصل (من خلال إنتاجها وطريقة تمثيلها).

يَهْدَفُ إِعَادَةُ التَّمْدِجَةِ ثَلَاثِيَّةِ الأَبْعَادِ إِلَى إِنْشَاءِ نَمَازِجِ ثَلَاثِيَّةِ الأَبْعَادِ إِنْطِلَاقًا مِنْ مَعْلُومَاتٍ مُسْتَخْرَجَةٍ مُبَاشَرَةً مِنْ مَشَاهِدٍ حَقِيقِيَّةٍ. وَهِيَ تُشَكِّلُ بَدِيلًا جَدُّ مُهِمًّا لِلتَّقْنِيَّاتِ التَّقْلِيدِيَّةِ، إِذْ تُمَكِّنُنَا مِنْ تَحْسِينِ التَّمْدِجَةِ ثَلَاثِيَّةِ الأَبْعَادِ عَلَى مُسْتَوَى الدِّقَّةِ وَسُرْعَةِ التَّصْمِيمِ وَالْوَاقِعِيَّةِ.

نَظَرًا لِتَعَدُّدِ الطُّرُقِ المُنَاحَةِ لِإِعَادَةِ التَّمْدِجَةِ ثَلَاثِيَّةِ الأَبْعَادِ فَإِنَّ مُحَاوَلَةَ جَمْعِهَا فِي دِرَاسَةٍ وَاحِدَةٍ مِنْ الصَّعُوبَةِ بِمَكَانٍ، وَعَلَيْهِ إِزْتِمَاتُنَا تَصْنِيفُهَا وَفُق مَعَايِيرَ اقْتِرَحْنَاهَا لِلتَّفْرِيقَةِ بَيْنَهَا.

كَمَا اقْتِرَحْنَا أَيْضًا طَرِيقَةً جَدِيدَةً لِإِعَادَةِ التَّمْدِجَةِ ثَلَاثِيَّةِ الأَبْعَادِ إِنْطِلَاقًا مِنْ صُورٍ حَقِيقِيَّةٍ، مِنْ خِلَالِ الْجَمْعِ بَيْنَ اثْنَيْنِ مِنَ الأَسَالِيبِ المَعْتَمَدَةِ، حَيْثُ يَهْدَفُ الأُسْلُوبُ الأَوَّلُ أَسَاسًا إِلَى إِسْتِنْتِاجِ (تَشْكِيلِ) غِلَافٍ بَصَرِيٍّ تَقْرِيبِيٍّ لِلْجِسْمِ المُرَادِ إِعَادَةَ تَمْدِجِهِ، لِإِسْتِعْمَالِهِ لِأَحِقًا فِي إِنْشَاءِ التَّمُودِجِ الفُعَلِيِّ. أَمَّا الثَّانِي فَيَسْتَعْمِلُ عِدَّةَ صُورٍ حَقِيقِيَّةٍ مُلْتَقِطَةً بِطَرِيقَةِ السْتَرِيُو (stéreo)، لِئَسْتَخْرِجَ مِنْهَا مَعْلُومَاتٍ خَاصَّةً بِالأَبْعَادِ وَالإِضَاءَةِ، تَكُونُ لِأَزِمَةِ إِسْتِكْمَالِ التَّمُودِجِ ثَلَاثِيَّ الأَبْعَادِ المُتَحَصِّلِ عَلَيْهِ، وَالْقِيَامِ بِعَمَلِيَّاتٍ مِثْلِ التَّلْوِينِ وَالتَّعْبِئَةِ.

كَلِمَاتُ البَحْثِ : إِعَادَةُ التَّمْدِجَةِ ثَلَاثِيَّةِ الأَبْعَادِ، نِظَامُ الرُّؤْيَةِ سْتَرِيُو، الغِلَافُ البَصَرِيُّ، التَّعْبِئَةُ وَالتَّمْدِجَةُ إِنْطِلَاقًا مِنَ الصُّورِ.

Résumé

Les images représentent, un support considérable d'informations que ce soit pour la réflexion (*aide à la décision, conception, etc.*) ou pour la communication (*production, représentation, etc.*).

Les techniques classiques de synthèse d'images s'attachent toutes à produire des vues d'une scène de modèle géométrique 3D. Leur principe consiste à augmenter un modèle géométrique 3D d'une scène tridimensionnelle, de certaines informations photométriques : couleurs, textures, matériaux, et interactions avec la lumière. Pour ces applications, il est nécessaire d'effectuer une première étape de modélisation manuelle de chaque élément de la scène à synthétiser, et une étape de rendu pour générer les images finales. Ce type de synthèse est gourmand en termes de temps et fournit des résultats de qualité artificielle.

Il est alors proposé de synthétiser de nouvelles vues uniquement à partir de vues réelles de la scène, en simulant le déplacement de la caméra qui les a acquises. Les techniques permettant de synthétiser de nouvelles vues à partir de vues réelles d'une scène sont appelées communément les méthodes de modélisation et rendu basés image (*Image Based Modeling and Rendering : IBMR*).

La reconstruction 3D, consiste à générer un modèle 3D à partir d'informations issues de la scène réelle. Elle constitue une alternative très intéressante, du fait qu'elle vise à améliorer la modélisation d'environnements en 3D, au niveau de la précision et de la rapidité de conception, ainsi qu'au niveau du réalisme.

Puisque le rassemblement des méthodes de reconstruction 3D disponibles dans un état d'art est très difficile (*vue leurs diversité*), nous les avons classés selon différents critères qui permettent de les regrouper en catégories.

Nous avons aussi proposé une nouvelle approche pour la reconstruction 3D à partir d'images ; en combinant deux méthodes d'IBMR : La première est la stéréovision et la deuxième est la technique appelée (*enveloppe visuelle*).

Mots clés : Reconstruction 3D, Stéréovision, Enveloppe visuelle, Rendu et modélisation basés image.

Abstract

The images represent a considerable information support either for reflection (decision aide, designing, etc...) or for communication (production, representation, etc...).

Classically, the images synthesis techniques, aims to produce views of a 3D geometric model scene. Their principle is to augment a 3D geometric model of a three-dimensional scene, with some photometric information: colors, textures, materials, and interactions with light. For those applications, it is necessary to make a first manual modeling step for each element of the scene to synthesize, followed by a rendering step in order to generate the final image. This kind of synthesis is greedy in terms of time and provides an artificial quality results.

It is then proposed to synthesize new views using only real views of the scene, by simulating the displacement of the camera which took the real views. The techniques for synthesizing new views from real views of a scene are commonly known (*Image-Based Modeling and Rendering*: IBMR).

The 3D reconstruction consists to generate a 3D model from information inspired from a real scene. It is a very interesting alternative, because it aims to improve the modeling of 3D environments, in terms of accuracy and design speed, as well as the level of realism.

Since the collection of 3D reconstruction methods available in a state of art is very difficult (for their diversity), we have classified them according to different criteria to group them into categories.

We have also proposed a new approach for 3D reconstruction from images, by combining two IBMR methods: The first is the stereovision and the second technique is called (*visual hull*).

Key words: 3D reconstruction, Stereovision, Visual hull, Image based modeling and rendering.

Introduction générale

1.	SYNTHESE D'IMAGES CLASSIQUE	1
1-1	PHASE DE MODELISATION	1
1-2	PHASE DE RENDU	2
2.	PROBLEMATIQUE	3
2-1	LA RECONSTRUCTION 3D, POUR QUOI ?	4
3.	DOMAINES D'APPLICATIONS DE LA RECONSTRUCTION 3D	5
3-1	ENTRAINEMENT	5
3-2	SIMULATION	5
3-3	COMPRESSION DE DONNEES	5
3-4	IMAGERIE MEDICALE	5
3-5	COMMERCE	6
3-6	CONSERVATIONS	6
3-7	AUTRES APPLICATIONS	6
4.	ORGANISATION DU MEMOIRE	6

Les possibilités d'exploitation des images (*quelle que soit leur origine*), se sont multipliées de manière importante suite à l'émergence récente des nouvelles technologies informatiques de traitement et d'acquisition de l'information. Ces images représentent, en fait, 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.*).

De nouveaux adorables outils de manipulation permettant d'exploiter directement ces images de manière presque exclusivement numérique, ont vu le jour grâce à l'évolution des capacités de traitement, aussi bien matérielles que logicielles. Fournissant ainsi une qualité équivalente, sinon meilleure, que celle obtenue par les techniques plus traditionnelles.

Dans cette introduction, nous essayons d'exposer 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.

1. Synthèse d'images classique

L'intérêt des techniques classiques de synthèse d'images est de produire des vues aussi réalistes que possible. Leur principe consiste à augmenter un modèle géométrique 3D par certaines informations photométriques [80]: (*couleurs, textures¹, matériaux, et interactions avec la lumière*).

Il est donc nécessaire d'effectuer une étape de modélisation manuelle de chaque élément de la scène à synthétiser [34]. Suivie des principales étapes de création des images 3D:

- la position et la trajectoire de la caméra et de la cible,
- le positionnement et le réglage des lumières,
- la création et l'affectation des textures,
- le choix du mode de rendu (*Répartition de la lumière et génération des ombres*),
- le calcul des images (*Affichage des scènes créées sur un écran de visualisation*).

1-1 Phase de modélisation

La **modélisation 3D** est l'étape en infographie² 3D qui consiste à créer, par un logiciel de modélisation 3D, un objet en trois dimensions ; par ajout, soustraction et modifications de ses constituants.

Assisté par un logiciel spécialisé, le concepteur décrit et place les objets manuellement dans le volume de la scène [65]. Il utilise un formalisme approprié pour décrire les caractéristiques géométriques (*formes*) 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

¹ Une "**texture**" désigne 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...*).

² L'**infographie** est la science de l'image numérique, elle définit les graphismes créés et gérés par ordinateur.

son relief¹. Puis, sont définies les différentes lumières éclairant la scène et les caméras desquelles sont prises les vues.

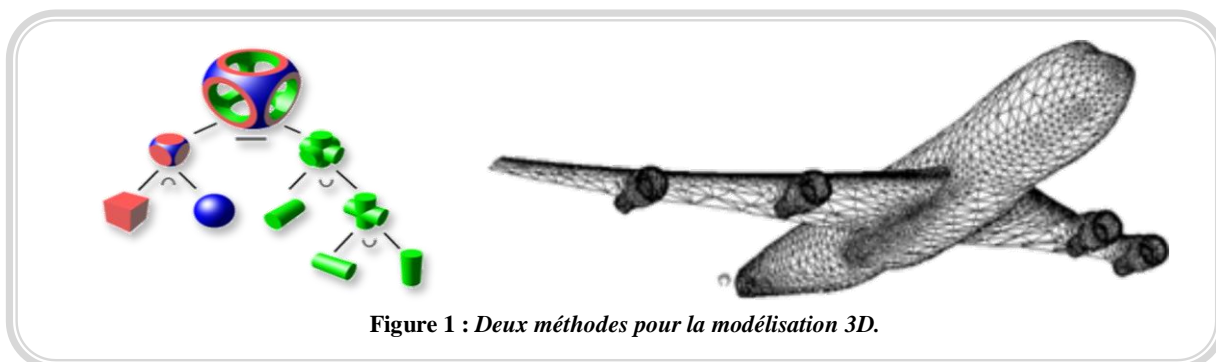


Figure 1 : Deux méthodes pour la modélisation 3D.

On appelle maquette numérique ; le modèle de l'objet ainsi constitué (*représentation informatique à partir d'informations géométriques*). La méthode la plus classique de représentation consiste à décomposer l'objet en « **Facettes** », ou polygones, qui -mis bout à bout- permettent de rendre compte de l'enveloppe extérieure d'un solide [80]. Cette représentation est purement géométrique du fait qu'elle ne met pas en évidence les caractéristiques optiques de l'objet. Plus une maquette comporte de polygones, plus l'image qui en résulte est précise (*Figure 2*).

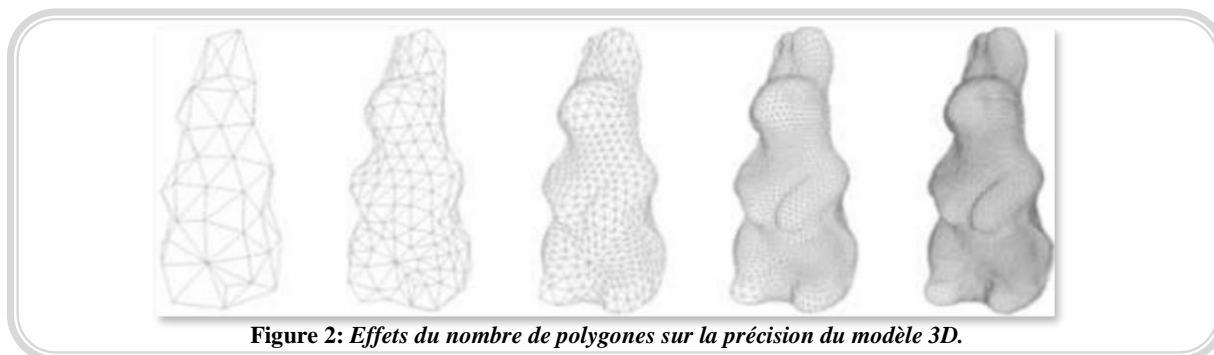


Figure 2: Effets du nombre de polygones sur la précision du modèle 3D.

1-2 Phase de rendu

Le calcul de rendu permet de fabriquer véritablement l'image de synthèse. Le but est de transformer la description 3D de la scène en une image 2D en fonction des différents éléments de la scène et du point de vue duquel elle est regardée. Des images d'un réalisme maximal (*vis-à-vis du modèle*) sont générées par le système de synthèse d'images, et ce par l'application des lois physiques de propagation de la lumière, prise en compte des réflexions, réfractions, diffusions, et interactions avec les matériaux composants la scène.

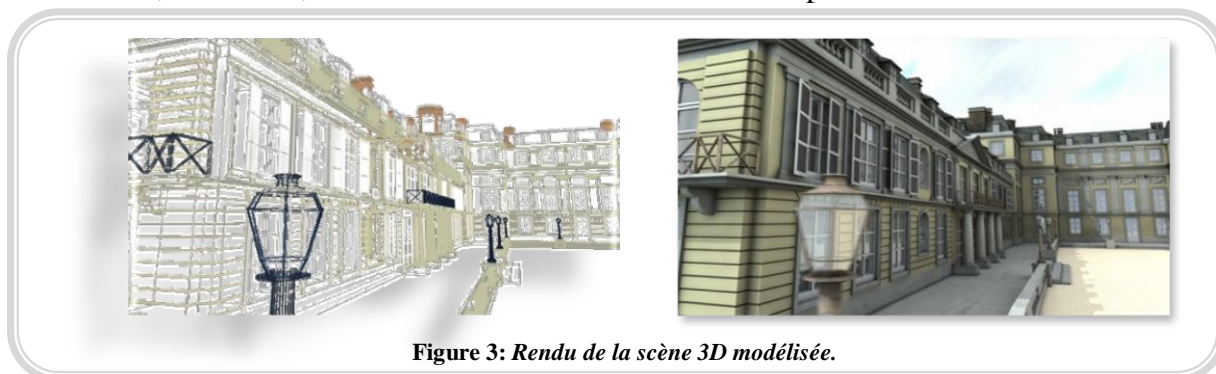


Figure 3: Rendu de la scène 3D modélisée.

¹ Le **relief** désigne une image (*dessin, film, etc.*) donnant l'illusion d'une troisième dimension, visible en stéréoscopie.

Diverses techniques de rendu sont disponibles [65] et permettent par exemple d' :

- Avoir un « **Réalisme lumineux** » en simulant le parcours inverse de la lumière en optique géométrique. Ce qui constitue la base de la technique du « **Lancer de rayons** » (*les rayons se reflètent ou se réfractent selon les matériaux*).
- Avoir un « **Réalisme feutré** » en analysant la lumière comme échange d'énergie entre surfaces ; calculé à partir des propriétés de réflectivité des matériaux, ce qui permet d'obtenir des lumières tamisées et des pénombres. C'est le principe de la « **Radiosité** ».

2. Problématique

Pour les objets fabriqués par l'homme, les méthodes de modélisation classiques donnent des résultats très satisfaisants et de haute qualité (*photographique et cinématographique*).

Mais, elles s'avèrent être inapplicables à des scènes réelles (*vastes et complexes*) qui contiennent un grand nombre de détails (*cas de la forêt*), ou tout simplement un trop grand volume de données (*cas de la ville*), difficilement gérable en pratique [16].

L'apparition de techniques de synthèse de plus en plus réalistes et de plus en plus complexes, a diminué les aspects synthétiques et artificiels des objets produits en fournissant un photoréalisme excellent, mais cela pose plusieurs types de problèmes :

- Les méthodes classiques de synthèse d'images sont gourmandes en **temps de calcul** [16]. Si la scène est trop complexe, synthétiser 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.
- Une scène complexe demande une **modélisation fastidieuse** [74], pouvant se chiffrer en hommes-années¹. 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.
- Un autre constat porte sur la **collecte longue et difficile de données** [16]. Si l'on souhaite, par exemple, modéliser un immeuble existant, il nous faudra utiliser des relevés métriques pris sur le terrain afin de rester le plus fidèle possible à la réalité. L'obtention de ces données et leur saisie peut être un travail harassant.
- En plus ; l'utilisation de ces outils nécessite souvent une **phase d'apprentissage** plutôt longue de la part de l'utilisateur [15], et celui-ci ne maîtrise pas toujours toute la puissance du système, du fait de la grande diversité de fonctions et d'options qu'il propose. La phase de modélisation est donc, la plupart du temps, confiée à un spécialiste en modélisation 3D.

Même si les problèmes rencontrés sont en général différents, les méthodes utilisées pour leur résolution sont souvent partagées entre plusieurs disciplines, et la reconstruction² 3D constitue l'une des approches les plus utilisées :

¹ **homme-année** : unité correspondant au travail d'une personne pendant un an.

² La **reconstruction 3D** consiste à générer un modèle 3D à partir d'informations issues de scènes réelles. (*Plus de détails plus loin*).

- En **Photogrammétrie**¹, on utilise des techniques de photographie stéréoscopique², qui sont des techniques apparues avant l'informatique. Des images chevauchées sont utilisées pour déterminer les formes. L'idée de base est la triangulation³ : Si l'on a deux images de différents points de vue du même point 3D, sa projection 2D se trouve à l'intersection des deux rayons formés par le centre de projection et ce point 3D. 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 calibrés⁴ avec précision.
- La **Vision par Ordinateur et la Robotique**, font de la reconstruction 3D un de leurs principaux champs de recherche [7]. Les méthodes utilisées sont l'inférence de structures à partir du mouvement (*structure from motion* [16], [60]), la stéréovision [82], les cartes de profondeurs⁵, etc. Il est en effet crucial en robotique de pouvoir reconstruire en temps réel un environnement 3D pour la navigation d'une machine. Et 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.
- L'intérêt pour **l'infographie**, est de pouvoir créer plus rapidement des modèles plus réalistes qui correspondent à des objets dont on a des photographies, des plans, des dessins, etc. Les systèmes de modélisation sont souvent sophistiqués, complexes et longs à maîtriser, d'où un intérêt récent pour la reconstruction à partir d'images. De plus, les applications nécessitent de plus en plus de réalisme (*environnements immersifs, réalité augmentée⁶, effets spéciaux, ...*) et les modèles traditionnels sont souvent trop (*propres*), ou (*simples*). 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.

2-1 Intérêts de la reconstruction 3D

La reconstruction 3D consiste à générer un modèle 3D à partir d'informations issues d'une scène réelle. Elle constitue une alternative très intéressante par rapport aux méthodes classiques de modélisation 3D, du fait qu'elle vise à [78] améliorer la modélisation d'environnements en 3D (*au niveau de la précision et du degré du réalisme*), à faciliter la conception et à augmenter sa rapidité.

La synthèse de nouvelles vues à partir de photographies est l'une des approches les plus prometteuses. Les vues disponibles de la scène contiennent des informations géométriques et des informations de textures et couleurs sous une forme déjà rendues ; car les objets sont éclairés par une source de lumière réelle, ce qui offre un double avantage:

¹ La **photogrammétrie** est une technique qui consiste à effectuer des mesures dans une scène en simulant la vision stéréoscopique humaine.

² La **stéréovision** est une méthode de mesure qui consiste à se servir de la prise d'images (*photographiques ou numériques*) sous différents angles de vue pour déterminer les dimensions, les formes ou les positions d'objets.

³ La **triangulation** est une technique permettant de déterminer la position d'un point en mesurant les angles entre ce point et d'autres points de référence dont la position est connue.

⁴ La **calibration** consiste à connaître les paramètres intrinsèques extrinsèques de caméras.

⁵ Des images qui montrent comment se distribuent les hauteurs.

⁶ La notion de **réalité augmentée** désigne les systèmes qui rendent possible la superposition d'un modèle virtuel 3D ou 2D à la perception que nous avons naturellement de la réalité et ceci en temps réel.

- **Eviter la modélisation d'objets effectuée à la main (*complexe et fastidieuse*).**
- **Et accélérer l'étape de rendu.**

Plusieurs méthodes de reconstruction 3D sont proposées dont la littérature. Et l'enjeu effectif, est alors de minimiser au maximum les temps de calcul ou d'améliorer la qualité des scènes modélisées.

3. Domaines d'applications de la reconstruction 3D

Les applications pouvant bénéficier de la reconstruction 3D, sont toutes les applications usuelles de la synthèse d'images : CAO¹, Réalité Virtuelle², Réalité Augmentée, entraînement, simulation, commerce, loisirs, compression de données, et beaucoup de recherches sont en cours dans ces sens [46]:

3-1 Entraînement

Pour toutes sortes d'interventions en milieux hostiles : nucléaire, militaire, spatial ; il est nécessaire de préparer et d'entraîner les hommes à évoluer dans leur futur milieu d'action ; en les plongeant dans une situation simulée le plus fidèlement possible, ils apprennent donc à repérer la topographie du site et de se familiariser à la géographie du lieu à l'aide d'une représentation graphique tridimensionnelle réaliste.

3-2 Simulation

L'intérêt des simulations a été démontré dans des contextes industriels, parfois de façon spectaculaire. Et comme il est fastidieux de produire de tels modèles manuellement, des systèmes automatiques de génération à partir d'images ont ici grand intérêt.

3-3 Compression de données

Calculer de nouvelles vues à partir de quelques images d'une scène, a une application immédiate en compression de données [58]. 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.

3-4 Imagerie médicale

Depuis de nombreuses années, la manipulation d'images médicales a véritablement aidé le chirurgien dans la pratique de son art [9]. L'utilisation de caméras vidéo, d'ordinateurs, et d'appareils de mesure sophistiqués pour préparer et réaliser les interventions, a permis d'améliorer le confort des patients, de diminuer les complications et de faire d'importantes économies de santé publique. La complexité des données 3D, en imagerie médicale nécessite le développement de modèles intégrant de multiples connaissances et de grandes quantités d'informations, et la synthèse d'images à partir d'images est donc complètement adéquate.

¹ La **conception assistée par ordinateur** (CAO) comprend l'ensemble des logiciels et des techniques de modélisation géométrique permettant de concevoir, de tester virtuellement - à l'aide d'un ordinateur et des techniques de simulation numérique - et de réaliser des produits manufacturés et les outils pour les fabriquer.

² La **réalité virtuelle** est une simulation informatique interactive immersive, visuelle, sonore et/ou haptique, d'environnements réels ou imaginaires.

3-5 Commerce

Avec Internet s'est développée la possibilité d'effectuer des transactions financières à distance, et de pratiquer le commerce électronique. Alors, les catalogues agrémentés de photos et de visualisations plus ou moins animées ou interactives, permettent de bien visualiser puis de choisir les produits.

3-6 Conservations

La capture automatique de modèles à partir d'images, peut être utilisée pour la restitution de maquettes numériques d'édifices ou d'objets d'antiquité [39]. Ceci peut simplifier la navigation et les interactions avec les sources documentaires qui lui sont associées. Ces informations et les représentations géométriques multiples des objets constituant l'édifice peuvent être restituées de manière dynamique en fonction de différents points de vue utilisateurs.

3-7 Autres applications

On peut encore citer le domaine des loisirs, et de toutes les applications exigeant un grand nombre d'images: visite de musées virtuels, visualisation de maisons ou d'équipements ménagers pour la vente, de lieux de vacances dans une agence de voyage, jeux vidéo.

4. Organisation du mémoire

La reconstruction 3D est une tâche ardue et une large bibliographie traite ce sujet. Dans le reste de ce mémoire nous essayons d'introduire ce domaine. La description du contenu de chaque chapitre est comme suit:

Dans le chapitre I, nous passons en revue les méthodes de synthèse d'images à partir d'images les plus connues dans la littérature et nous les illustrons par des exemples.

Le chapitre II traite les notions préliminaires liées à l'enveloppe visuelle¹, la stéréovision et la mise en correspondance qui forment la base de notre travail.

Au cours du chapitre III, nous proposons une classification multicritères des méthodes disponibles de reconstruction 3D.

Nous proposons ensuite, dans le chapitre IV, une nouvelle approche de reconstruction 3D d'objets. Notre approche combine deux techniques d'IBMR²; la première est l'enveloppe visuelle et la deuxième est celle de la reconstruction par stéréovision. Nous proposons aussi un algorithme parallèle pour construire l'enveloppe visuelle.

Nous présentons les résultats expérimentaux obtenus par l'application de notre approche, et enfin, une conclusion et des perspectives possibles pour notre travail sont énoncées.

¹ **L'enveloppe visuelle** est une technique qui permet de délimiter approximativement le volume qui englobe un objet.

² Images Based Modeling and Rendering.

Chapitre II

Synthèse d'images à partir d'images

1.	SYNTHESE D'IMAGES CLASSIQUE	1
1-1	PHASE DE MODELISATION	1
1-2	PHASE DE RENDU	2
2.	PROBLEMATIQUE	3
2-1	LA RECONSTRUCTION 3D, POUR QUOI ?	4
3.	DOMAINES D'APPLICATIONS DE LA RECONSTRUCTION 3D	5
3-1	ENTRAINEMENT	5
3-2	SIMULATION	5
3-3	COMPRESSION DE DONNEES	5
3-4	IMAGERIE MEDICALE	5
3-5	COMMERCE	6
3-6	CONSERVATIONS	6
3-7	AUTRES APPLICATIONS	6
4.	ORGANISATION DU MEMOIRE	6
6-3	MORPHING/INTERPOLATION	14
6-3-1	<i>Le morphing</i>	14
6-3-2	<i>Interpolation de point de vue (View Interpolation)</i>	15
6-3-3	<i>Déformation de point de vue (View Morphing)</i>	15
6-4	INTERPOLATION DE RAYONS LUMINEUX ET FONCTION PLENOPTIQUE	16
7.	CONCLUSION	17

1. Introduction

Contrairement à la synthèse d'images traditionnelle, les techniques d'IBMR créent de nouvelles vues directement à partir d'images existantes. Ceci a l'avantage d'être plus économe et de pouvoir créer de nouvelles images photo-réalistes. Pour si faire, il est important de comprendre et de bien modéliser les transformations géométriques qui permettent la formation des images.

Dans cette partie nous présentons, dans la section 2, la formation de l'image et le modèle de caméra. Nous essayons ensuite de faire un survol des méthodes d'IBMR les plus connues dans la littérature, et ce pour avoir une idée sur les techniques utilisées pour la synthèse d'images à partir d'images.

2. Modèle de la camera et formation de l'image

2-1 Modèle de la caméra

Le modèle géométrique le plus couramment employé en vision [24] pour décrire la formation de l'image est le modèle du «trou d'épingle» ou « sténopé » (*pinhole*), illustré par la *Figure I-1(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 I-1(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 $R = (O_c, x, y, z)$ soit direct.

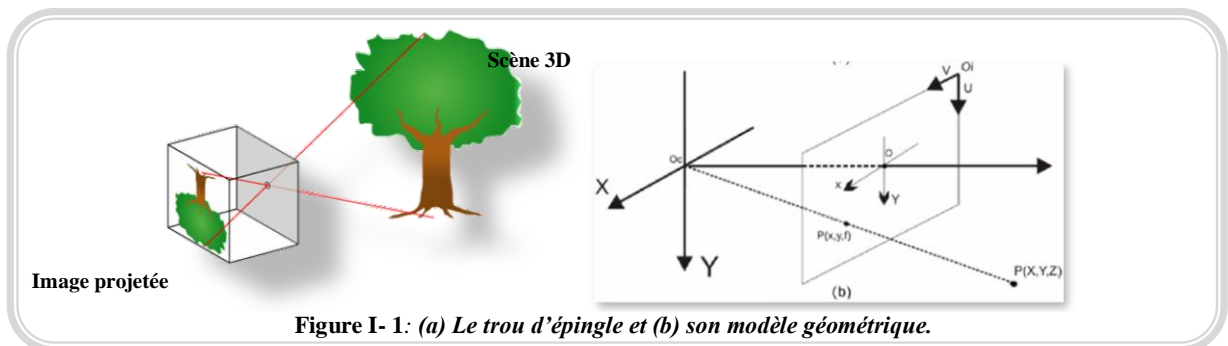


Figure I- 1: (a) Le trou d'épingle et (b) son modèle géométrique.

2-2 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 [59]. 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 : $P_S (X, Y, Z, 1)^T$ (point scène en repère scène)
- Repère de la Caméra:
Point Caméra : $P_c (x, y, z, 1)^T$ (point scène en repère caméra)
Point Image : $p_r (x_r, y_r, 1)^T$ (point image en repère caméra)
- Repère de l'Image :
Point Image : $p_i (u, v, 1)^T$ (point image en repère image)

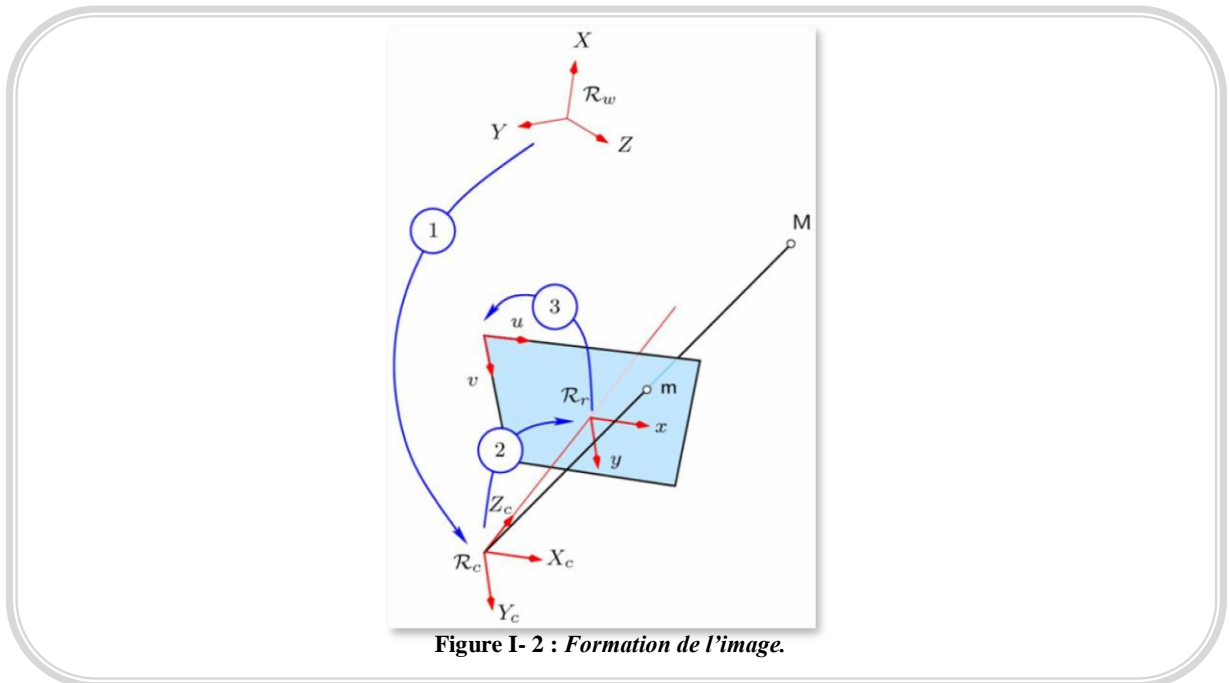


Figure I- 2 : Formation de l'image.

Pour passer des coordonnées définies dans le repère de la scène aux coordonnées images exprimées en pixels¹, trois phases sont nécessaires [24]:

$$(X \ Y \ Z) \xrightarrow{T} (X_c \ Y_c \ Z_c) \xrightarrow{P} (x \ y) \xrightarrow{C} (u \ v)$$

- (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**.
- (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 (elles sont toujours exprimées dans le repère caméra). Cette projection est une projection perspective dans le modèle de trou d'épingle.
- (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:

¹ **PIXEL**, Son nom provient de la locution anglaise **picture element**, qui signifie « élément d'image ».

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

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

$$p_i = MP_s$$

Où : $M = {}_r^i C \cdot {}_s^c T$

${}_r^i C$: est la matrice des paramètres intrinsèques,

${}_s^c T$: est la matrice des paramètres extrinsèques,

M : est la matrice de projection projective d'une image appelée aussi matrice de transformation rigide.

3. Le Calibrage

Le calibrage d'une caméra [17] 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 leurs projections dans une image.

Le calibrage de caméra est un domaine de recherche à part entière, très vaste, mais on peut classer ses méthodes en deux catégories [55]: Calibrage classique et le calibrage automatique :

3-1 Calibrage classique

Le principe de base est de photographier une mire (*Figure I-3*) 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 [43]. Il est évident qu'il est nécessaire d'utiliser plus de correspondances que possible afin d'améliorer la précision. Les données utilisées (*extraites des images produites par l'appareil*) sont en général des points. Mais certaines méthodes utilisent aussi des lignes ou des ellipses [26].

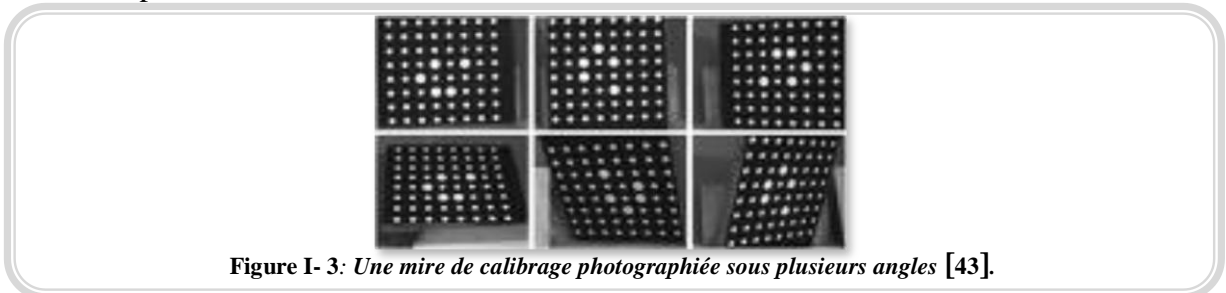


Figure I- 3: Une mire de calibrage photographiée sous plusieurs angles [43].

3-2 Calibrage automatique ou auto-calibrage

La calibration automatique [55] 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 épipolaire¹. Il est montré dans [55], qu'au moins trois vues différentes, gardant quand même des correspondances, sont nécessaires pour résoudre les équations qui mènent aux paramètres.

Utilisent des paires stéréo, les mouvements de caméras, ou des éléments du monde dont on connaît les propriétés projectives, comme par exemple des faisceaux de droites parallèles.

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

La synthèse d'images 3D à partir d'images s'affronte au fait 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. L'utilisation des 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, constitue tout l'art de la synthèse [49].

Une alternative consiste à synthétiser de nouvelles vues uniquement à partir des vues réelles de la scène de départ en simulant le déplacement de la caméra qui a pris les vues réelles. **L'apport d'une telle technique est d'éviter totalement l'étape de modélisation propre à tous les systèmes de synthèse d'images.** Ainsi la scène n'est pas définie par un modèle géométrique tridimensionnel, mais par des vues (*bidimensionnelles*).

Les techniques permettant de synthétiser de nouvelles vues à partir des vues réelles d'une scène sont appelées communément méthodes de modélisation et rendu basées images (*Images Based Modeling and Rendering: IBMR*). Certains auteurs font la distinction entre les méthodes de modélisation et les méthodes de rendu. En général, le terme de « **Rendu basé images** » (*Images-Based Rendering : IBR*) est utilisé pour décrire les techniques qui produisent un nouvel ensemble d'images représentant des vues virtuelles de cet environnement à partir d'un ensemble d'images d'un environnement (*Figure I-4*).

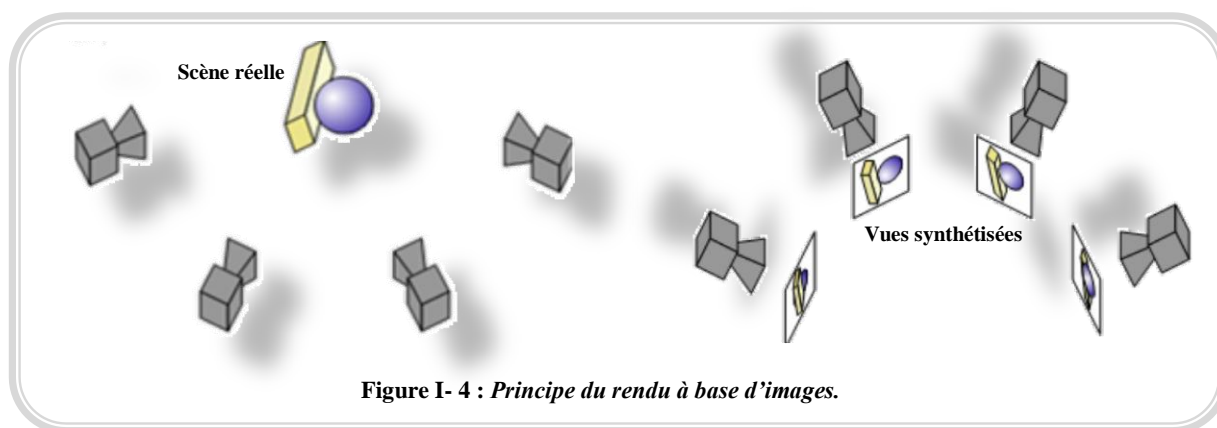
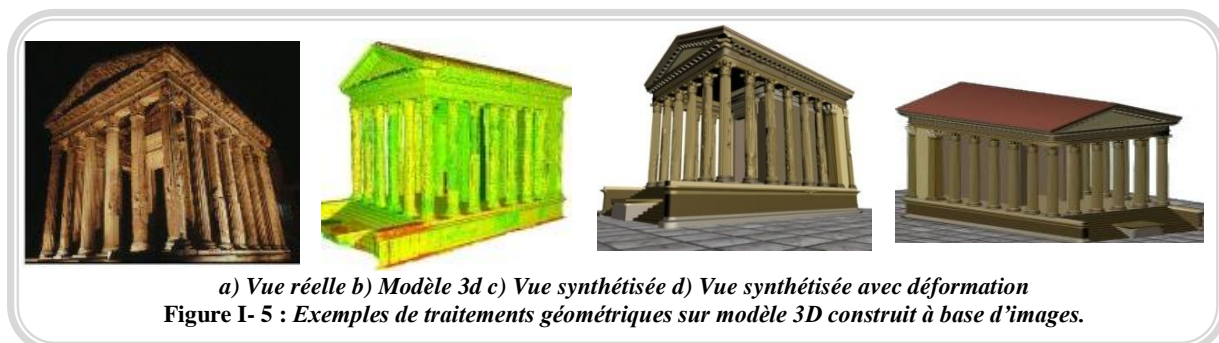


Figure I- 4 : Principe du rendu à base d'images.

¹ Liens entre des points de deux images prises à partir de différentes positions.

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, en simulant le déplacement de la caméra qui a pris les photographies (*si on dispose d'un modèle 3D de la scène*) on peut synthétiser de nouvelles vues depuis n'importe quels points de vue. On peut même appliquer des traitements géométriques plus généraux (*déformations, simplifications, augmentations ou des incrustations*) sur une représentation 3D d'un objet, en la manipulant par des outils standards, comme des logiciels de synthèse d'images, ou des éditeurs de modèles.

5. Méthodes de rendu/modélisation à base d'images (IBMR)

Pour synthétiser de nouvelles vues, 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 pixel¹, paramètres des caméras*) d'une image.

On peut distinguer ces techniques à partir de la façon dont les informations sont utilisées. Une ou plusieurs images sont utilisées, déformées et combinées selon différentes approches (*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 pour la synthèse de nouvelles vues à partir d'images réelles [74]:

- Les techniques fondées exclusivement sur le **rendu par les images** [14]. Ces techniques tentent de générer des vues de synthèse à partir d'un ensemble d'images originales. Elles utilisent, interpolent ou déforment directement l'ensemble des vues originales pour inférer une nouvelle vue ; mais n'estiment pas la vraie structure 3D.
- Les techniques basées images / géométrie [35] 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, excepter le cas du 3D scanning² qui essaie de retrouver complètement la structure 3D sous-jacente.
- Les techniques hybrides [57] combinant les techniques de **modélisation standard avec des vues de la scène réelle**, soit pour tirer parti des avantages des différentes approches, soit pour accélérer certaines techniques de rendu existantes.

¹ Par **profondeur de pixel**, on désigne le nombre de bits d'information fournis pour chaque point de l'écran.

² L'utilisation d'un scanner 3D qui analyse un objet du monde réel ou de l'environnement pour recueillir des données sur sa forme ou son apparence (*couleur par exemple*).

Nous nous contentons, pour notre travail, de décrire quelques méthodes basées uniquement sur les images.

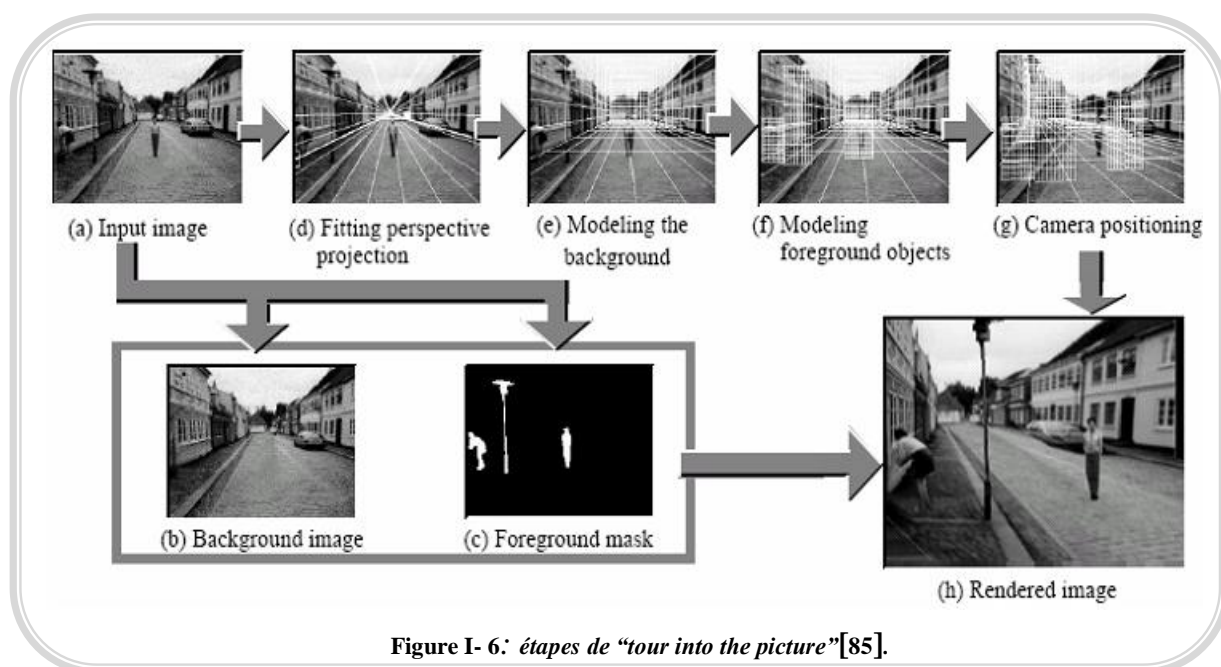
6. Techniques de rendu purement à base d'images

6-1 Imposteurs

Le but des imposteurs [83] est d'obtenir une grande complexité visuelle au moindre coût, c'est une classe de techniques qui effectuent un pré-rendu hors-ligne d'une partie de la scène. Cette partie sera plaquée (*entant que texture*) par la suite sur la scène. Ainsi, à la place de l'objet complexe est affichée son image.

Les plus simples des imposteurs sont les « Panneaux d'affichage » (*billboards*). Leur principe consiste à utiliser un plan texturé (*représentant un objet complexe*) et de le placer dans la scène. Typiquement, ils sont employés pour représenter des objets très difficiles à modéliser explicitement (*tels que des nuages ou des arbres*). On distingue ceux toujours parallèles à l'écran, ceux toujours orientés vers l'observateur et ceux qui sont de simples plans fixes dans l'espace de la scène.

Les panneaux d'affichage modélisent les objets au premier plan de l'image et sont déplacés lors de l'animation, accompagnés d'une légère déformation de l'image afin de donner l'impression de pénétration dans l'image (*tour into the picture*) [85].



6-2 Forme à partir de X (Shape from X)

Cette catégorie de techniques permet de reconstituer la forme des surfaces par différents procédés.

6-2-1 Forme à partir de contour (Shape from contour)

Le principe (*du Shape from contour*) [67] est de reconstituer les normales des surfaces visibles à partir de leurs contours. Il est alors possible de reproduire la surface à partir d'un champ de normales.

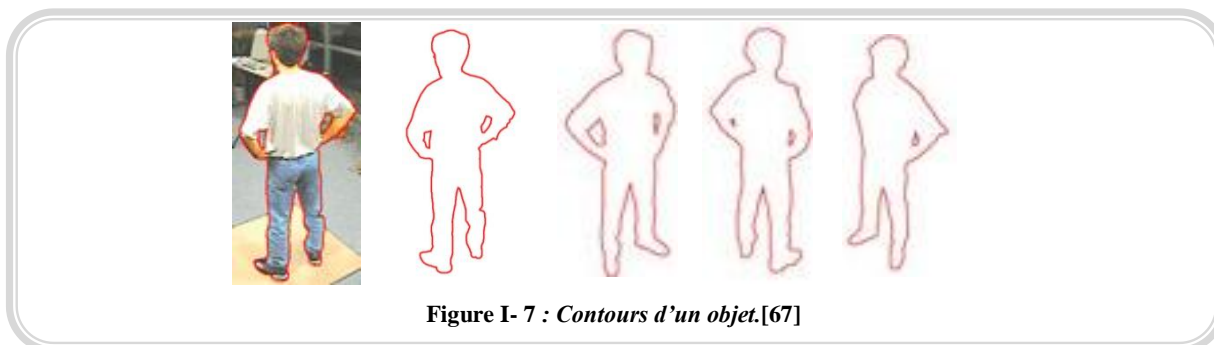


Figure I- 7 : Contours d'un objet.[67]

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.

6-2-2 Forme à partir de l'ombrage (Shape from shading)

Cette catégorie de techniques [68] utilise des hypothèses fortes sur la nature des objets ou les conditions d'illumination pour reconstituer la forme des surfaces.

L'intensité lumineuse que réfléchit un objet, peut être utilisée pour calculer sa structure tridimensionnelle. Si on connaît ; les positions de la caméra et de la source lumineuse, les lois de réflectance de l'objet, il est possible de calculer la forme de l'objet.

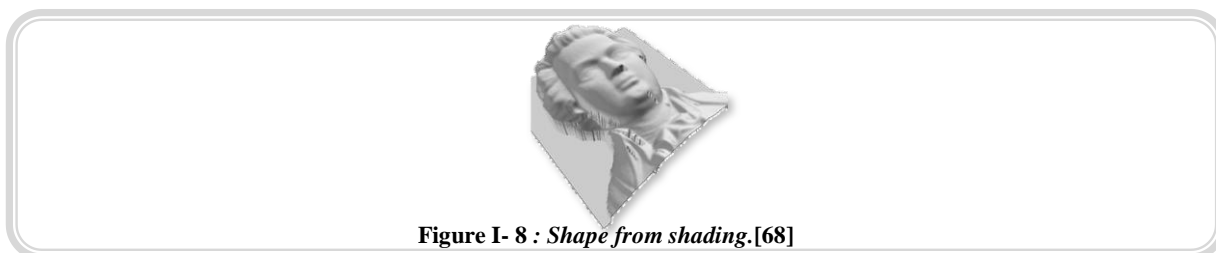


Figure I- 8 : Shape from shading.[68]

6-2-3 Forme à partir de silhouette (Shape from silhouette)

L'**enveloppe visuelle** [3] est l'intersection des cônes des silhouettes¹ d'entrée. Le volume qui résulte de l'intersection est une limite approximative de la forme de l'objet. Plus le nombre de silhouettes utilisées est grand, 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. Cela 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.

¹ Une **silhouette** est l'information de contour qui permet de définir une sorte d'inéquation en disant que l'objet se trouve exactement à l'intérieur d'un volume.

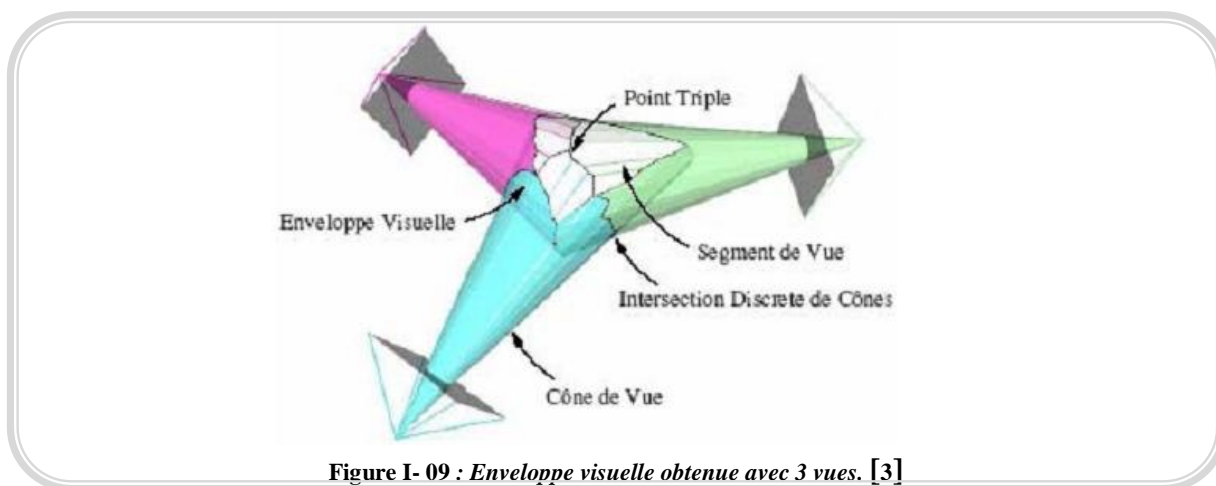


Figure I- 09 : Enveloppe visuelle obtenue avec 3 vues. [3]

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. Il existe beaucoup d'algorithmes fournissant des approximations de l'enveloppe visuelle.

6-2-4 Forme à partir de texture (Shape from texture)

La texture peut permettre de séparer certains objets du fond [1]. Pour ce qui est de la reconstruction 3D, les variations de texture sur l'image peuvent donner une bonne indication du relief, mais seulement dans le cas où cette texture est homogène.

Il est nécessaire, dans une image réelle, de procéder à une segmentation des éléments de texture pour pouvoir avoir accès à leur tailles et espacements. Les imprécisions liées à cette étape sont souvent génératrices d'erreurs.



Figure I- 10 : Extraction de la forme du tissu à partir de la texture.[1]

6-3 Morphing/Interpolation

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

6-3-1 Le morphing

Le morphing [54] 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 initiale et une finale*), sur lesquelles on établit des points de contrôle en correspondance 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*¹, *splines*²) entre les éléments de contrôle les plus proches.

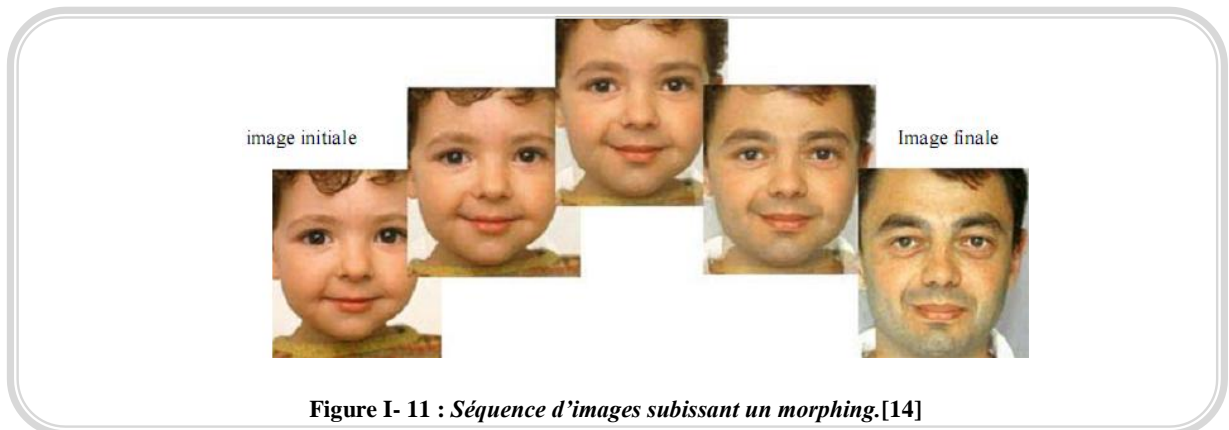


Figure I- 11 : Séquence d'images subissant un morphing.[14]

6-3-2 Interpolation de point de vue (View Interpolation)

Le principe consiste à synthétiser de nouvelles vues à partir d'images de référence, pour lesquelles des correspondances permettant de déplacer les points d'une image à l'autre sont établies [71]. 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.

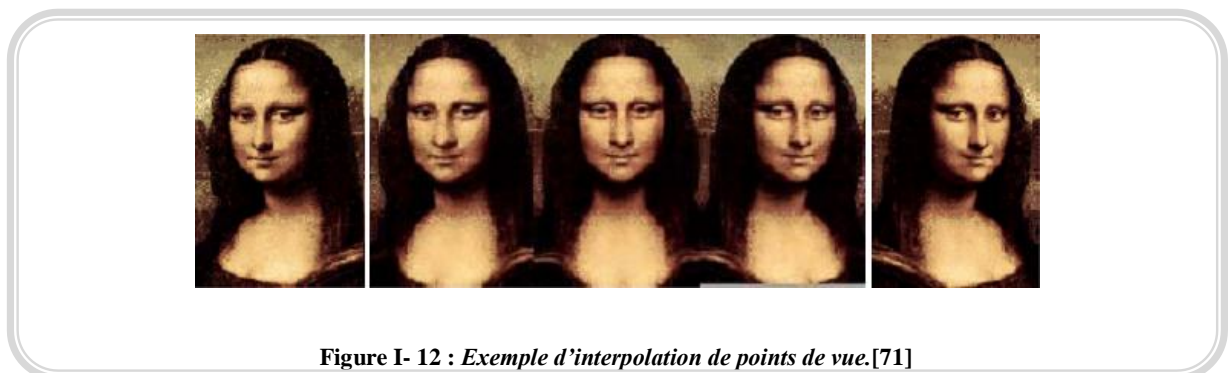


Figure I- 12 : Exemple d'interpolation de points de vue.[71]

6-3-3 Déformation de point de vue (View Morphing)

La déformation de point de vue [79] consiste à pré-déformer les images de référence avant de les interpoler linéairement. La pré-déformation 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 re-déformée pour donner l'image finale (*post-warping*).

¹ L'interpolation **bilinéaire** est une extension de l'interpolation linéaire, elle consiste à réaliser une interpolation linéaire dans un sens, puis de nouveau dans l'autre sens.

² Une Spline est une fonction définie par morceaux par des polynômes.

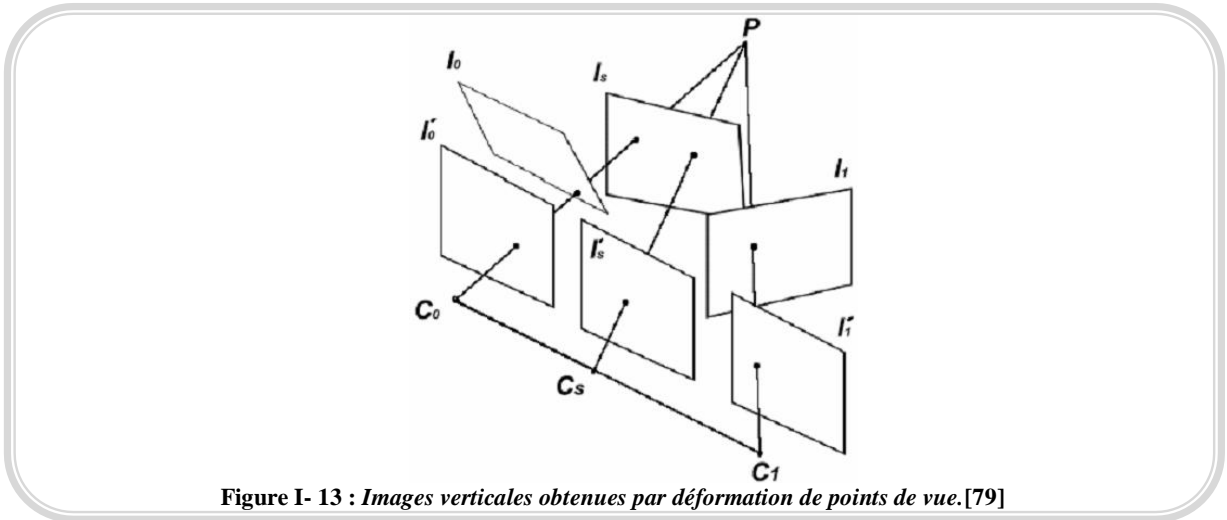


Figure I- 13 : Images verticales obtenues par déformation de points de vue.[79]

6-4 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 [40]. Cet espace est dénommé light-field (*champ lumineux*) [47]. 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.

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.

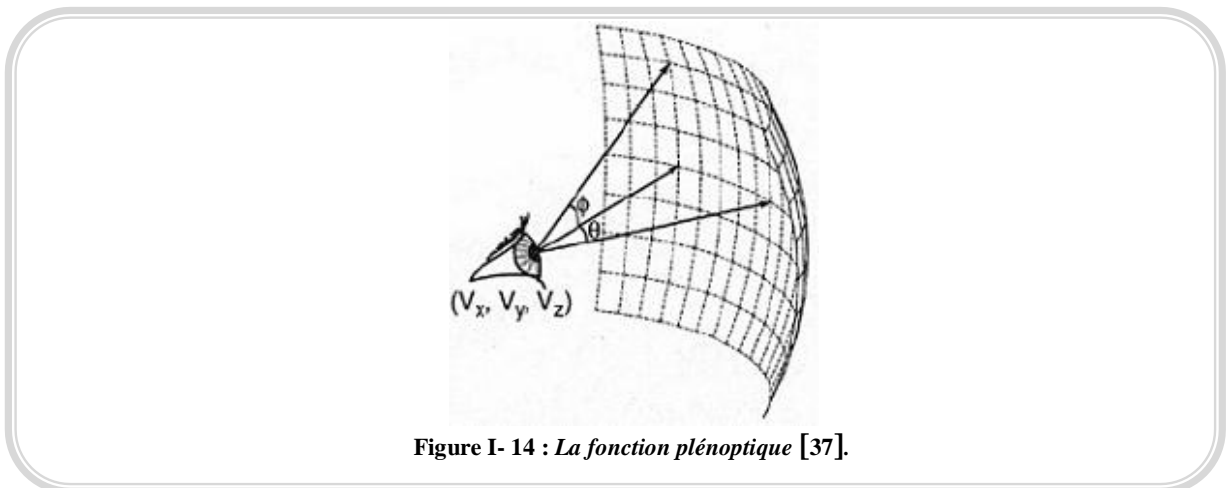


Figure I- 14 : La fonction plénoptique [37].

6-4-1 Fonction plénoptique

La fonction plénoptique [37] 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 (*éclairage constant*), la fonction se réduit à 5 dimensions :

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

Malgré ces simplifications, la quantité de données nécessaire pour décrire une telle fonction est tellement grande que les méthodes décrites 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 et Lumigraphes [72] 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 [73] peuvent simuler la fonction plénoptique. Si le point de vue est fixé, la fonction plénoptique peut alors être représentée par un panorama 2D [13].

7. Conclusion

Nous avons passé en revue les méthodes de synthèse d'images à partir d'images les plus connues dans la littérature ; illustrées par des exemples très significatifs. Ces méthodes sont diverses et dépendent souvent des moyens mis en œuvre.

Nous allons présenter dans le chapitre suivant, le principe de construction de l'enveloppe visuelle d'un objet, quelques approches de mise en correspondance et finalement la stéréovision.

Chapitre III

Enveloppe visuelle, Appareillement et stéréovision

1.	INTRODUCTION	18
2.	L'ENVELOPPE VISUELLE	18
3.	LA MISE EN CORRESPONDANCE	18
3-1	METHODES DE MISE EN CORRESPONDANCE	19
3-1-1	<i>Méthodes locales d'appariement</i>	20
3.1.1.1	Mesures de similarité.....	20
3.1.1.2	Forces et lacunes des méthodes locales	22
3-1-2	<i>Méthodes globales d'appariement</i>	23
3.1.2.1	Programmation Dynamique.....	23
3.1.2.2	Max-flow /min-cut.....	24
4.	LA STEREOVISION	26
4-1	SYSTEME DE STEREOVISION	27
4-2	GEOMETRIE EIPOLAIRE ET RELATION ENTRE DEUX IMAGES	28
4-2-1	<i>La contrainte epipolaire</i>	28
4-2-2	<i>La matrice fondamentale</i>	29
4-2-3	<i>La matrice essentielle</i>	30
5.	CORRESPONDANCE STEREO	31
5-1	CHOIX DES PRIMITIVES STEREOSCOPIQUES.....	32
5-2	CONTRAINTES DE LA MISE EN CORRESPONDANCE	32
5-2-1	<i>Contraintes géométriques</i>	32
5.2.1.1	Contrainte epipolaire	32
5.2.1.2	Contrainte de limites de disparité	33
5.2.1.3	Contrainte d'ordre.....	33
5.2.1.4	Contrainte d'unicité.....	34
5-2-2	<i>Contraintes figurales</i>	34
5.2.2.1	Disparité locale constante.....	34
5.2.2.2	Continuité figurale.....	34
6.	CONCLUSION	35

1. Introduction

Dans cette partie nous présentons, en premier lieu, la technique de l'enveloppe visuelle qui sera utilisée dans notre travail pour restreindre la zone de recherche de correspondants. Dans la section 3, nous exposerons quelques techniques de mise en correspondance. Les notions de base de la stéréovision seront détaillées dans la section 4 avant de conclure ce chapitre.

2. L'enveloppe visuelle

Supposons que l'on dispose de plusieurs silhouettes d'un même objet correspondant aux points de vue de différentes caméras. Ces silhouettes sont utilisées pour reconstruire une forme approximative de ce dernier.

Le principe est de mettre un objet devant un fond connu, de prendre une séquence d'images depuis des points de vue entourant cet objet et ensuite de traiter chacune de ces images afin d'y extraire le fond de la forme. Grâce aux différents points de vue, on englobe l'objet dans un volume appelé enveloppe visuelle.



L'enveloppe visuelle 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 on utilise de silhouettes ; plus l'enveloppe visuelle convergera vers un volume plus serré et sera plus proche de la forme de l'objet réel.

Il s'agit d'approches surfaciques [27] si l'on vise à reconstruire la surface de l'enveloppe visuelle en fournissant des points isolés ou un maillage. Et si on s'intéresse au volume délimité par l'enveloppe visuelle, et qu'on se base sur des discrétisations de l'espace ; on parle dans ce cas d'approches volumiques [62].

Les approches surfaciques peuvent être précises comparées aux approches volumiques, 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 des algorithmes et des calculs impliquant les cônes de vue, dont le lieu d'intersection est mal défini.

3. La mise en correspondance

La mise en correspondance [45] est le premier pas vers la reconstruction. En effet ; pour la reconstruction à partir d'images, 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.

La mise en correspondance de deux images stéréoscopiques consiste à établir une relation point à point entre celles-ci. Elle nécessite une représentation condensée des images pour

réduire les calculs et permettre d'appliquer des critères de similarité [28]. 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 appairer des points, en revanche cette correspondance peut s'avérer relativement difficile à réaliser d'un point de vue purement algorithmique.

Les deux types d'objets principalement utilisés sont les contours d'objets et les régions homogènes [53].

- Les *contours*¹ représentent des zones de discontinuité lumineuse représentées sous forme de suite de points connexes, de segments de droites ou de courbes. Les critères de ressemblance sont basés sur la forme, ou sur la longueur et l'orientation de segments qui les composent.
- Les *régions homogènes* sont caractérisées par leurs propriétés d'homogénéité généralement basées sur la luminance. Les méthodes de segmentation en régions [61] sont classées en deux catégories: les méthodes de division (*top down*) qui réalisent une partition de l'espace des luminances et les méthodes de croissance de régions (*bottom up*) qui utilisent l'information de luminance et les relations spatiales entre les points de l'image. Les critères de ressemblance sont basés sur la luminance et les caractéristiques inertielles des régions.

Pour initialiser le processus de mise en correspondance de deux images stéréoscopiques, dans le cas de systèmes de prise de vue non calibrés, il faut donc extraire un certain nombre d'indices visuels et les appairer afin d'obtenir des couples de points appelés *amers* qui sont les images homologues d'un même point de l'espace [52].

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 [7].

3-1 Méthodes de mise en correspondance

La mise en correspondance est un domaine vaste, où deux approches différentes se confrontent en traitement d'images [53].

- 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 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 une information "dense" par la mise en correspondance de la réflectance de tous les éléments (*points*) de la scène.

¹ Limite externe de la surface d'un corps, surtout en parlant des objets arrondis. Ligne de séparation des éléments d'une image.

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.

Les méthodes de fusion d'images stéréoscopiques sont basées sur:

- l'élargissement de la correspondance par recouvrement de deux régions homologues et l'utilisation de la cohérence des relations d'adjacence [30],
- la mise en correspondance des droites épipolaires homologues par programmation dynamique [31] en calculant le chemin de coût minimum dans un graphe de luminance,
- le déplacement d'une fenêtre de corrélation le long de droites épipolaires homologues. Cette méthode consiste à rechercher pour une zone de l'image de référence un maximum de corrélation sur une fenêtre mobile dans l'image homologue. Le maximum de ressemblance correspond à une corrélation maximale entre les pixels des deux images.

Toutes ces méthodes sont très sensibles aux variations de radiométrie (*ou variation de l'intensité lumineuse entre les deux images pour un même point*), et demandent des temps de calcul très importants.

3-1-1 Méthodes locales d'appariement

Les méthodes présentées dans cette section utilisent des primitives pour évaluer les appariements [28]. 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.

3.1.1.1 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 $I_i(u, v)$ 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 p_i la moyenne de l'intensité dans la fenêtre de l'image i , le coefficient de corrélation normalisée 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 -1 et 0*) 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 II-2*).

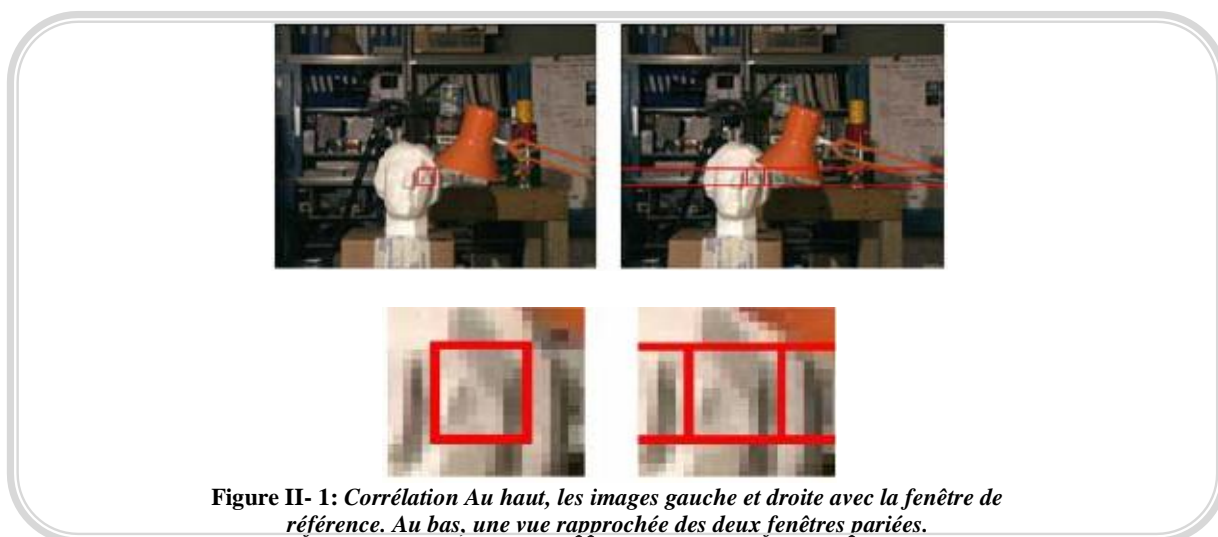


Figure II- 1: *Corrélation Au haut, les images gauche et droite avec la fenêtre de référence. Au bas, une vue rapprochée des deux fenêtres pariées.*

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 [88] 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.

Une seconde approche tente plutôt d’estimer le mouvement entre les deux images de la paire stéréo pour en déduire une mesure de disparité [13]. 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 (E_t). 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 E_t . 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 [31]. De plus, le fait d'utiliser uniquement la composante horizontale du gradient peut engendrer des imprécisions.

3.1.1.2 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 performant moins des zones de texture uniforme. Prenons par exemple le cas extrême où les deux images représentent un mur blanc mat 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 [32]. 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-3*. On observe qu'une section cache certaines parties de la scène 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 des techniques, consiste à détecter les occultations pour ensuite effectuer une interpolation dans les zones de disparités inconnues en fonction des disparités voisines.

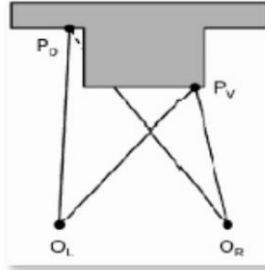


Figure II- 2: Exemple d'occultation.

3-1-2 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 appairer. 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.

3.1.2.1 Programmation Dynamique

Une première approche utilise la programmation dynamique pour résoudre un problème de minimisation sur une ligne complète de l'image [31]. 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.

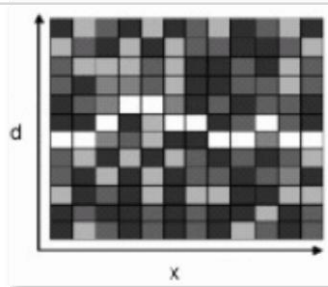


Figure II- 3: Exemple de chemin minimal (en blanc) dans l'espace (x, d) . Les niveaux de gris représentent les valeurs de dissimilarité.

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 k_P 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-5) dont la somme des coûts d'appariement est minimisée. On cherche donc à minimiser la fonction suivante:

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

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, d'où le nom, pour des questions de performance.

Contrairement aux méthodes locales, cette approche assure une valeur de disparité en tout point de l'image de référence, parfois au détriment de la précision de celles-ci.

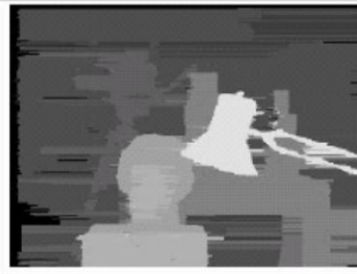


Figure II- 4: Exemple de résultat obtenu par la méthode de programmation dynamique. On remarque les traits horizontaux irréguliers entre les lignes.

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-5). 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*).

3.1.2.2 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-6). 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.

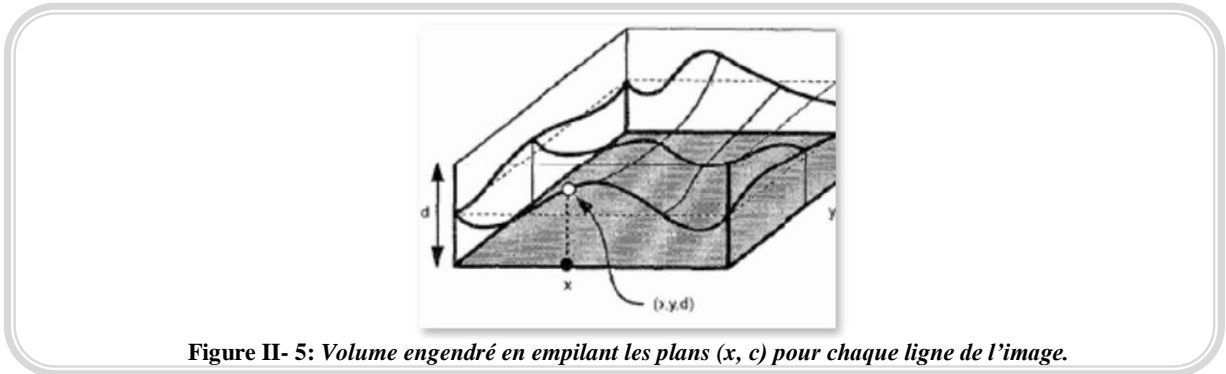


Figure II- 5: Volume engendré en empilant les plans (x, c) pour chaque ligne de l'image.

Il est donc proposé [75][76] de le formuler en tant que calcul d'un flux maximal à travers un volume présenté sous forme de graphe (*Figure II-7*). 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 à rectifier les images : il suffit de calibrer les caméras par rapport au référentiel de la scène.

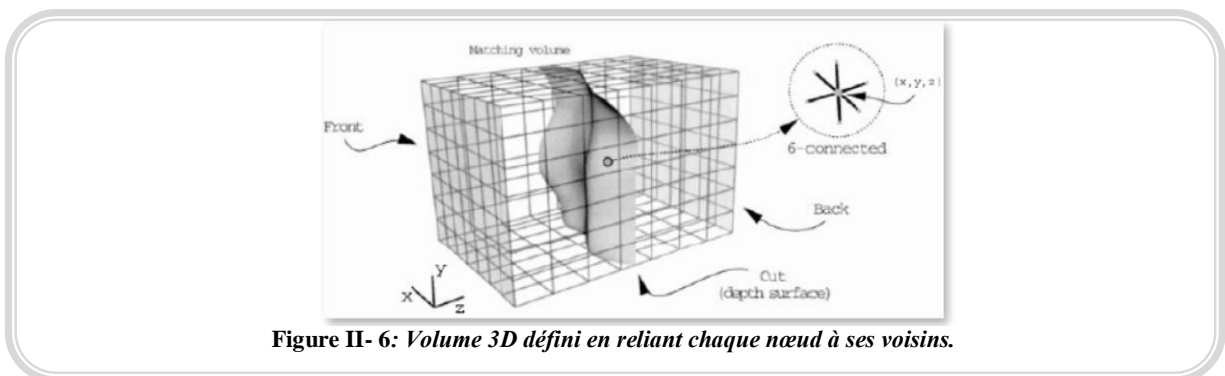


Figure II- 6: Volume 3D défini en reliant chaque nœud à ses voisins.

On ajoute une source s et un drain t au volume (*Figure II-8*). Le flux quitte la source, traverse le volume et rejoint le drain, à l'intérieur du graphe, chaque nœud 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 nœuds du plan $z = z_{max}$. Le volume est donc défini par l'ensemble des nœuds V :

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

Où s est la source, t le drain et V^* est l'ensemble des autres nœuds:

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

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.

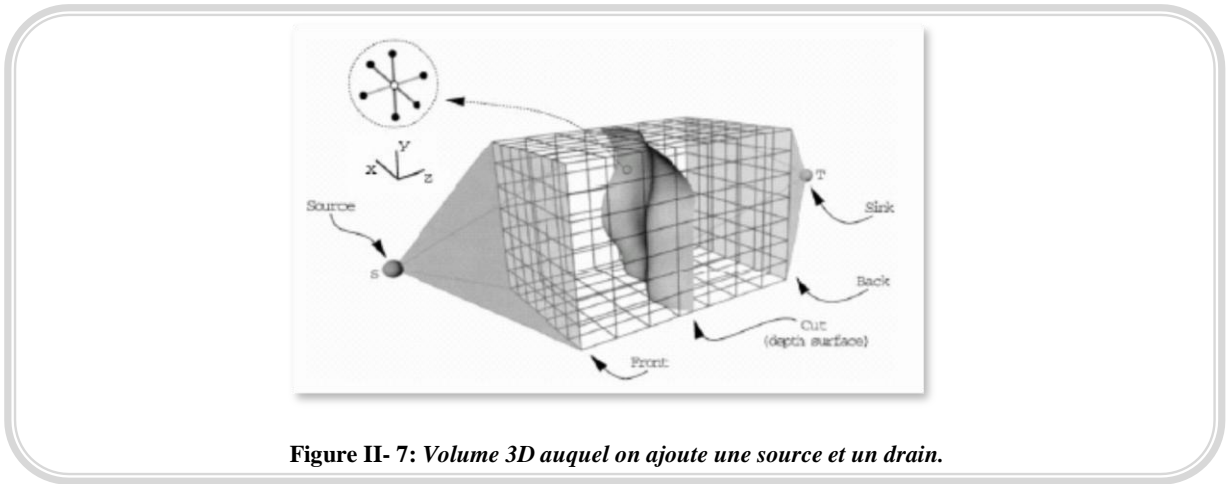


Figure II- 7: Volume 3D auquel on ajoute une source et un drain.

4. La stéréovision

En vision artificielle¹, le monde perçu est un univers en deux dimensions. En effet, une caméra cause la perte de l'information de profondeur puisque elle effectue une opération de projection perspective (*transforme le monde tridimensionnel de la scène en une représentation bidimensionnelle*). Etant donné un point physique de la scène et si la transformation de projection est connue, il est donc possible de connaître précisément la position de sa projection dans l'image. En revanche, il existe une infinité de points de la scène portés par une droite qui vérifie la transformation inverse pour un point donné de l'image.

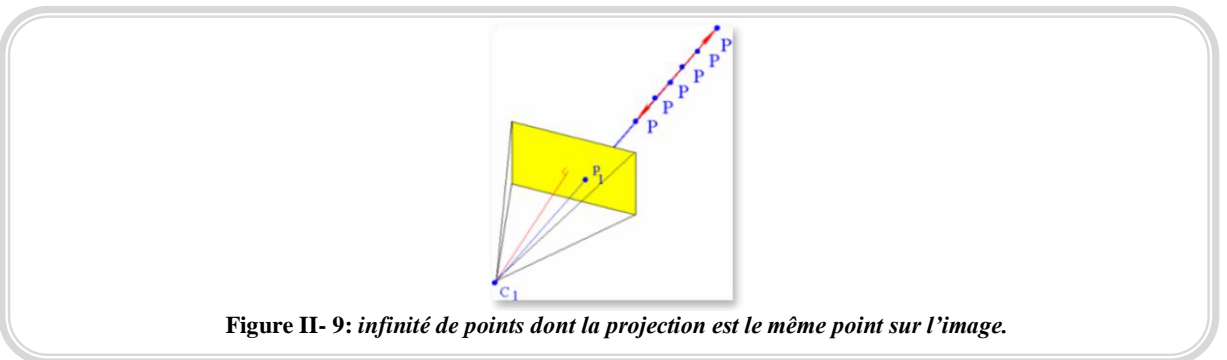


Figure II- 9: infinité de points dont la projection est le même point sur l'image.

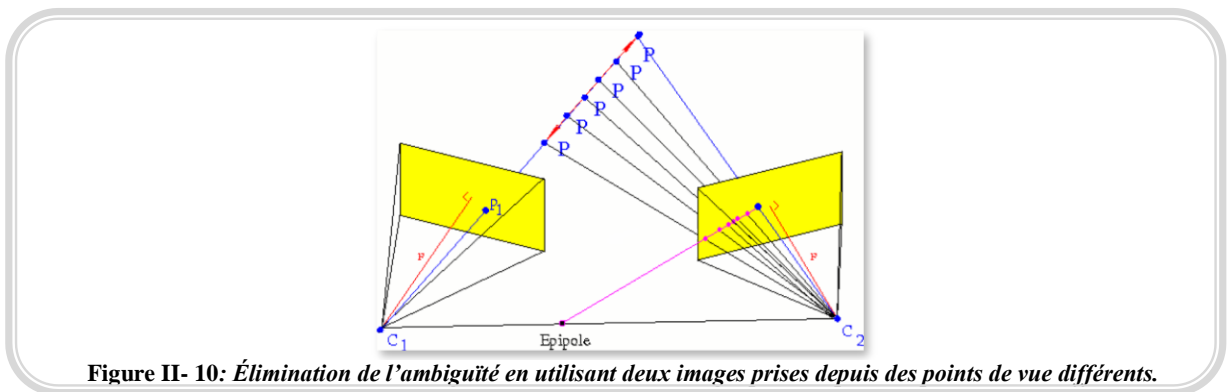
Comment, dans ces conditions, retrouver la dimension perdue? Pour déterminer la coordonnée tridimensionnelle manquante ; un complément d'information est nécessaire. La

¹ La vision par ordinateur (aussi appelée **vision artificielle**), est une branche de l'intelligence artificielle dont le but est de permettre à une machine de comprendre ce qu'elle «voit » lorsqu'on la connecte à une ou plusieurs caméras.

solution envisagée, se trouve dans **la combinaison de plusieurs images prises de points de vue différents** [77].

Alors des points de différentes images sont mis en correspondance grâce à des algorithmes appropriés, et c'est cette partie du traitement qui pose le plus de problèmes [28]. Ces algorithmes de stéréo-correspondance sont très sensibles au bruit¹, aux déformations perspectives et aux éclairages, ce qui les rend moins fiables.

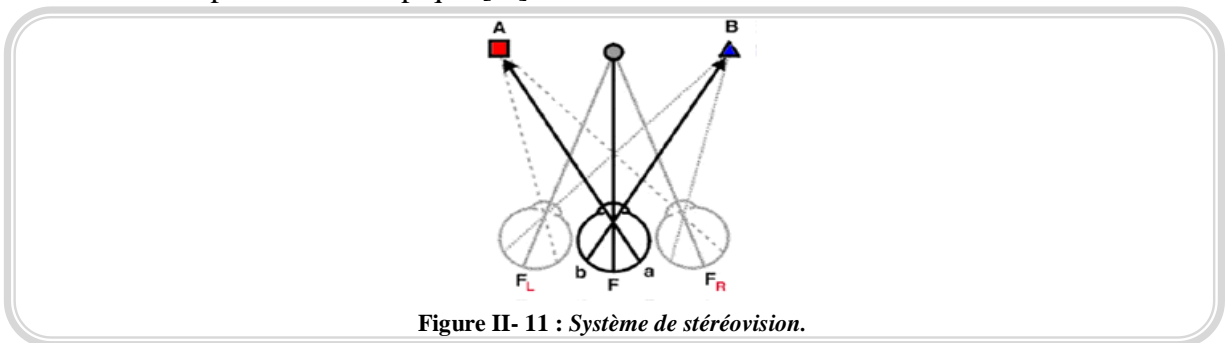
De nombreuses recherches sont encore menées pour améliorer ces méthodes de correspondance. La stéréovision s'intéresse à l'interprétation de deux vues distinctes de la scène afin de résoudre l'ambiguïté de la profondeur. Il s'agit de calculer les coordonnées 3D d'un point à partir de ses deux projections dans les deux images d'origine (*en connaissant le modèle de projection² de chaque caméra et la relation spatiale entre elles*).



4-1 Système de stéréovision

La stéréovision vise à réaliser une tâche similaire à la vision humaine [18], à l'aide d'un ou plusieurs périphériques de capture d'image (*par exemple un appareil photo ou un caméscope numérique*).

On utilise deux caméras, vidéo ou CCD³, disposées comme les yeux humains, qui vont donner deux images d'une scène (*Figure II-11*). Ces images sont formées de pixels et constituent une paire stéréoscopique [88].



Les images de la paire prises par des caméras écartées, ne sont pas sans relation. Ainsi un décalage d'un certain nombre de pixels entre les deux images, sera observé. Ce décalage est

¹ Le **bruit** en photo ou vidéo numériques est un défaut parasite dégradant la qualité de l'image.

² Le **modèle géométrique** qui permet de décrire la formation de l'image.

³ **CCD** (*Charge-Coupled Device*) , soit « Appareil à Transfert de Charges ». Désigne un capteur qui convertit la lumière en charges électriques.

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 et l'information de profondeur inférée.

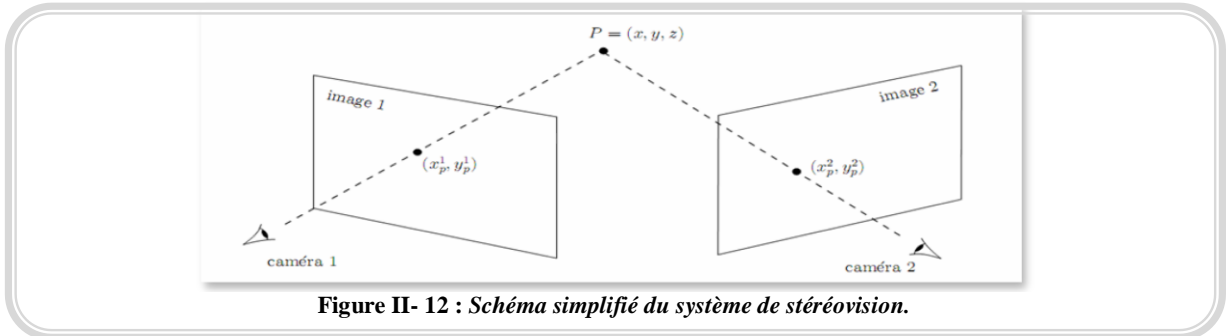


Figure II- 12 : Schéma simplifié du système de stéréovision.

Le principe de la stéréovision paraît relativement simple, mais 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, en d'autres termes ; **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.
- **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.**
- **Pour tout les couples de points mis en correspondance calculer la position 3D du point de la scène réelle par triangulation [69].**

4-2 Géométrie épipolaire et relation entre deux images

Considérons deux images A et B provenant de deux caméras perspectives observant la même scène. Soient M^A et M^B les matrices de projections correspondantes à ces deux images. Un point P de la scène se projette en $X^A = M^A P$ et $X^B = M^B P$.

On s'intéresse ici aux liens qui existent entre X^A et X^B . Ces liens sont caractérisés par la (*géométrie épipolaire*) [32].

4-2-1 La contrainte épipolaire

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

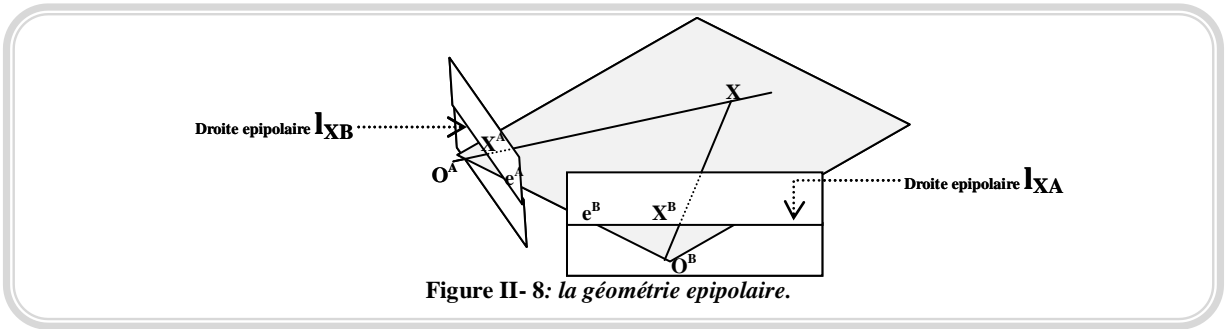


Figure II- 8: la géométrie épipolaire.

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.

$$e^A = M^A O^B$$

$$e^B = M^B O^A$$

Les épipoles jouent un rôle fondamental dans la vision stéréoscopique.

4-2-2 La matrice fondamentale

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

$$X^A F X^B = 0$$

Où F est une matrice 3x3

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

$$L^B = F X^A$$

$$L^A = F^t X^B$$

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 $X^A F X^B = 0$. Elle doit être vérifiée pour tous les couples (X_i^A, X_i^B) , on obtient donc l'équation suivante:

$$X_i^A \cdot X_i^B \cdot f_{11} + X_i^A \cdot X_i^B \cdot f_{12} + X_i^A \cdot X_i^B \cdot f_{13} + X_i^A \cdot X_i^B \cdot f_{21} + X_i^A \cdot X_i^B \cdot f_{22} + X_i^A \cdot X_i^B \cdot f_{23} + X_i^A \cdot X_i^B \cdot f_{31} + X_i^A \cdot X_i^B \cdot f_{32} + X_i^A \cdot X_i^B \cdot f_{33} = 0$$

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 $F = 0_{3 \times 3}$ mais elle n'est pas très intéressante. Pour éviter de trouver ce résultat, on va utiliser une astuce qui consiste à supposer f_{33} non nul. Ce n'est pas forcément très stable numériquement mais c'est pratique. Puisque F est invariant par facteur d'échelle, on pose $f_{33} = 1$. Le système à résoudre devient alors:

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

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 d'avoir 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 être 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:

$$F e^A = 0$$

$$F^t e^B = 0$$

4-2-3 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 X de l'image A, on peut exprimer la contrainte épipolaire comme suit:

$$X_c^A = {}^A C^{-1} X^A$$

Où A_rC est la matrice des paramètres intrinsèques de l'image A et X_c^A les coordonnées du point X dans le repère caméra. Par simple analogie avec ce qui a été vu précédemment, la contrainte épipolaire devient:

$$(X_c^B)^t \cdot {}^B_rC^t \cdot F \cdot {}^A_rC \cdot X_c^A = 0;$$

Soit :

$$(X_c^B)^t \cdot E \cdot X_c^A = 0 ;$$

Avec:

$$E = {}^B_rC^t \cdot F \cdot {}^A_rC$$

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 propres¹ non nulles de E ont même valeurs. La matrice essentielle peut être estimée, théoriquement, à partir de cinq correspondances.

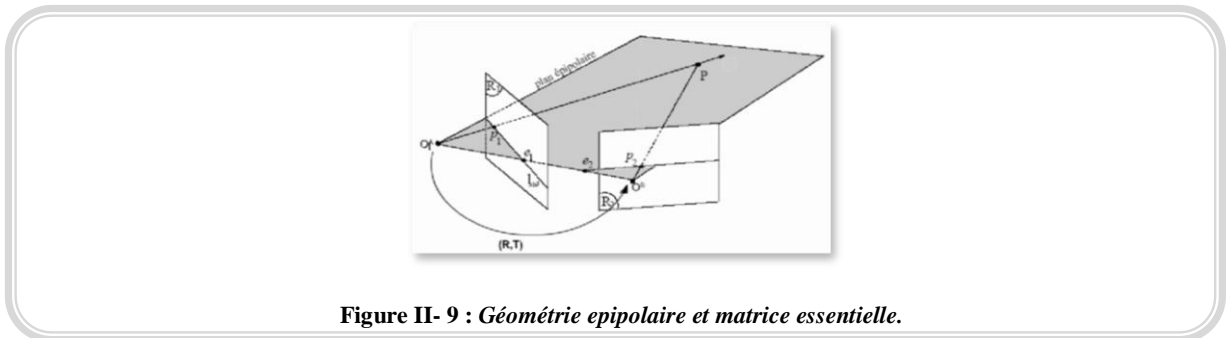


Figure II- 9 : Géométrie épipolaire et matrice essentielle.

5. Correspondance stéréo

Une des explications les plus communes de notre perception de la profondeur dans une scène est qu'on utilise la différence entre ce qu'on voit avec l'œil gauche et avec l'œil droit. Ce procéder est plus difficile à réaliser par des machines, et quelques problèmes doivent être surmontés :

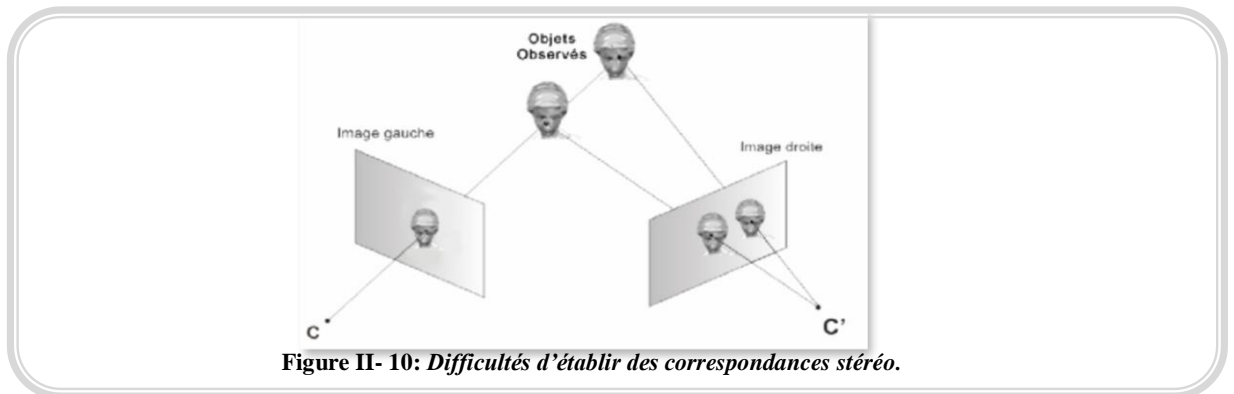


Figure II- 10: Difficultés d'établir des correspondances stéréo.

¹ Une valeur propre est associée à un vecteur propre est le facteur de modification de taille, c'est-à-dire le nombre par lequel il faut multiplier le vecteur pour obtenir son image.

5-1 Choix des primitives stéréoscopiques

Le choix des éléments (*ou primitives*) image à mettre en correspondance est crucial. Les techniques de segmentation d'images permettent l'extraction d'une gamme variée de primitives telles que [61]:

- 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. 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.

5-2 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 [28]. 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.

5-2-1 Contraintes géométriques

En plus des critères de ressemblance utilisables pour mettre en correspondance les objets de deux images, la structure du système de prise de vue stéréoscopique définit une contrainte géométrique qui réduit (*au maximum*) le processus de recherche pour un point de son correspondant.

5.2.1.1 Contrainte 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 $O^A P$. Les correspondants potentiels de X^A dans l'image de droite sont donc nécessairement situés sur la projection de la droite $O^A P$ dans l'image de droite.

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 épi-pôles soient à l'infini.

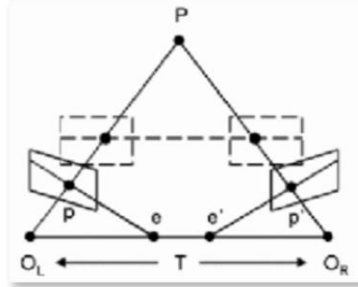


Figure II- 11: Rectification de la paire stéréo.

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.

5.2.1.2 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 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.

5.2.1.3 Contrainte d'ordre

Les contraintes épipolaires et de disparité permettent de réduire le nombre d'appariements possibles entre les primitives de l'image gauche et les primitives de l'image droite. 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.

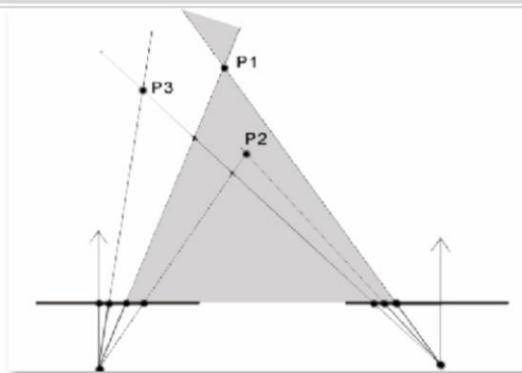
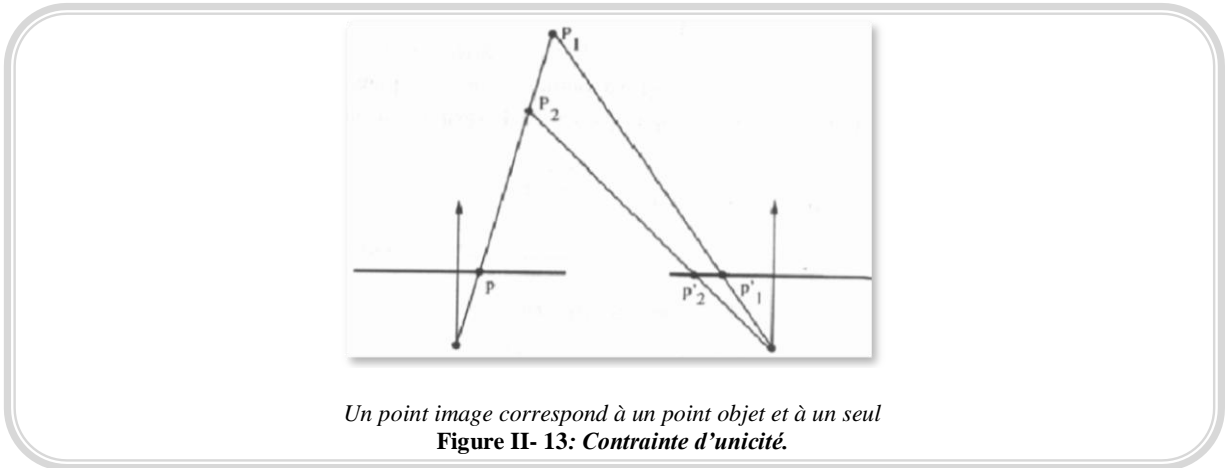


Figure II- 12: Contrainte d'ordre P3 et P1 ont le même ordre mais P1 et P2 sont inversés dans l'image de gauche.

5.2.1.4 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 où elle décrit un cas limite de la contrainte d'ordre. Lorsque deux points P_1 et P_2 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.



5-2-2 Contraintes figurales

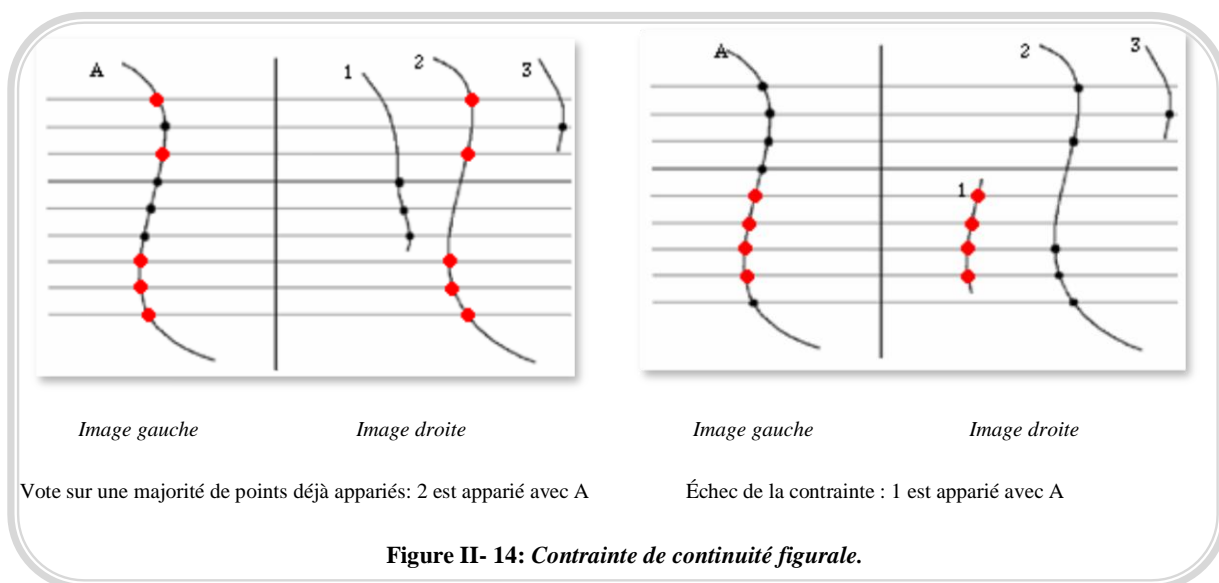
Les contraintes figurales sont définies comme telles car elles n'ont pas à l'instar des contraintes géométriques de base théorique, néanmoins, elles peuvent elles aussi imposer une contrainte d'ordre géométrique. On trouvera aussi dans cette catégorie toute contrainte spécifique à une primitive particulière.

5.2.2.1 Disparité locale constante

Lorsqu'une scène comporte des objets arrondis 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 ce fait, on limite par là même la variété des objets à reconstruire.

5.2.2.2 Continuité figurale

Cette contrainte limite les variations de disparité le long des contours. On évite ainsi autant que possible 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.



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.

6. Conclusion

Nous avons présenté dans ce chapitre le principe de construction de l'enveloppe visuelle, les approches de mise en correspondance qui constituent la phase cruciale et délicate de tout système de vision stéréoscopique qui est considéré comme la première étape de la reconstruction 3D à partir d'images.

Nous proposons dans le chapitre suivant une classification multicritères des différentes méthodes de reconstruction 3D.

Chapitre III

Classification multicritères des méthodes de reconstruction 3D

1.	INTRODUCTION	36
2.	CLASSIFICATION MULTICRITERES DES METHODES DE RECONSTRUCTION 3D	36
2-1	LE DEGRE DE GENERALITE DE LA MODELISATION 3D :	37
2-1-1	<i>Reconstruction projective</i>	37
2-1-2	<i>Reconstruction affine</i>	37
2-1-3	<i>Reconstruction métrique</i>	37
2-1-4	<i>Reconstruction euclidienne</i>	37
2-2	LE NIVEAU D'AUTOMATICITE DU SYSTEME	38
2-2-1	<i>La reconstruction automatique</i>	38
2-2-2	<i>La reconstruction semi-automatique</i>	38
2-2-3	<i>La reconstruction manuelle</i>	39
2-3	LE TYPE D'APPROCHE UTILISEE	39
2-3-1	<i>Approche descendante pixellaire</i>	39
2-3-2	<i>Approche ascendante</i>	40
2-3-3	<i>Les approches hybrides</i>	40
2-4	DEGRE DE PRECISION	41
2-4-1	<i>Exactitude du modèle géométrique</i>	41
2-4-2	<i>Exigences de description</i>	41
2-4-3	<i>L'apparence visuelle</i>	42
2-4-4	<i>Plusieurs aspects à la fois</i>	42
2-5	LA COMPLEXITE DES SCENES RECONSTRUITES	43
2-5-1	<i>Grandes scènes</i>	43
2-5-2	<i>Scènes moyennes</i>	43
2-5-3	<i>Petites scènes</i>	43
2-6	LA MOBILITE DES SCENES	44
2-6-1	<i>Scènes mobiles</i>	44
2-6-2	<i>Scènes statiques</i>	44
2-7	LA DISSOCIABILITE DES OBJETS	44
2-7-1	<i>Reconstruction mono-objet</i>	44
2-7-2	<i>Reconstruction multi-objet</i>	44
2-8	LE TYPE DE RESULTAT ENVISAGE	45
2-8-1	<i>Reconstruction 2D/2D</i>	45
2-8-2	<i>Reconstruction 2D/3D</i>	45
2-9	ACTIVES OU PASSIVES	45
2-9-1	<i>Les méthodes actives</i>	45
2-9-2	<i>Les méthodes passives</i>	48
2.9.2.1	Restitution à partir d'une seule image	48
2.9.2.2	Restitution à partir d'un couple d'images	49
2.9.2.3	Restitution à partir d'un bloc multi-image	49
3.	RECAPITULATIF	50
4.	CONCLUSION	51

1. Introduction

Chez l'homme ; le processus de reconstruction des formes est réalisé de manière plus ou moins consciente. Lorsqu'il s'agit de faire ce travail par une machine, cette tâche s'avère un problème assez délicat.

La vision par ordinateur est une discipline qui tente de simuler la vision humaine en établissant des modèles qui possèdent des propriétés proches de la perception visuelle humaine. C'est pour cela que la reconstruction 3D a donc suscité beaucoup de curiosité sur le plan scientifique et mathématique et a été traitée dans la littérature au cas par cas. De nombreuses méthodes ont été proposées mais elles sont souvent limitées à un domaine spécifique d'application.

Bien se positionner dans le domaine de la reconstruction 3D, avant d'aborder les problèmes techniques relatifs à une approche particulière, permet d'accélérer les recherches. Evitant ainsi de se confronter à des phénomènes secondaires vis-à-vis du but principal.

Cette étude nous a permis d'aborder le domaine de la reconstruction 3D, et à chaque phase, de nouvelles informations ont été acquises. Le reste de ce chapitre va être consacré pour proposer une classification multicritère des méthodes de reconstruction 3D.

2. Classification multicritères des méthodes de reconstruction 3D

Le regroupement des méthodes de reconstruction 3D dans un simple état de l'art est très difficile ; voir même impossible vu leur grand nombre. Nous allons alors essayé d'établir une classification selon des critères de différentes natures afin de regrouper toutes les méthodes existantes. Cette classification peut avoir un impact direct sur le choix du système qui sera proposé dans tout travail de reconstruction. Pour classer les méthodes de reconstruction 3D nous proposons les critères suivants:

- **Le degré de généralité de la modélisation 3D,**
- **Le niveau d'automatisme du système,**
- **Le type d'approche utilisé,**
- **Le degré de précision,**
- **Le type des données utilisées,**
- **La complexité des scènes reconstruites,**
- **La mobilité des scènes,**
- **La dissociabilité des objets,**
- **Le type de résultat envisagé,**
- **L'interaction avec la scène.**

Puisque ces techniques diffèrent en fonction des données d'entrée qu'elles peuvent traiter et du type de représentation géométrique qu'elles peuvent générer, cette étude est déterminante pour le choix du système qui sera proposé dans tout travail de reconstruction.

2-1 Le degré de généralité de la modélisation 3D :

Le modèle de surface utilisé dépend évidemment de l'application envisagée.

2-1-1 Reconstruction projective

Utilisée lorsque les coordonnées des points sont définies dans un repère projectif [66]. La reconstruction est alors définie à une transformation projective près. Une telle reconstruction ne contient pas d'informations critiques (*absence de notion d'angle, de longueur et de parallélisme*)[56][25].



Limage originale de la maison (à gauche) et une image du modèle après la suppression de l'escalier

Figure III- 2: L'image originale de la maison (à gauche) et une image du modèle.

2-1-2 Reconstruction affine

Utilisée lorsque les coordonnées des points sont définies dans un repère affine [66]. La reconstruction est alors définie à une transformation affine près. Une telle reconstruction contient des informations affines (*rapport de longueurs et parallélisme*)[56][36].



Figure III- 2 : Deux images réelles et une image du modèle 3D obtenu.

2-1-3 Reconstruction métrique

Utilisée lorsque les coordonnées des points sont définies dans un repère métrique [66]. La reconstruction est alors définie à une transformation métrique près. Une telle reconstruction contient des informations métriques (*Changement d'échelle*)[38][52].



Figure III- 3 : Résultat d'une reconstruction métrique.

2-1-4 Reconstruction euclidienne

Utilisée lorsque les coordonnées des points sont définies dans un repère euclidien [66]. Cette reconstruction est la plus complète. Elle contient toutes les informations que l'homme a l'habitude de manipuler (*notions de distances et d'angles*)[2][64][25].



Figure III- 4 : Exemple de reconstruction euclidienne.

2-2 Le niveau d'automatisme du système

2-2-1 La reconstruction automatique

Dans ces approches ; la reconstruction est faite automatiquement, l'utilisateur n'intervient donc pas durant le processus, son seul travail consiste à préparer la scène à reconstruire (*Positionnement des objets, mise en marche des appareils d'acquisition...etc.*) [10] [20] [78][81].

Actuellement, il n'est pas encore possible d'automatiser complètement le processus de reconstruction à partir d'images dont les points de vue sont très éloignés [19]. Le principe de base étant la triangulation, les algorithmes de détection automatique de caractéristiques et de correspondances entre images ont besoin de photographies proches (*cohérence spatiale*) pour que ces détections aient plus de chances de réussir (*ce qui rend cette méthode impossible à utiliser pour modéliser de grands environnements car le nombre d'images nécessaires serait trop grand*).

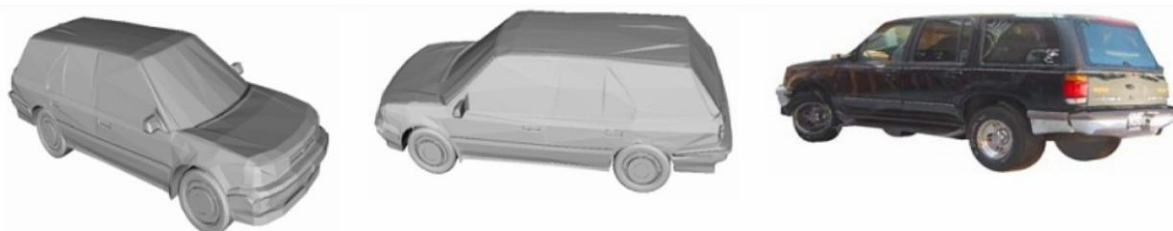
2-2-2 La reconstruction semi-automatique

L'approche assistée nécessite des interactions avec l'utilisateur. Mais ces actions, contrairement à la modélisation traditionnelle, sont en général peu nombreuses et simplifiées. Le but est de faire réaliser à l'homme des tâches difficiles à automatiser, et faisant appel à la connaissance qu'il a du monde réel (*préciser les arêtes du modèle, les contraintes géométriques, ...*), et à la machine des tâches plus contraignantes (*calculs, projections, ...*).

Dans ces méthodes, l'utilisateur a la possibilité de designer des points ou des droites sur une ou plusieurs images approximant ainsi le modèle 3D de l'objet à reconstruire [4][6] [38]. Il peut même avoir à sa disposition des modèles génériques qui seront augmentés jusqu'à l'obtention d'une bonne estimation du modèle 3D [34][41]; le reste du travail sera fait par le logiciel assistant.



(A gauche) Un modèle de base mis sur une image réelle,
(A droite) Le modèle rectifié selon les besoins,



(A gauche) Modèle obtenu.
(A droite) Modèle texturé.

Figure III- 5 : Reconstruction semi-automatique.

2-2-3 La reconstruction manuelle

L'utilisateur fait recours aux méthodes de synthèse classique, sauf que les objets à modéliser sont des objets réels ; il prend en charge tout le processus de synthèse et essaye de donner une approximation de l'objet réel (*mesures, couleurs, textures...etc.*). Cette approche est surtout utilisée en cas d'indisponibilité des informations pertinentes à la reconstruction 3D envisagée (*destruction de site par exemple*)[40].

2-3 Le type d'approche utilisée

Nous allons maintenant présenter trois méthodologies différentes permettant de reconstruire la surface d'objet de la scène. Ces trois méthodes se différencient par le type de primitive utilisée pour l'appariement des images ou pour la modélisation de la forme.

2-3-1 Approche descendante pixellaire

Elle cherche à appairer de manière massive l'ensemble des pixels/points de l'image sur une grille régulière, soit dans l'espace image, soit dans l'espace objet. Elle est bien adaptée pour caractériser la surface interne des objets.

Dans les approches "Top-Down" [70][33], des connaissances à priori sur les scènes sont introduites. On part ainsi d'un ensemble de modèles connus à l'avance pour alors choisir celui qui présente la meilleure cohérence avec les données. Ces approches sont généralement plus robustes que les approches "Bottom-Up", mais se trouvent limitées par la librairie de modèles définie à priori.



Figure III- 6 : Reconstruction descendante.

2-3-2 Approche ascendante

À base de maillages de primitives (*points 3D, segments ou courbes 3D, etc.*). Elle consiste à extraire des primitives 3D généralement éparses et à reconstruire la surface par un maillage (*e.g. triangulation, ... etc.*) s'appuyant sur les primitives extraites. Elle est bien adaptée pour caractériser les limites saillantes de surface des objets ou structures au sein des objets.

Les approches "bottom-up" [51] partent de primitives pour proposer la reconstruction finale, sans prendre en compte d'éventuelles connaissances à priori sur les scènes que l'on cherche à reconstruire.

Les approches "bottom-up" sont donc génériques et nécessitent des données de bonne qualité. Un inconvénient souvent inhérent à ce type d'approches est le manque de robustesse.

En effet, l'extraction de primitives est sujette aux problèmes de sur-détection et de sous-détection qui influence le résultat final. De plus, l'absence (*ou du moins le manque*) de connaissances à priori sur les scènes génère parfois des résultats peu réalistes à cause d'une combinatoire importante. Dans la majorité des processus "bottom-up", les primitives extraites sont des segments 3D ou/et des plans 3D.



Figure III- 7 : Reconstruction descendante.

2-3-3 Les approches hybrides

Consistent, dans un premier temps, à formuler des hypothèses à partir de primitives par une approche "bottom-up", puis dans un second temps, à vérifier ces hypothèses par une approche "top-down", et enfin ajuster des hypothèses de modèles géométriques ou morphologiques élémentaires dans les données. Cela suppose que l'on a une très bonne connaissance à priori du monde observé et que l'ensemble du monde peut être décrit avec la grammaire des formes.

Ces approches [42] peuvent être également composées ou mises en compétition pour un objet donné en fonction de l'adéquation du modèle sous-jacent à la réalité de l'objet. En effet, certains objets de scènes (*surfaces et limites de surfaces courbes*) ne peuvent être parfaitement décrits avec une grammaire de formes basiques, mais peuvent très bien l'être par une approche de type maillage de primitives qui est très polymorphe.

2-4 Degré de précision

Durant la génération de la maquette 3D d'une scène, les besoins d'une précision géométrique détaillée, la disponibilité de tous les détails, les performances d'affichage en temps réel et le photoréalisme sont aujourd'hui des aspects fondamentaux. Mais, au delà de la simple application d'un procédé technique, qu'il soit simple ou intégré, d'autres dimensions d'ordre méthodologique s'imposent pour la détermination d'une stratégie d'acquisition et de traitement des données spatiales.

Pour bien évaluer la validité d'une approche, il est nécessaire d'identifier les objectifs de représentation qui conduisent les différentes expériences. En effet, le choix d'un type de représentation véhicule les modalités de collecte de données et détermine leur traitement.

Dans cette optique, on peut citer quatre catégories d'approches principales :

2-4-1 Exactitude du modèle géométrique

Ces approches [19][48] s'appuient sur l'application d'une démarche standard de reconstruction 3D à partir d'un balayage laser 3D. Une attention particulière est donnée à l'évaluation de l'exactitude géométrique du modèle résultant de la phase de reconstruction. Toutes (*ou presque*) les informations acquises en phase de relevé sont exploitées pour la génération d'un modèle polyédrique décrivant les surfaces de façon la plus cohérente possible aux points saisis.

Parmi les problèmes communs à toutes les solutions commerciales employées. Une première remarque concerne la quantité des données à traiter. Les algorithmes utilisés ont de grosses difficultés pour le calcul de l'ensemble des surfaces décrivant l'objet entier. Un autre problème concerne la qualité des surfaces résultantes. Les artefacts présents sur la géométrie restituée (*présence de trous, discontinuités, etc.*) dépendent principalement de divers facteurs accidentels qui influencent la propagation du faisceau laser en phase d'acquisition. De plus, le faisceau laser au moment de l'impact avec la surface est réfléchi dans différentes directions qui dépendent de la rugosité superficielle locale : cela altère de façon considérable la qualité du maillage généré. Enfin le type de matériau impacté par le faisceau détermine l'intensité du signal de retour. Cela conduit parfois à une perte de données et une dégradation conséquente du nuage de points quand il n'y a pas de retour du signal (*white noise*).

La structuration de la maquette en plusieurs niveaux de résolution constitue une voie pour la conservation et le partage de la restitution géométrique d'un édifice. Les raisons techniques qui conduisent ce traitement sont en effet contraintes à l'exploitation des modèles 3D générés dans des systèmes de navigation temps réel en ligne.



(A gauche) le modèle 3D obtenu, (A droite) modèle 3D texturée avec une partie en zoom
Figure III- 8 : Exactitude du modèle géométrique.

2-4-2 Exigences de description

Ce genre d'approches [21] est caractérisé par une stratégie d'acquisition et de traitement conçues en cohérence avec un objectif de représentation. En général il s'agit d'acquérir un

ensemble de points pertinents pour restituer les seules informations nécessaires à la description d'un aspect spécifique (*restitution des caractères constructifs, analyse de la forme architecturale, lecture de l'état de conservation des matériaux, etc.*).

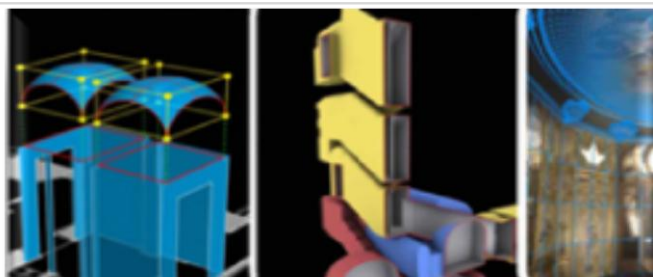


Figure III- 9 : Exemples de reconstructions descriptives.

2-4-3 L'apparence visuelle

Plusieurs techniques [21] de synthèse d'images sont disponibles aujourd'hui pour simuler l'apparence visuelle des formes basées sur le calcul de l'interaction entre sources lumineuses, caractéristiques des matériaux et géométrie des objets.

Ces approches se concentrent sur la restitution de l'aspect visuel des surfaces qui composent l'objet d'étude. Le niveau de description géométrique exploitée pour le rendu est très variable : il peut être représenté par les simples points d'un nuage, par un réseau approximatif de polygones, ou par des mailles polyédriques très fines.



Figure III- 10 : Restitution de l'apparence visuelle d'un édifice.

2-4-4 Plusieurs aspects à la fois

Ces approches sont basées sur l'utilisation complémentaire ou sur l'intégration de différents procédés techniques. Plusieurs méthodes d'acquisition sont exploitées pour la restitution des différents aspects qui participent à la représentation exhaustive de l'objet étudié. En phase d'acquisition, la saisie d'informations métriques et photométriques est prise en compte en fonction des différents niveaux de complexité géométrique que l'objet d'étude présente et par rapport aux contraintes spécifiques du système de représentation choisi.

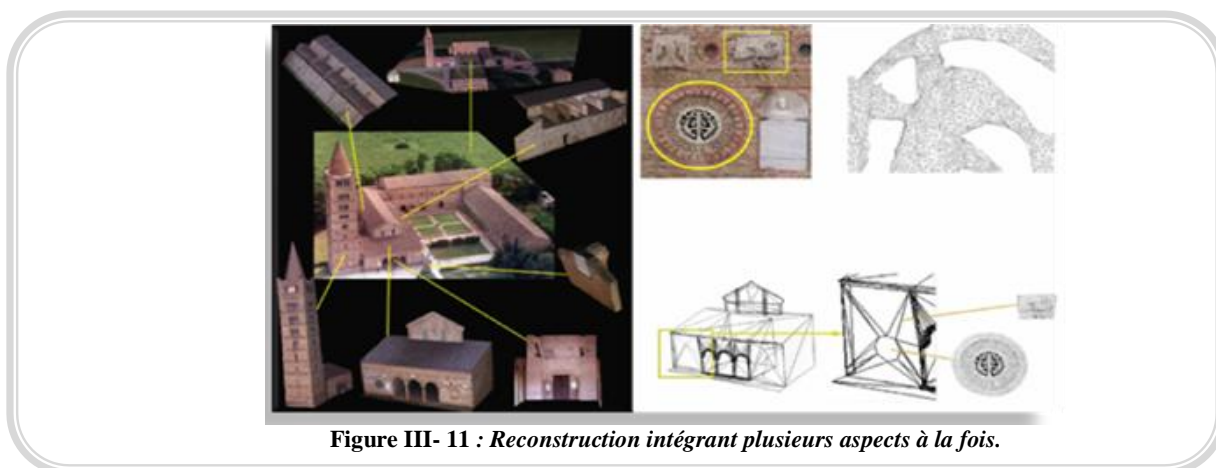


Figure III- 11 : Reconstruction intégrant plusieurs aspects à la fois.

2-5 La complexité des scènes reconstruites

2-5-1 Grandes scènes

Les techniques utilisées pour reconstruire de grandes scènes [29] (*des villes par exemple*) sont très sophistiquées, elles font intervenir les satellites, les avions, les données GPS...etc. Et ont leurs propres algorithmes de détection de forme et de plaquage de textures.

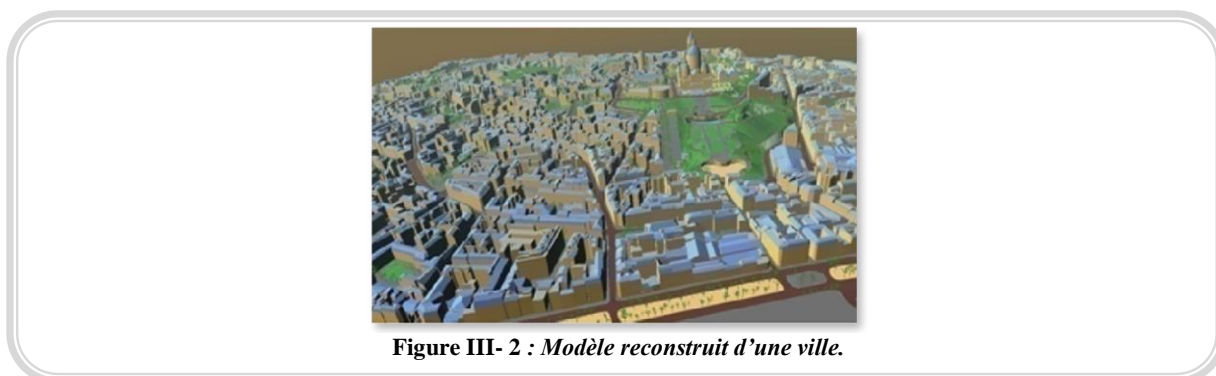


Figure III- 2 : Modèle reconstruit d'une ville.

2-5-2 Scènes moyennes

Reconstruire des scènes d'ordre moyen [29], tel qu'un bâtiment, nécessite l'utilisation de techniques facilitant l'acquisition des données nécessaires à la reconstruction, il est par exemple impossible de faire tourner un temple et d'y prendre des photos par une caméra fixe.

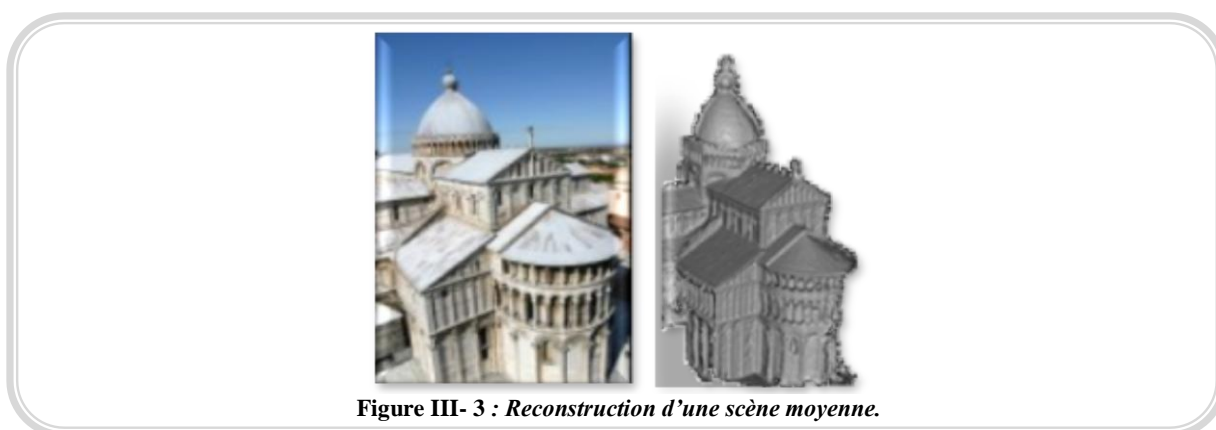


Figure III- 3 : Reconstruction d'une scène moyenne.

2-5-3 Petites scènes

Ici, l'acquisition des projections 2D de l'objet à étudier n'est pas difficile [12] et les temps requis pour rassembler les données essentielles pour la reconstruction sont relativement plus petits.



Figure III- 4 : Objets de petites tailles.

2-6 La mobilité des scènes

2-6-1 Scènes mobiles

Pour l'analyse de mouvement [44][60], quand le modèle de mouvement est connu, il faut estimer ces paramètres et dans ce cas les problèmes portent sur la détermination de la zone de recherche. Quand le modèle est inconnu, il faut estimer le vecteur de vitesse en chaque point et on est confronté au problème de la robustesse des algorithmes.

2-6-2 Scènes statiques

La reconstruction de scènes fixes [48] ne fait pas intervenir la notion de temps. Les études menées dans ce cadre visent essentiellement à améliorer les algorithmes ou d'obtenir des résultats plus réalistes que possibles.

2-7 La dissociabilité des objets

2-7-1 Reconstruction mono-objet

Tient à reconstruire la scène une fois pour toute sans la décomposer en sous modèles [41]. La scène est constituée d'un ensemble d'objets indissociables ou simplement d'un seul objet.

2-7-2 Reconstruction multi-objet

Une reconstruction graduelle consiste à assembler des modèles reconstruits pour aboutir à la reconstruction d'une scène complexe, en procédant par la reconstruction séparée des objets constituant la scène et par la suite, en rassemblant les modèles générés dans un seul modèle [34].

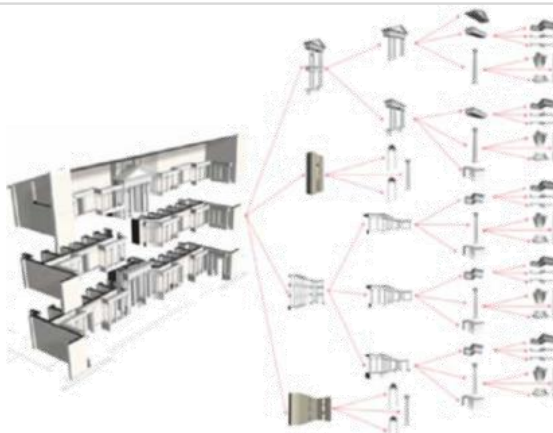


Figure III- 5 : Regroupement des éléments reconstruits.

2-8 Le type de résultat envisagé

2-8-1 Reconstruction 2D/2D

En général, ces techniques produisent un nouvel ensemble d'images représentant des vues virtuelles à partir d'un ensemble d'images d'un environnement; sans passer par le modèle 3D [50].



Figure III- 6 : Vue panoramique.

2-8-2 Reconstruction 2D/3D

Lorsque l'on génère un modèle 3D, on peut synthétiser de nouvelles vues depuis n'importe quels points de vue en simulant le déplacement de la caméra [86]. 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 standards, comme des logiciels de synthèse d'images, ou des éditeurs de modèles.



Scène réelle



Modèle 3D



Modèle 3D texturé



Modèle 3D rectifié

Figure III- 7 : Reconstruction 2D/3D.

2-9 Actives ou passives

Parmi les nombreuses méthodes mises au point pour calculer les informations 3D, on peut distinguer deux grandes classes [48]:

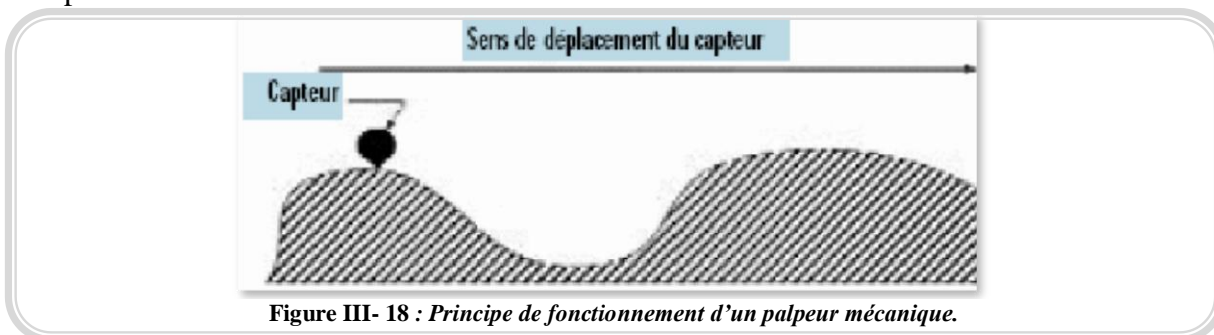
2-9-1 Les méthodes actives

Les méthodes actives [48] qui utilisent des capteurs ayant une action directe sur l'environnement étudié: systèmes acoustiques (*relatifs au son*), tactiles (*qui réagissent au toucher*), télémètres, ... etc. Ils acquièrent généralement l'information spatiale.

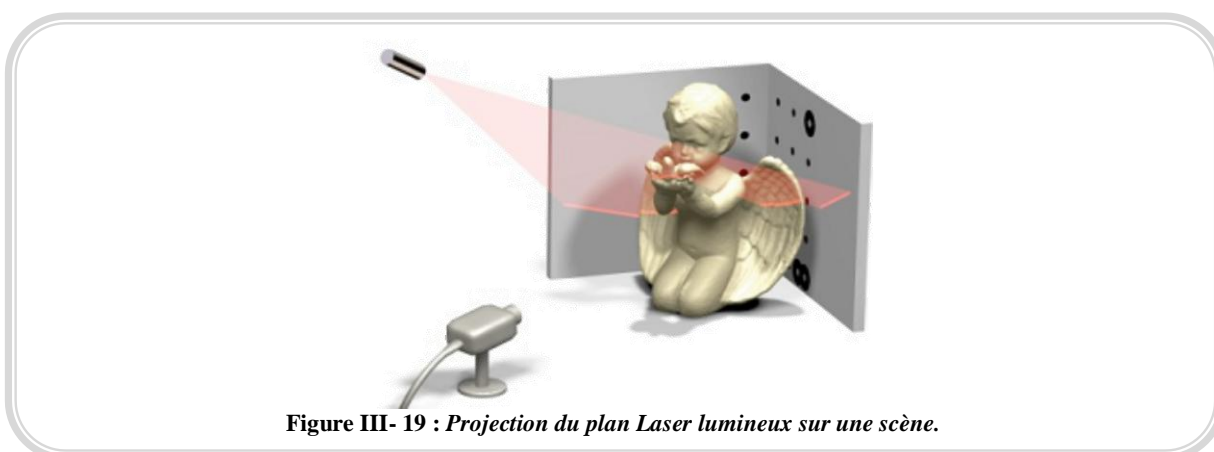
Les Machines à Mesurer Tridimensionnelles (*M.M.T.*) [84] (*équipements de "copie mécanique"*), sont à l'origine de l'acquisition d'objets en vue d'une reconstruction 3D. Ces outils s'appuient sur la technologie alors bien maîtrisée de la détermination de points dans l'espace. En effet, les coordonnées 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.

Amener le palpeur sur chaque point voulu et mesurer les données 3D, nécessite énormément de temps. Les palpeurs modernes sont connectés à des ordinateurs et les mesures sont faites automatiquement. 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.

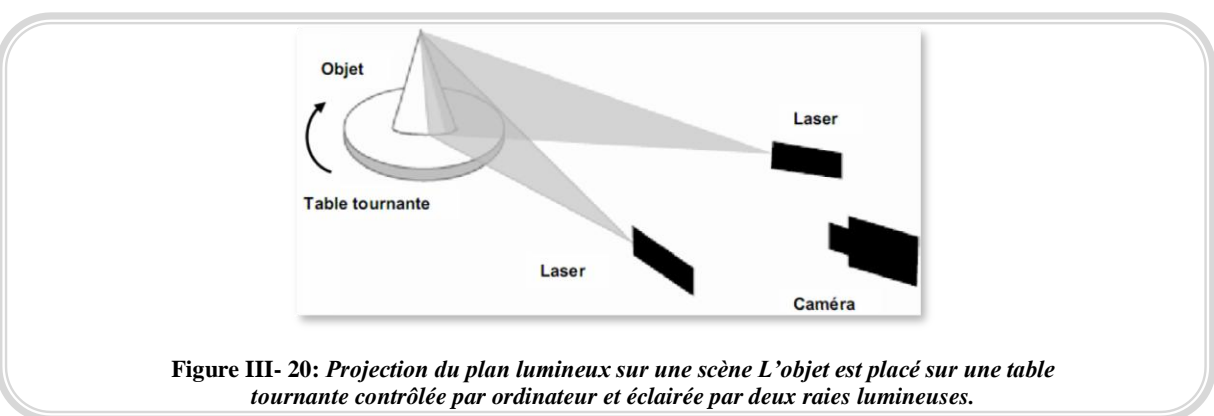
Le coût de ces appareils est très élevé surtout lorsqu'ils sont constitués de pièces mécaniques de haute précision, mais ils permettent de numériser tout type d'objet, indépendamment de leur forme ou de leur matière.



Les méthodes de télémétrie laser [48] 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. Le schéma de principe du télémètre laser est donné en *Figure III.19*.



Si on connaît avec précision la position de la source de lumière et le modèle de la caméra utilisée, il est alors possible de calculer les coordonnées tridimensionnelles des points de la scène ainsi éclairés par simple triangulation. 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 III.20*).



On peut également utiliser un éclairage contrôlé ou une lumière structurée illuminant la scène d'une manière particulière. Le principe 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. 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. On observera alors le résultat de l'éclairage de la scène par cette source particulière.

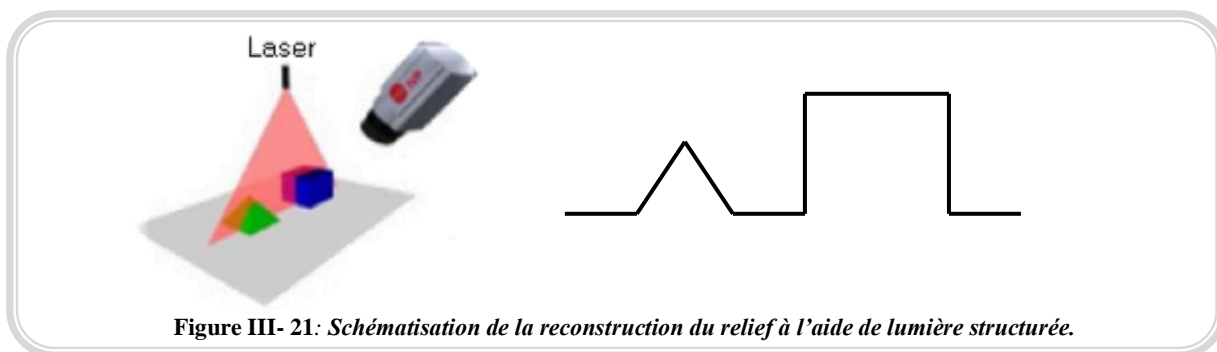


Figure III- 21: Schématisation de la reconstruction du relief à l'aide de lumière structurée.

Les méthodes de reconstruction basées sur la lumière structurée englobent de nombreuses techniques dont, en particulier, la triangulation laser, 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*).

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.

Le processus de triangulation est donc ici un peu plus complexe puisqu'il faut appairer la grille apparente (*celle projetée sur la scène*) avec la grille théorique (*le motif structurant proprement dit*).

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 III- 22: Lumière structurée illuminant un visage.

Les méthodes actives fournissent des mesures précises et peuvent être 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 réfléchissantes.

Néanmoins, les méthodes actives sont très utilisées dans de nombreuses applications, et particulièrement dans le domaine de la robotique.

Il y-a 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, les ultrasons nocifs aux animaux, ...*) soit de la profondeur de la scène observée (*images satellites*).

2-9-2 Les méthodes passives

Les **méthodes passives** [20] 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.

La stéréovision est la méthode 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. La stéréovision représente le système minimal pour réaliser cette triangulation.

Il s'agit alors de déterminer quelles sont les informations pertinentes à extraire de ces images pour en reconstruire la géométrie de la scène. On est donc confronté directement à la complexité apparente des images. Ces méthodes diffèrent en fonction de la configuration de l'acquisition : mono, stéréo ou multi images.



Figure III- 23 : Reconstruction passive.

2.9.2.1 Restitution à partir d'une seule image

La restitution à partir d'une seule image [4][38] isolée de parties planes d'un objet s'appuie sur la mesure d'un certain nombre de point. Pour développer un redressement (*ou une rectification*), on considère, par approximation, que la surface est définie uniquement par des plans identifiés par un minimum de quatre points d'appui. L'homographie est l'équation projective qui établie la relation entre le plan de l'objet et le plan de l'image [5].

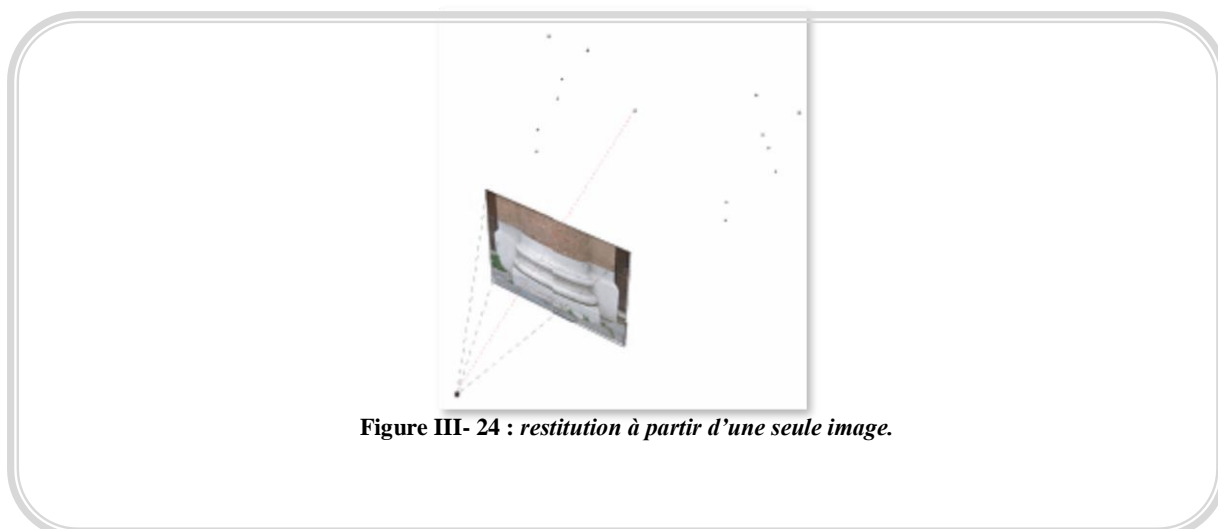


Figure III- 24 : restitution à partir d'une seule image.

2.9.2.2 Restitution à partir d'un couple d'images

La restitution à partir d'un couple d'images utilise une configuration stéréo [8][88]. En phase d'acquisition, les chambres (*avec une orientation interne connue et étalonnée et une distorsion négligeable de l'objectif*) sont presque parallèles entre elles. Dans les solutions de restitution numérique, l'observation stéréoscopique consiste en un système d'asservissement en temps réel qui agit sur l'un des clichés et qui prend en compte les paramètres d'orientation internes et externes des perspectives. Ce dispositif commande les déplacements en translation du deuxième cliché en fonction de la position observée sur le premier de façon à ce que les deux points observés soient toujours homologues. Ce mécanisme permet à l'opérateur de réaliser des représentations bidimensionnelles de façades, ou des modèles filaires et surfaciques en trois dimensions. Les formes arbitraires peuvent être reconstruites par des techniques de corrélation d'images automatiques. Ces techniques s'appuient sur l'analyse d'image et permettent d'extraire de façon automatique ou semi-automatique un ensemble de points détectés sur les variations chromatiques des surfaces.

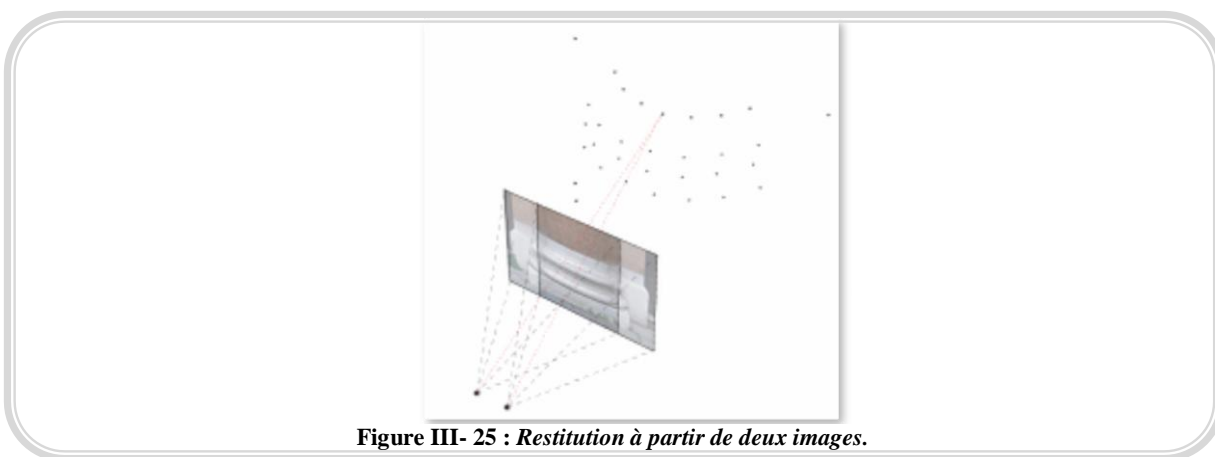


Figure III- 25 : Restitution à partir de deux images.

2.9.2.3 Restitution à partir d'un bloc multi-image

La restitution à partir d'un bloc multi-images [81][11][62][63] peut utiliser des photographies convergentes, horizontales, verticales ou obliques. La restitution se base sur l'ajustement des blocs d'images par la méthode des faisceaux (*conditions de colinéarité*) à partir d'un ensemble de correspondances (*points, lignes*) sélectionnées sur les photographies. Un des principaux avantages des méthodes basées sur ce principe est la possibilité d'exécuter un auto-étalonnage de la chambre directement à partir de photos de l'objet. En effet, une première solution de calibration déterminée, l'ajout de correspondances permet d'affiner le calcul des paramètres (*externes et internes*) par minimisation d'erreur.

Cette approche n'est donc pas limitée à des chambres métriques ou étalonnées ce qui la rend beaucoup plus flexible. Les résultats de la photogrammétrie multi-images peuvent être des modèles filaires en 3D ou des photomodèles (*représentations géométriques enrichies par les textures*).

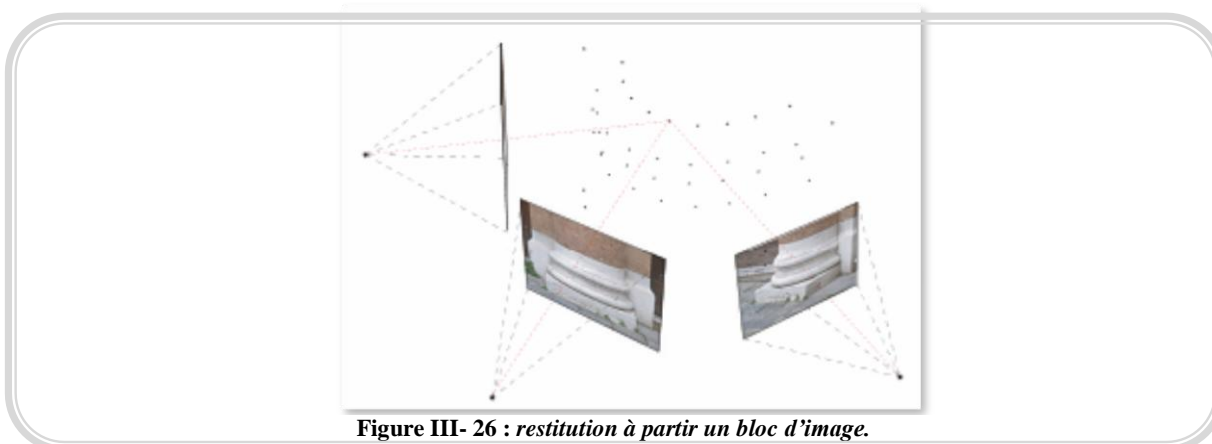


Figure III- 26 : restitution à partir un bloc d'image.

3. Récapitulatif

Le tableau suivant présente la classification multicritères proposée des méthodes de reconstruction 3D, il met en évidence les différents critères à prendre en charge pour tout travail de reconstruction 3D et permet d’avoir une idée des démarches et résultats envisagés:

Tableau III- 1 : Tableau présentant la classification multicritère proposée.

<i>Classement selon</i>	<i>Type de reconstruction</i>
Le degré de généralité de la modélisation 3D.	Projective
	Affine
	Métrique
	Euclidienne
Le niveau d’automaticité du système.	Automatique
	Semi-automatique
	Manuel
Le type d’approche utilisé.	Top-Down
	Bottom-Up
	Hybrides
Le degré de précision.	Exactitude du modèle géométrique
	Exigences de description
	Apparence visuelle
	Plusieurs aspects à la fois
Le type des données utilisées.	Optiques (mono-scopie, stéréoscopie, multi-scopie)
	LASER
	RADAR, rayons X, produits dérivés, etc.
La complexité des scènes reconstruites.	Grandes
	Moyennes
	Petites
La mobilité des scènes.	Mobiles
	Statiques
La dissociabilité des objets.	Mono-objet
	Multi-objets
Le type de résultat envisagé.	2D/2D
	2D/3D
L’interaction avec la scène.	Active
	Passive

4. Conclusion

Nous avons proposé, au cours de ce chapitre, une classification des méthodes disponibles de reconstruction 3D. Notre principal objectif a été de présenter les différents critères à prendre en charge pour tout travail de reconstruction 3D, sans avoir à affronter des aspects secondaires et imprévisibles par rapport au but initial.

En fait, cette classification, doit permettre aux personnes de bien choisir les techniques, les algorithmes, les outils et les moyens nécessaires pour aboutir à une reconstruction 3D. Elle permet aussi de prédire la qualité, la quantité et le type des résultats obtenus avant même d'aborder un sujet particulier ; minimisant ainsi le temps et les efforts consacrés à ce but.

Nous considérons que la bonne compréhension des critères de classification est primordiale pour éviter toutes confusions possibles. Il serait par exemple, impossible d'envisager une reconstruction 3D d'une ville en temps réelle (*Hypothèse contradictoire selon la classification présentée*).

Chapitre IV

Une méthode multi-
vue pour la
reconstruction 3D

1.	MOTIVATIONS	52
2.	APPROCHE MULTI-VUE POUR LA RECONSTRUCTION 3D	53
2-1	PREMIERE OPTIMISATION.....	53
2-2	DEUXIEME OPTIMISATION	54
3.	ETAPES DU PROCESSUS DE RECONSTRUCTION.....	57
3-1	SYSTEME DE PRISE DE VUES	57
3-2	RESTRICTION DE LA ZONE DE RECHERCHE DE CORRESPONDANTS	58
3-3	MISE EN CORRESPONDANCE	58
3-4	RECONSTRUCTION 3D.....	58
4.	ALGORITHME.....	59
5.	EXPERIMENTATION ET RESULTATS	60
5-1	IMAGES UTILISEES	60
5-2	PREPARATION DES SILHOUETTES	61
5-3	CONSTRUCTION DE L'ENVELOPPE VISUELLE	61
5-4	RECONSTRUCTION 3D.....	63
6.	AVANTAGES.....	65

1. Motivations

Après avoir étudié les méthodes de reconstruction 3D et en ayant proposé une classification, nous proposons dans ce chapitre une approche multi-vue pour la reconstruction 3D.

Les deux méthodes : enveloppe visuelle et stéréovision ont leurs inconvénients inhérents [3 article guelma]: L'enveloppe visuelle ne peut pas détecter les régions concaves, même si un grand nombre de silhouettes est utilisé. La stéréovision est instable et non fiable spécialement lors du processus de mise en correspondance, surtout pour les surfaces non texturées et les régions occluses [8 article guelma]. Cependant les deux méthodes sont bien complémentaires en nature et 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 [12][15][87].

La technique de l'enveloppe visuelle limite au maximum 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.

Dans ce travail nous traitons l'idée de la combinaison de l'enveloppe visuelle/stéréovision d'une façon différente, nous procédons différemment en ce qui concerne la combinaison des informations issues des deux méthodes pour la reconstruction de l'objet.

Nous proposons [22] deux optimisations de l'algorithme de mise en correspondance:

- La première consiste à utiliser les informations issues de la géométrie de l'enveloppe visuelle pour réduire la zone de recherche de correspondances.
- La seconde ; quand-à elle, vise à limiter la recherche par pixel sur un fragment de longueur d_{max_i} de la droite épipolaire pour la reconstruction stéréo. d_{max_i} n'est rien d'autre que la longueur égale à la disparité maximale (*calculable*) entre deux points des images de référence.

Ceci va nous permettre principalement de :

- Réduire le temps nécessaire pour reconstruire un objet.
- Obtenir des modèles de qualité optimale (*Eviter les correspondances en dehors de l'enveloppe visuelle*).
- Utiliser des photographies réelles (*Sans besoins à d'éventuelle segmentation*).

Dans ce qui suit, nous allons expliquer en détail l'approche de reconstruction 3D que nous proposons.

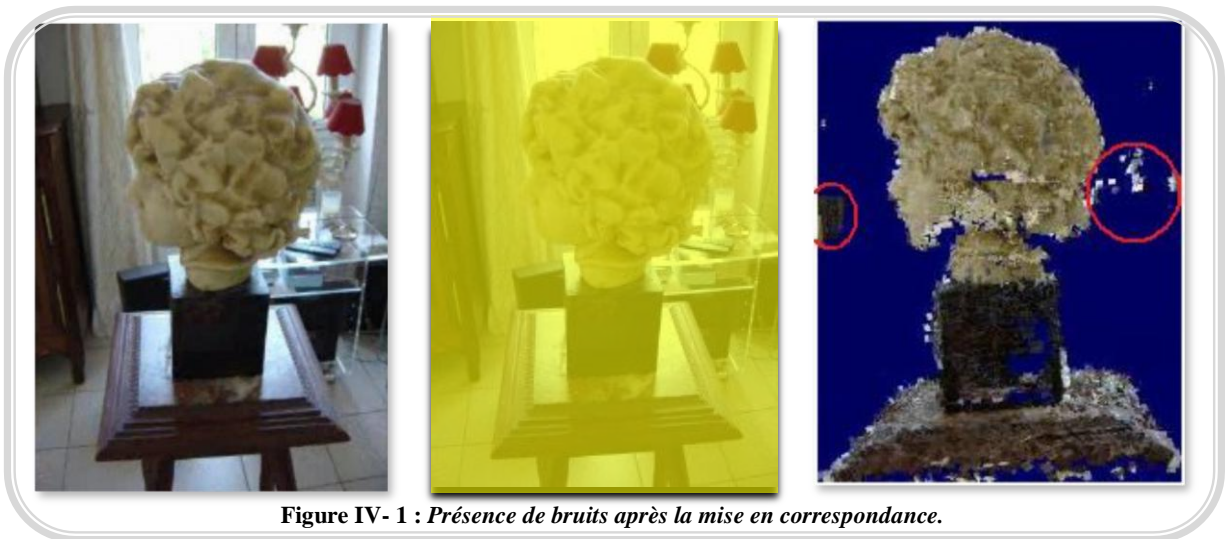
2. Approche multi-vue pour la reconstruction 3D

2-1 Première optimisation

La plupart des algorithmes de mise en correspondance utilisent la globalité de l'image lors des calculs, en d'autres termes, l'analyse se fait par un balayage (*vertical ou horizontal*) de tous les pixels de l'image. Cette solution présente deux inconvénients (*Figure IV-1*):

- Comme il est clair ; la zone de recherche est trop grande par rapport à l'objet effectif à reconstruire, ce qui augmente le temps de traitement,
- de plus, on constate bien les bruits (*Encerclés en rouge*) générés dans le résultat de mise en correspondance des points de l'objet à reconstruire, et qui nécessitent un traitement supplémentaire pour les éliminer.

La séquence d'images (*de gauche à droite*) représente respectivement l'objet à reconstruire, la zone de recherche (*en jaune*) et enfin le résultat de la mise en correspondance.



La méthode que nous proposons, utilise simplement l'enveloppe visuelle pour délimiter le champ de recherche, ce qui va éliminer les anomalies pouvant apparaître en dehors de la frontière extérieure du résultat et d'éviter les surcoûts des calculs supplémentaires.

Notre première optimisation consiste donc à limiter les calculs de correspondances aux points qui se trouvent exactement à l'intérieur de l'enveloppe; évitant ainsi des calculs supplémentaires et l'apparition de bruits dans le résultat final. Cette solution, permet néanmoins, d'acquérir l'objet sans être obligé à utiliser un arrière plan de couleur spécifique pour éventuelle segmentation (*séparation de l'objet de son environnement*). La figure IV- 2 illustre bien l'optimisation apportée à l'algorithme de mise en correspondance.

L'enveloppe visuelle va nous permettre de délimiter verticalement la zone de recherche de correspondants (*Figure IV- 3*); ainsi, les lignes à analyser seront positionnées exactement entre les deux lignes horizontales E_i et E_s (*limite inférieure et supérieure de l'enveloppe visuelle*).



Objet à reconstruire.



Zone de recherche.



Résultat de la mise en correspondance.

Figure IV- 2 : Minimisation de la zone de recherche en utilisant l'enveloppe visuelle.



Figure IV- 3 : Délimitation verticale (par les valeurs E_i et E_s) de la zone de recherche.

2-2 Deuxième optimisation

Dans notre système, nous proposons de mettre l'objet à reconstruire sur une table tournante (par pas angulaire fixe). Une caméra immobile capture les différentes vues de la scène.

Les silhouettes de l'objet vont être utilisées pour inférer les profondeurs des points se trouvant sur sa bordure.

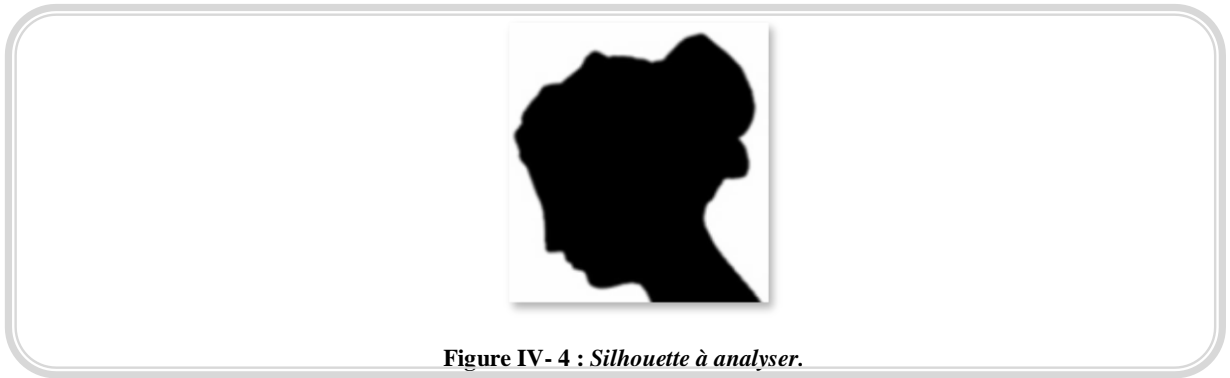


Figure IV- 4 : Silhouette à analyser.

Puisque l'objet tourne sur une table, son volume peut être délimité par une sphère comme le montre la figure suivante :

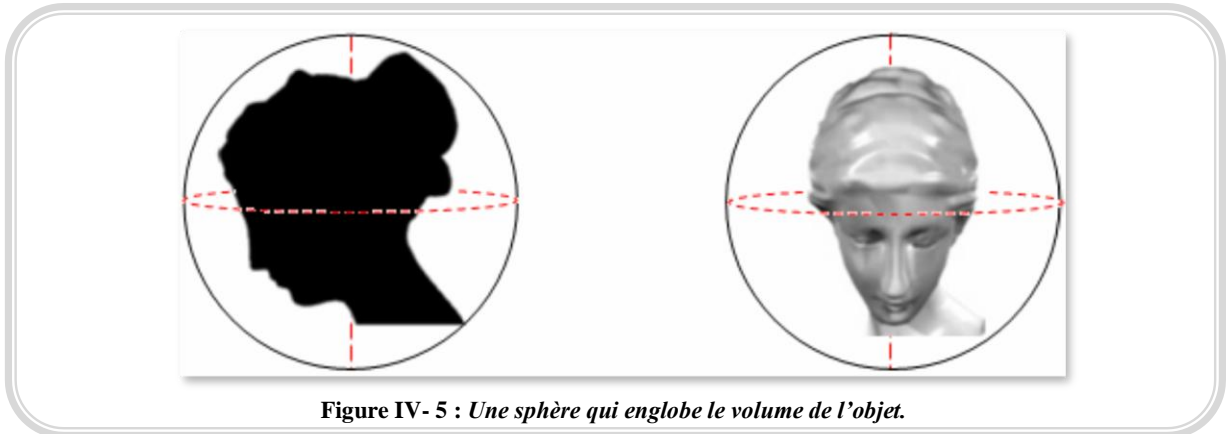


Figure IV- 5 : Une sphère qui englobe le volume de l'objet.

Notre but est de calculer des profondeurs P_i (où i est le numéro de ligne) par rapport à l'axe de rotation de l'objet, ces mesures nous aideront par la suite à construire effectivement l'enveloppe visuelle de l'objet.

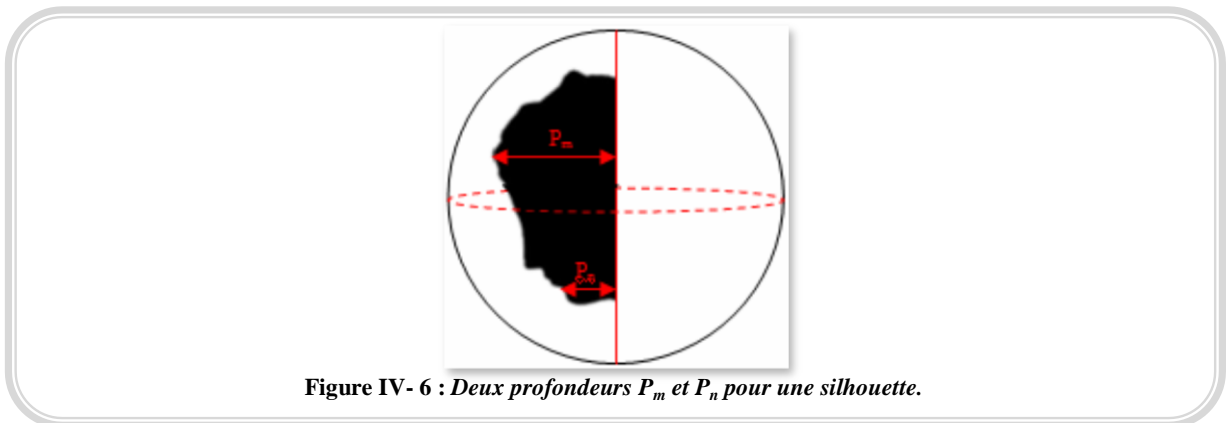


Figure IV- 6 : Deux profondeurs P_m et P_n pour une silhouette.

En procédant ainsi, et en utilisant toutes les silhouettes disponibles, nous obtenons un ensemble de profondeurs qui nous permettront de construire une enveloppe visuelle englobant exactement l'objet.

La Figure IV-7 nous rappelle comment calculer la longueur d d'un arc de diamètre p et limité par un angle α (en degrés).

$$d=(2\pi p)*(\alpha / 360) \quad (1)$$

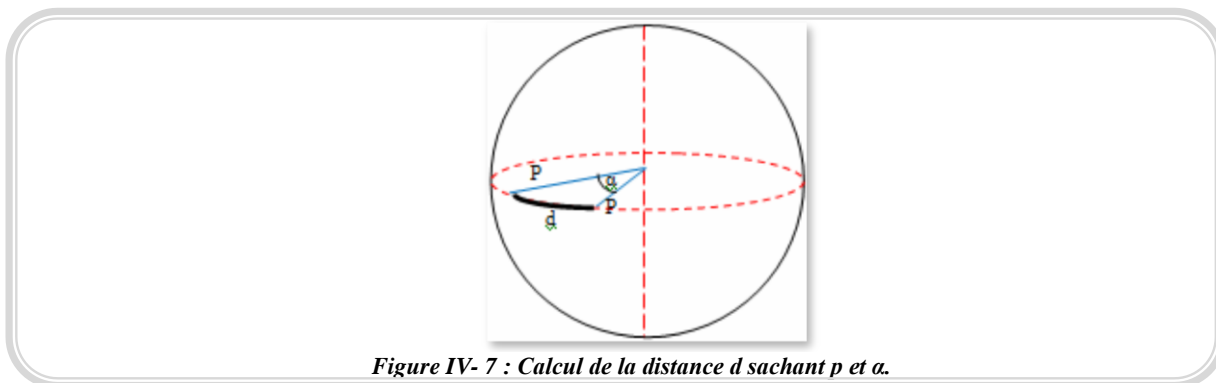


Figure IV- 7 : Calcul de la distance d sachant p et α .

Pour notre méthode, nous utilisons la stéréovision, non seulement pour reconstruire le modèle 3D de l'objet, mais aussi pour améliorer l'algorithme de mise en correspondance.

Chaque point de la surface de l'objet va former un cercle virtuel lors de sa rotation. La figure suivante montre le même principe appliqué à l'objet à reconstruire:

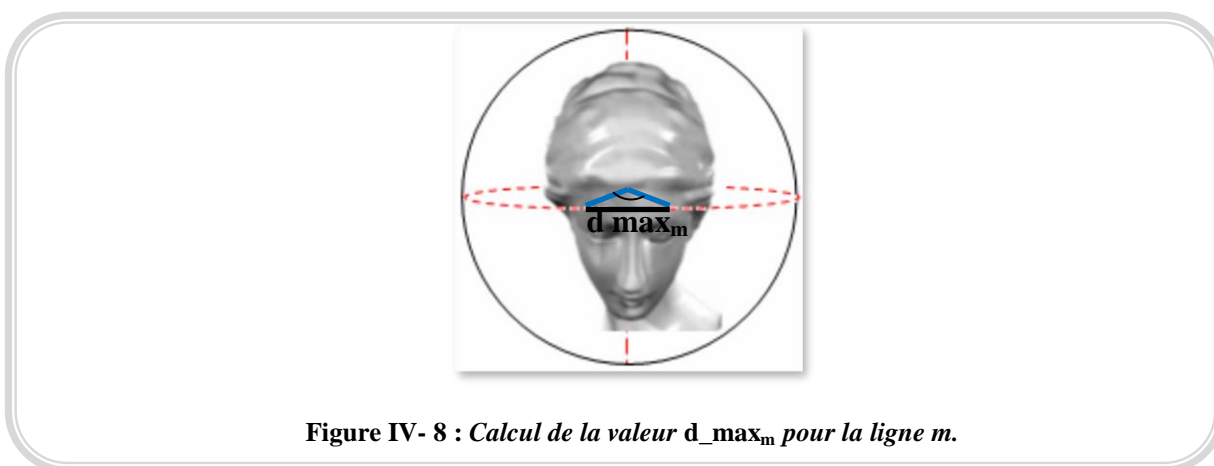


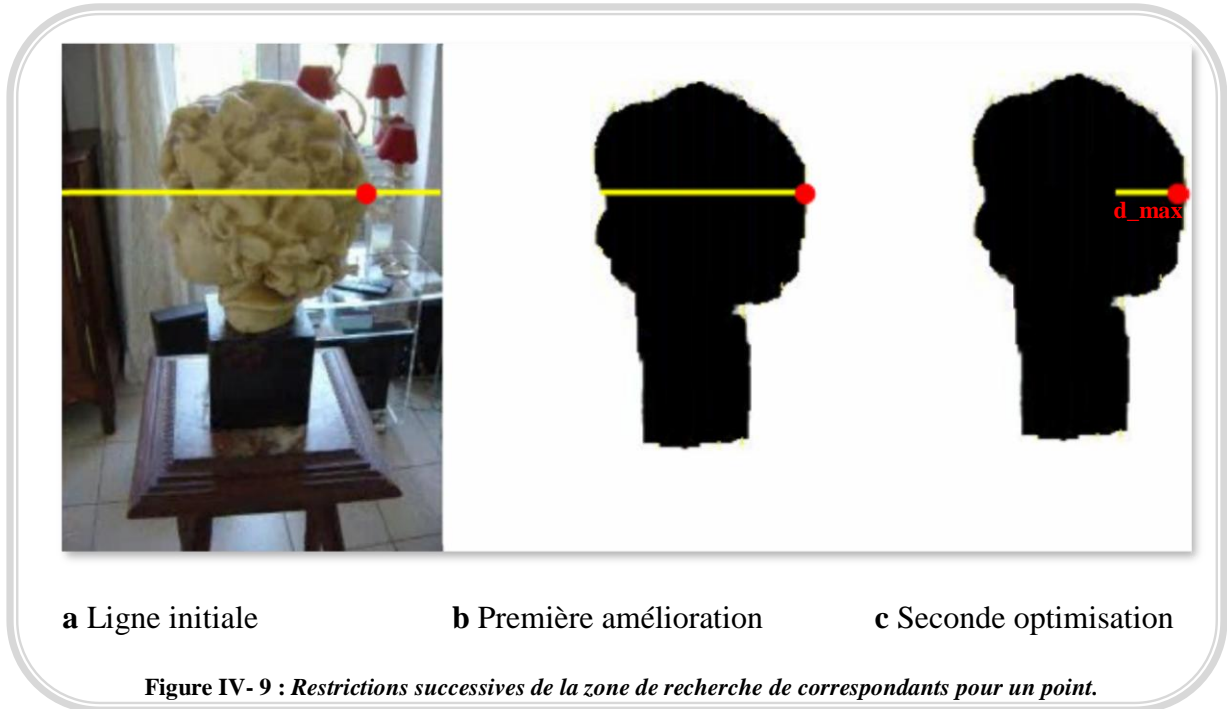
Figure IV- 8 : Calcul de la valeur d_{max_m} pour la ligne m .

Puisque la table tourne avec un pas angulaire fixe, et que nous connaissons les paramètres intrinsèques et extrinsèques de la caméra; nous pouvons calculer la distance entre la position initiale et finale d'un point se trouvant sur la bordure de l'objet. Cette distance constitue en fait la valeur d_{max_i} (*disparité maximale de la ligne epipolaire numéro i*). En d'autres termes ; le déplacement d'un point se trouvant à l'intérieur de l'enveloppe visuelle d'une image à l'autre ne peut, en aucun cas, dépasser la distance d_{max_i} puisque les points les plus éloignés de l'axe de rotation (*bordure de l'objet*) ont les plus grandes profondeurs p_{max} ; pour tout point de profondeur p et de distance d : $p \leq p_{max} \Rightarrow d \leq d_{max_i}$ (*selon l'équation 1*). Il sera alors inutile de calculer de correspondants pour ces points en delà de cette distance.

La distance d_{max_i} sera utilisée par la suite durant le processus de mise en correspondance pour limiter horizontalement le champ de recherche ; alors on ne va examiner qu'un fragment de largeur d_{max_i} de la droite epipolaire déterminé pour chaque profondeur calculée.

Comme nous l'avons mentionné, la plupart des techniques de mise en correspondance se basent sur un balayage de toute l'image. Pour un pixel donné, elles analysent tous les points de la ligne epipolaire qui le franchit. Notre première optimisation consiste à limiter le champ de recherche à un fragment qui résulte de l'intersection de la ligne epipolaire avec la silhouette. L'utilisation de la disparité d_{max_i} comme montrée précédemment, permet d'appliquer une nouvelle réduction à ce fragment.

La figure ci-dessous montre les restrictions successives appliquées sur la zone de recherche. L'image (a) présente la ligne initiale (*en jaune*) de recherche du correspondant d'un point (*en rouge*) qui s'étend sur la totalité de la largeur de l'image. L'image (b) présente la première amélioration apportée en utilisant l'enveloppe visuelle, la recherche se fait sur un fragment ne dépassant pas la frontière de l'enveloppe visuelle. La deuxième amélioration apparaît clairement dans l'image (c) ; la zone de recherche est limitée à une distance d_{max} à l'intérieur de l'enveloppe visuelle.



L'application de ces restrictions pour chaque point de l'image de référence permet de gagner beaucoup de temps au profit de la mise en correspondance ; donc de la globalité de la reconstruction 3D.

3. Etapes du processus de reconstruction

Notre méthode de reconstruction se déroule en quatre étapes :

- (1) Acquisition des photos,
- (2) Construction de l'enveloppe visuelle et restriction de la zone de recherche,
- (3) Mise en correspondance,
- (4) et reconstruction 3D effective.

Nous allons dans ce qui suit détailler chacune de ces phases.

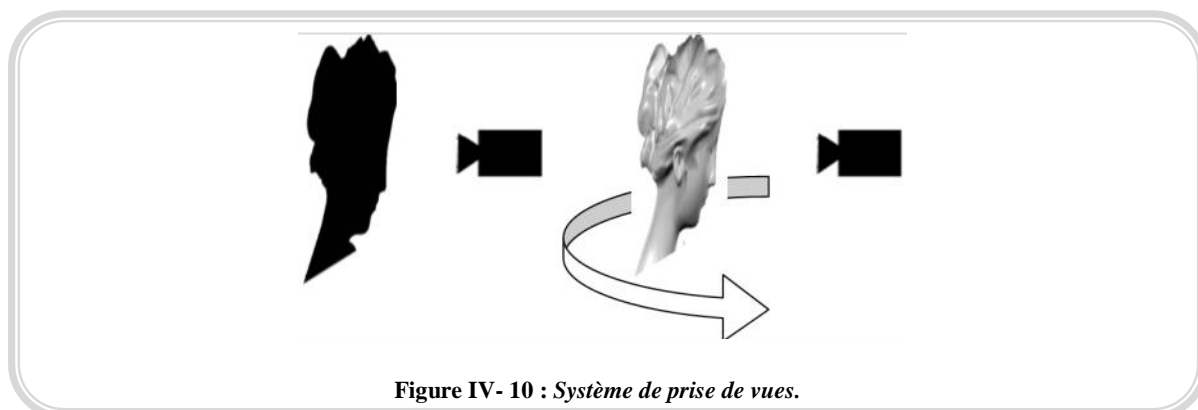
3-1 Système de prise de vues

L'objet à numériser est placé sur un support tournant, deux caméras identiques fixes (*dont les paramètres intrinsèques et extrinsèques sont connus*) vont servir pour l'acquisition simultanée des images de référence ainsi que des silhouettes. La table tourne et une image est prise par chacune des deux caméras à pas angulaire régulier (*10 degrés¹*).

¹ Plus l'angle de rotation est petit, plus l'enveloppe sera exacte.

On dispose des informations de calibrage des caméras utilisées :

- paramètres intrinsèques renseignés par le fabricant,
- et les paramètres extrinsèques qui sont issus de la géométrie de la trajectoire et de la distance entre la caméra, l'objet et la silhouette.



3-2 Restriction de la zone de recherche de correspondants

Les valeurs limites E_i et E_s (*extrémité inférieure et supérieure de l'enveloppe visuelle*) sont les mêmes pour toutes les images, et sont calculées une seule fois à partir de la première silhouette.

Pour chaque silhouette, nous pouvons calculer les profondeurs P_i (*comme expliqué dans la section 2.2*) qui représentent les positions 3D correspondantes qui seront ainsi mémorisées et affichées sur l'écran. Un nuage de points préliminaire représentant l'enveloppe visuelle est donc obtenu lorsque nous complétons un tour complet.

Il suffit alors pour restreindre horizontalement la zone de recherche, d'associer à chaque ligne epipolaire i de chaque photo les valeurs d_{max_i} calculées (*section 2.2*).

3-3 Mise en correspondance

L'algorithme de mise en correspondance va donc, pour chaque ligne i de l'image de référence de droite; chercher le correspondant de chaque point se trouvant entre les deux extrémités du fragment epipolaire -se trouvant à l'intérieur de l'enveloppe visuelle- sur une distance inférieure ou égale à d_{max_i} .

3-4 Reconstruction 3D

Une fois que la mise en correspondance entre deux pixels est établie par l'application d'une des techniques de reconstruction stéréoscopique, nous pouvons 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.

Ce processus est réalisé pour chaque point appartenant à l'enveloppe visuelle et mis en correspondance avec son homologue dans la zone de recherche d_{max_i} . Ainsi, le nuage de points obtenu dans la phase précédente est rectifié pour obtenir le modèle 3D exacte de l'objet.

4. Algorithme

Algorithme Reconstruction_3D

Début

Acquisition des photos et silhouettes.

Construction de l'enveloppe visuelle.

Pour chaque couple $(j, j+1)$ d'image de référence, $j \leftarrow 1$ à n faire

Pour chaque ligne i de l'image j , $i \leftarrow E_i$ à E_s faire

 - Calculer la valeur d_{max_i} à partir de la silhouette de l'image $j+1$.

Pour chaque pixel dans la zone recherche de l'image j faire

 - Rechercher son correspondant sur la zone de longueur d_{max_i} de l'image $j+1$.

Si correspondant trouvé alors

 - Calculer et sauvegarder sa position 3D.

Finsi

FinPour

FinPour

FinPour

 - Afficher les points 3D calculés.

Fin

5. Expérimentation et résultats

5-1 Images utilisées

Afin de se concentrer sur les aspects fondamentaux de la méthode que nous proposons, nous avons simulé le système de prise de vues par l'utilisation directe de 36 images calibrées (figure IV-11) d'un objet (*angle de rotation choisi 10°*).

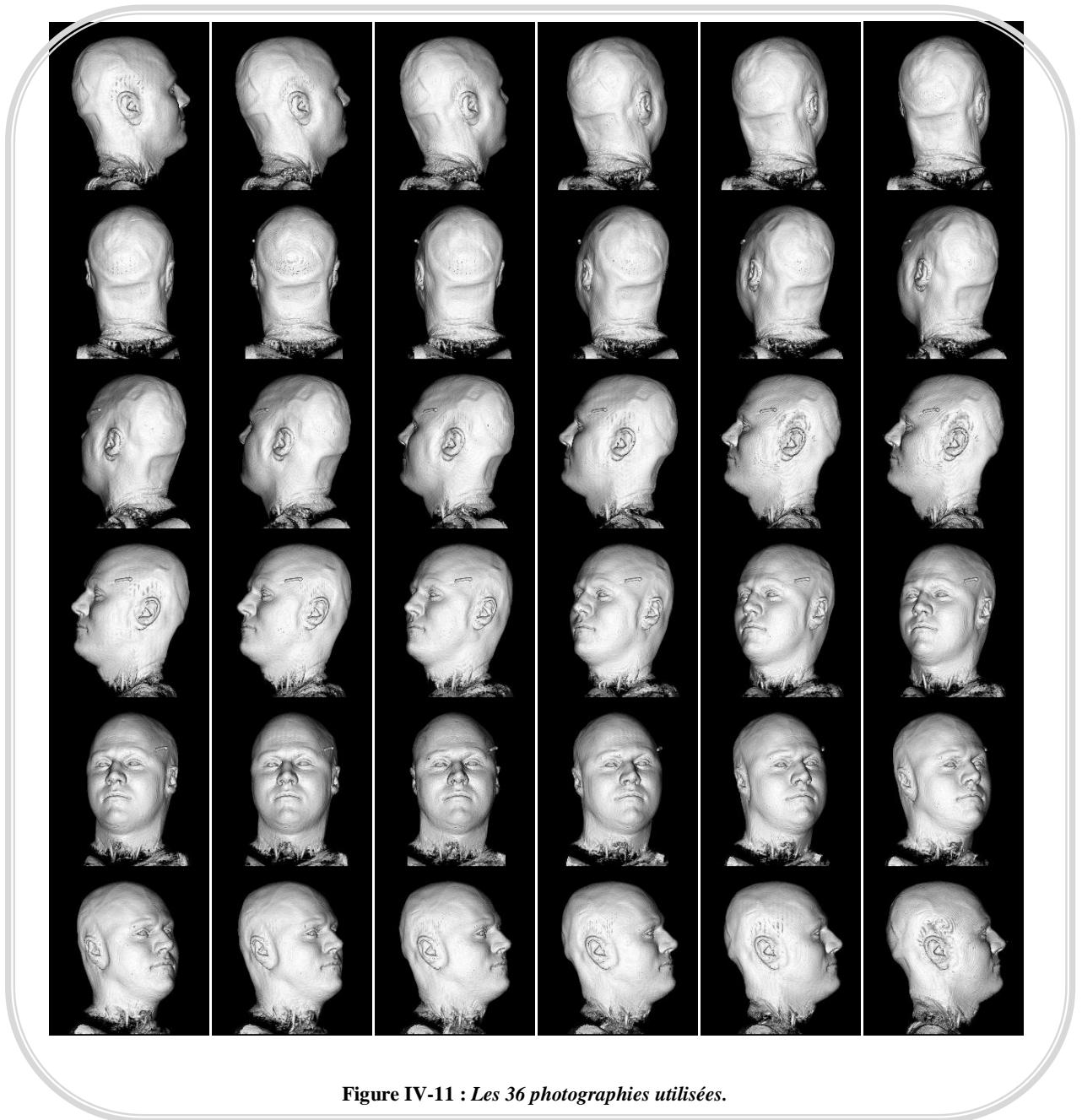


Figure IV-11 : Les 36 photographies utilisées.

5-2 Préparation des silhouettes

La préparation des silhouettes a été effectuée manuellement. Nous avons traité les photos initiales pour obtenir des images monochromes (*Noir/Blanc*) –du fait que les entrées de notre algorithme sont les photos et les silhouettes–.

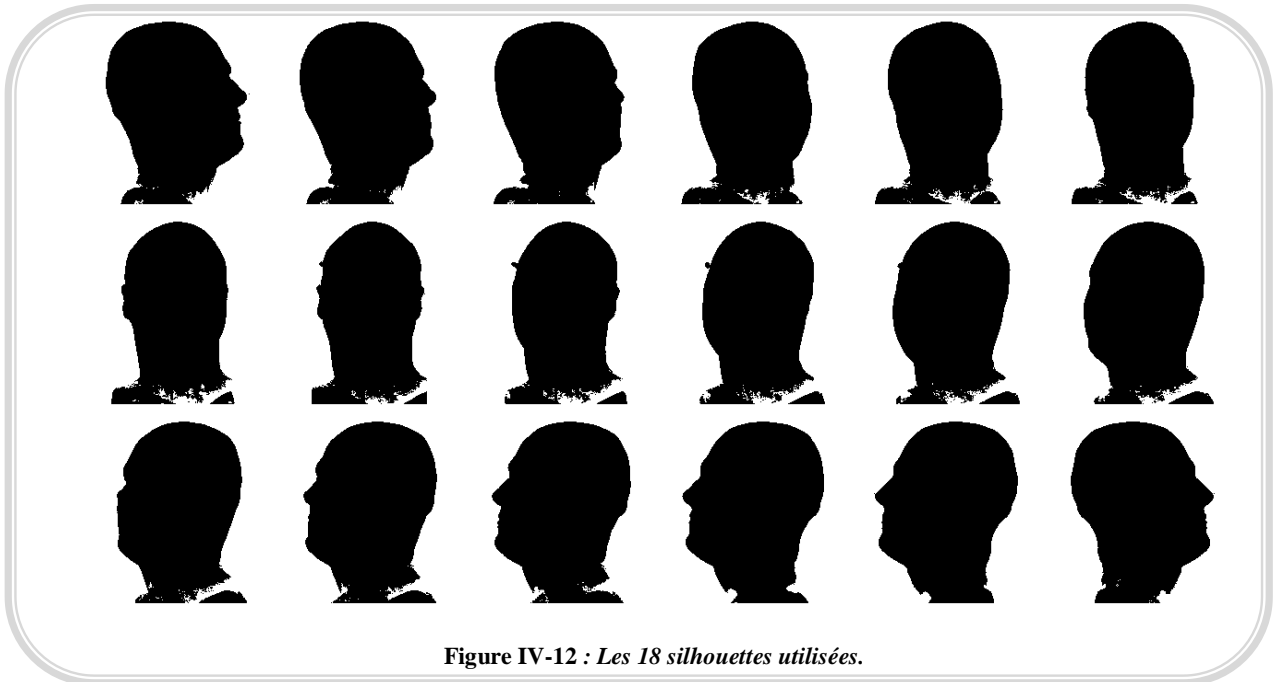


Figure IV-12 : Les 18 silhouettes utilisées.

5-3 Construction de l'enveloppe visuelle

Pour construire l'enveloppe visuelle de l'objet, nous proposons [] un algorithme qui permet de la générer instantanément.

Nous avons utilisé 18 silhouettes, chacune a été utilisée pour calculer les profondeurs des points se trouvant sur la bordure de l'objet par rapport au centre de rotation (*Comme montré précédemment*). Une silhouette k (*tel que $1 \leq k \leq 18$*) est utilisée pour calculer à la fois :

- les profondeurs relatives à la silhouette k ,
- et les profondeurs relatives à la silhouette $k+18$ ($10^\circ * 18 = 180^\circ$ soit un demi-tour) en prenant part de la symétrie inversée des silhouettes d'un objet (*figure IV-13*).

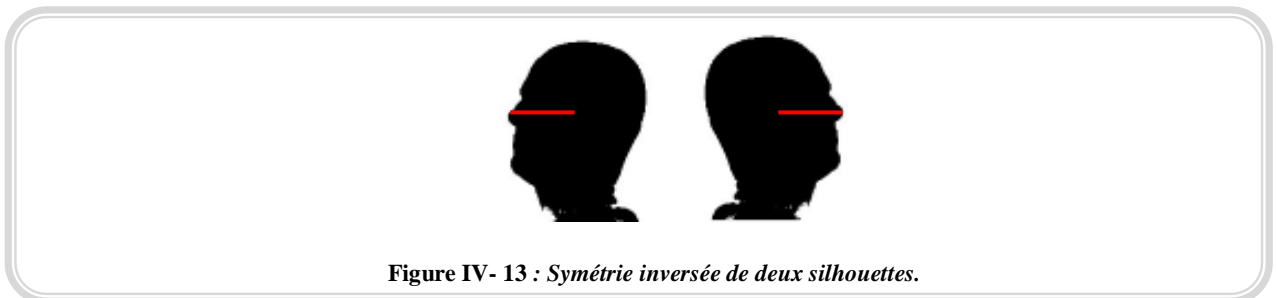


Figure IV- 13 : Symétrie inversée de deux silhouettes.

Les profondeurs extraites à partir de la partie gauche d'une silhouette k (*tel que $1 \leq k \leq 18$*) sont égales à celles extraites à partir de la partie droite de la silhouette $k+18$, donc, il est inutile de réfère le calcul.

La profondeur calculée P_i représente en fait la troisième coordonnée Z d'un point défini par (X, Y, Z) . Y est le numéro de la ligne utilisée pour calculer cette profondeur, et X est relatif au numéro de la silhouette analysée (*Rotation dans l'espace*). La position du point 3D ainsi calculée est mémorisée, et ce processus est réalisé pour chaque point appartenant à la bordure de la silhouette pour obtenir un nuage de points qui délimite le volume de l'objet. Par conséquent, un nuage de points (*figure IV-14*) est obtenu après le positionnement dans l'espace des différents points dont les profondeurs sont calculées.

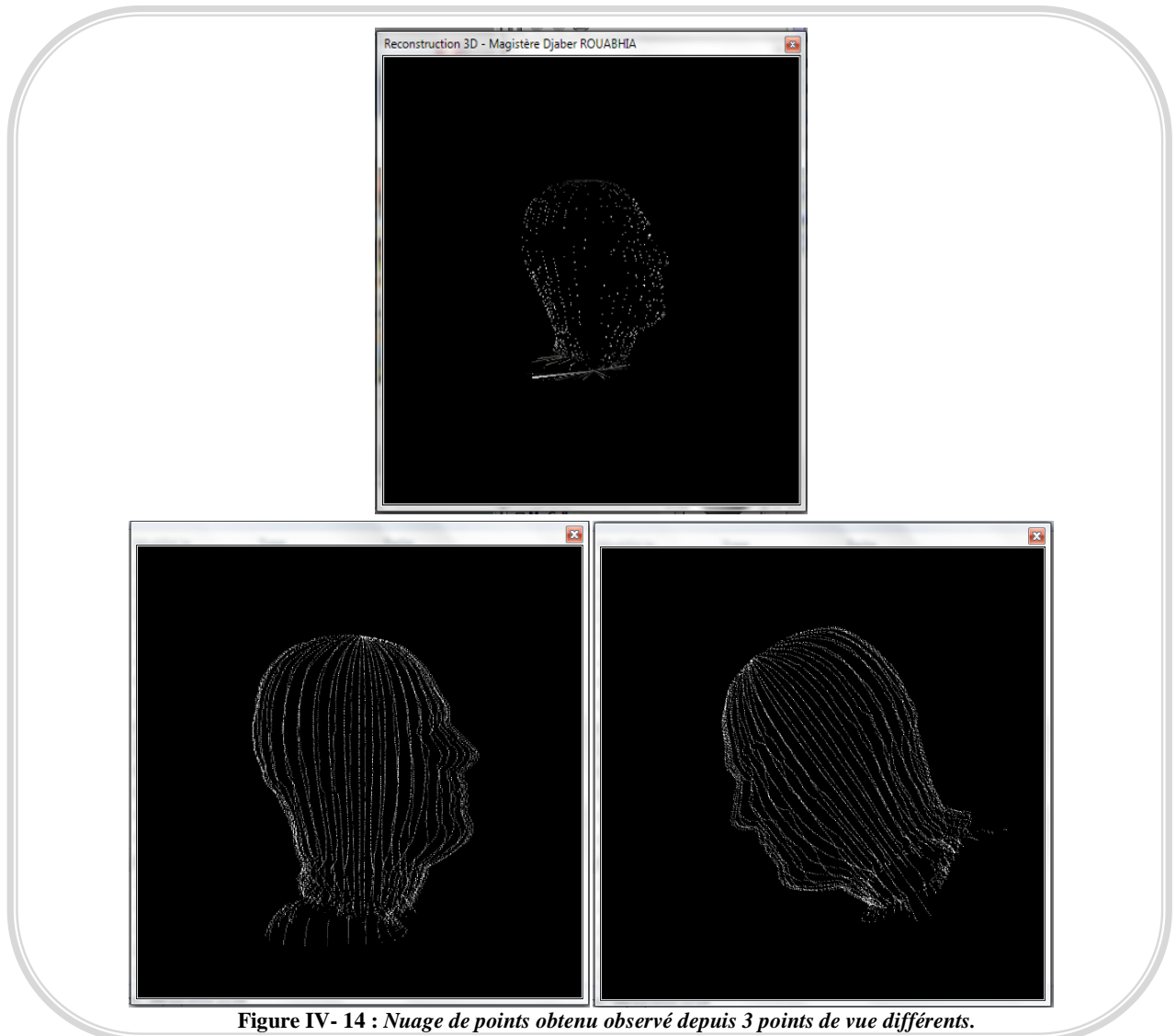


Figure IV- 14 : Nuage de points obtenu observé depuis 3 points de vue différents.

La mise en œuvre de l'algorithme de construction de l'enveloppe visuelle nous a permis d'obtenir des résultats que nous considérons comme acceptables:

- Une grande précision est fournie, et l'enveloppe obtenue est une très bonne approximation du volume de l'objet à reconstruire,
- Les temps de calculs sont très satisfaisants du fait que l'observateur ne détecte aucune lenteur et que les résultats sont instantanément visualisés.

L'enveloppe visuelle (*figure IV-15*) obtenue constitue une première approximation du volume de l'objet à reconstruire, et nous pouvons lui appliquer quelques opérations telles que les rotations dans l'espace (*selon les trois axes*) ou Agrandissement/Atténuation.

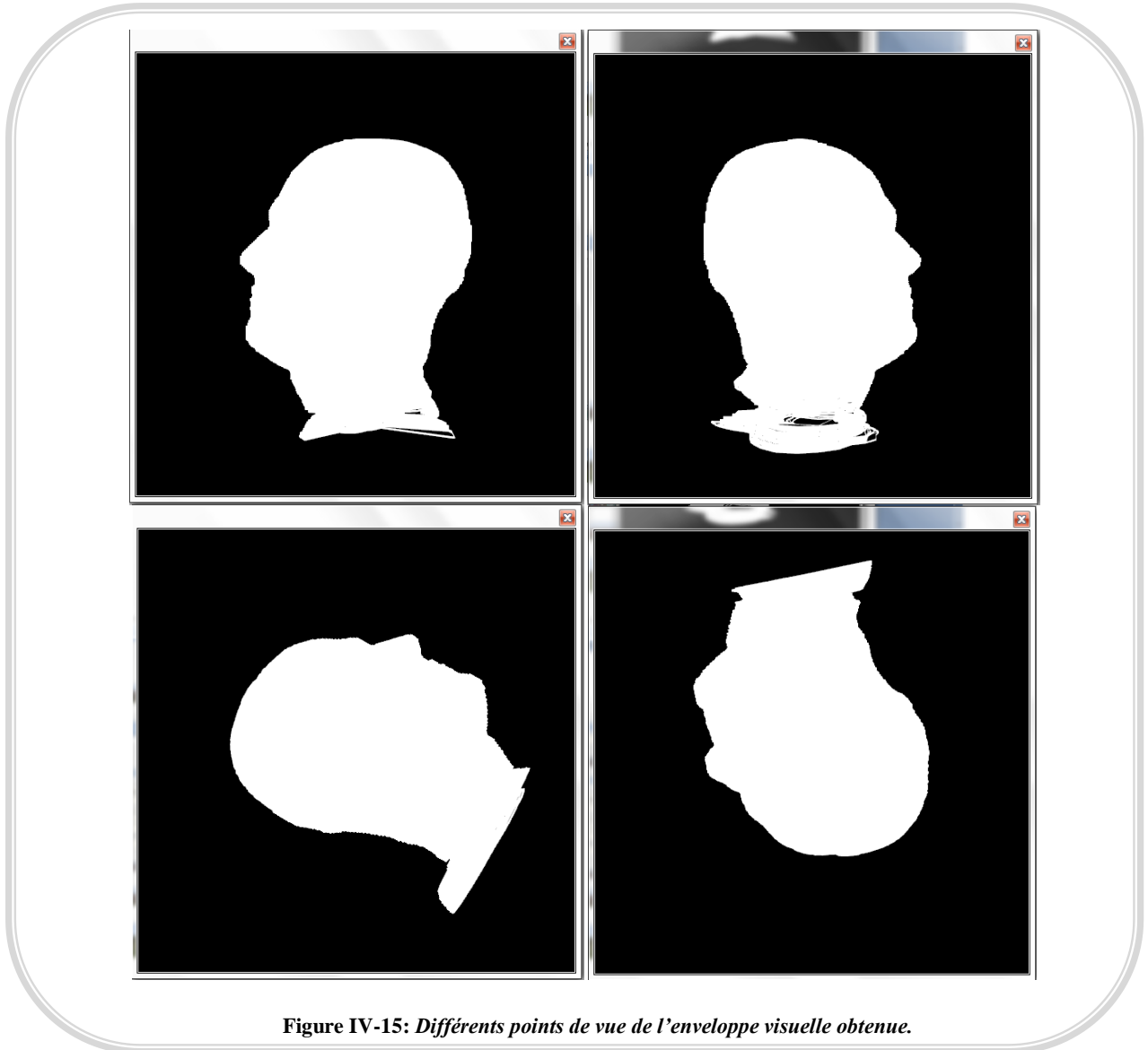


Figure IV-15: Différents points de vue de l'enveloppe visuelle obtenue.

5-4 Reconstruction 3D

Enfin, pour la reconstruction effective de l'objet, nous avons pu établir une première reconstruction partielle []. En récupérant (à partir des 36 images réelles) les couleurs des points calculés, nous avons obtenus des résultats instantanés donnant une reconstruction partielle (figure IV-16) très proche de la réalité avec une légère différence par rapport à l'objet à reconstruire.

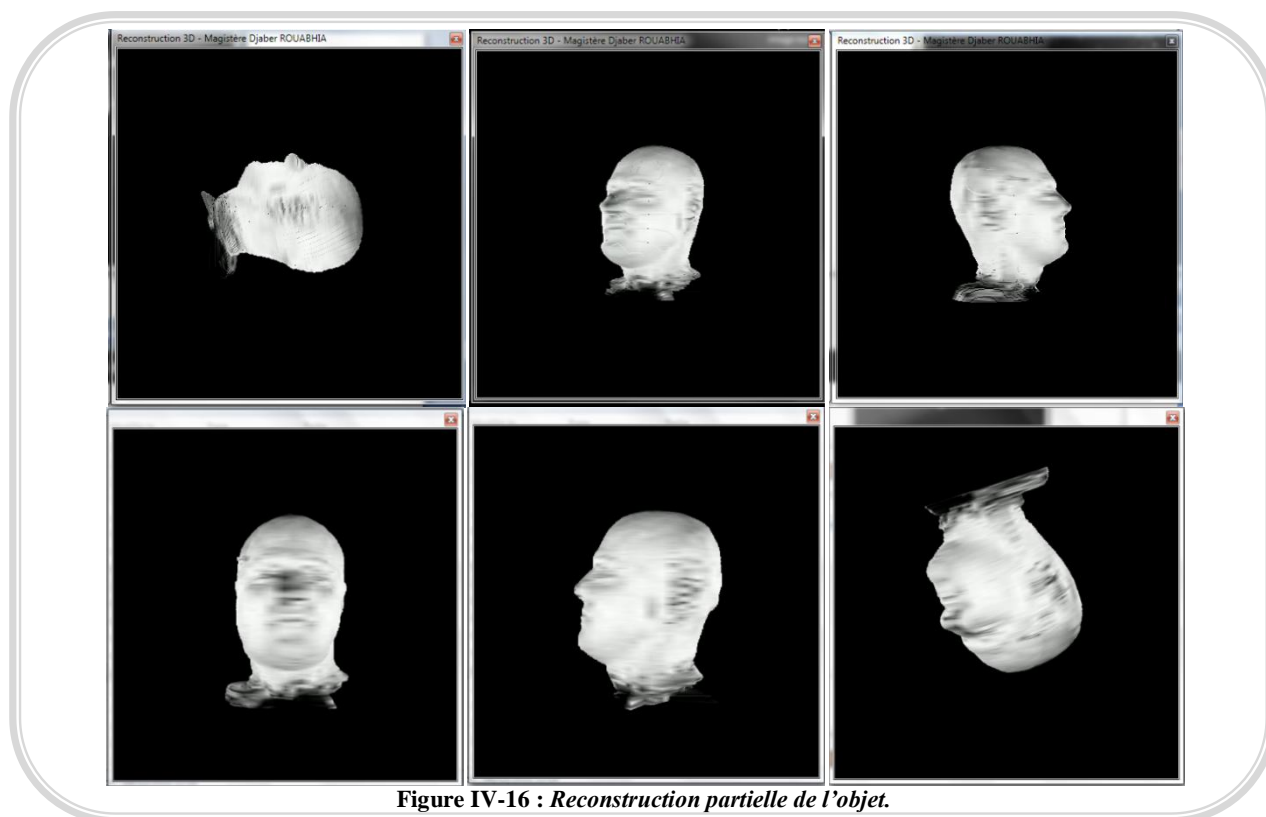


Figure IV-16 : Reconstruction partielle de l'objet.

Après avoir restreint la zone de recherche -par le calcul des valeurs de disparité maximale d_{max_i} -, la reconstruction 3D finale de l'objet peut être ainsi faite par l'application d'un des algorithmes de stéréovision disponibles (*qui sont très nombreux*) -voir [] par exemple- qui permettent de raffiner l'enveloppe visuelle et de détecter les régions concaves de l'objet.

Pour chaque silhouette, un tableau qui contient la disparité maximale calculée pour chaque ligne epipolaire de chaque silhouette est généré. Ce tableau peut être utilisé pour minimiser la zone de recherche de correspondants pour les algorithmes stéréoscopiques (*comme expliqué précédemment*).

Le tableau IV-1 montre bien les apports de l'application des restrictions proposées sur la zone de recherche. Nous avons choisi aléatoirement dix lignes epipolaires successives d'une silhouette quelconque. La première colonne contient le nombre de pixel (*sans aucune restriction*) à analyser lors de l'appariement. La deuxième colonne représente la largeur de la zone de recherche (*exactement à l'intérieur de l'enveloppe visuelle*) après l'application de la première optimisation. Et finalement, la troisième colonne montre clairement le degré de minimisation de la zone de recherche en utilisant notre deuxième optimisation (*usage de la disparité maximale*).

Comme il est clair, effectuer le test sur 95 pixels est beaucoup plus rapide par rapport à l'utilisation des 816 pixels originaux (*même par rapport à l'utilisation de 544 pixels après l'application de la première optimisation*). Ainsi, un grand temps est préservé et la reconstruction 3D est bien accélérée.

Nous nous sommes convaincus [22] que la restriction de la zone de recherche que nous avons proposée est très importante et va réduire au maximum les temps nécessaires pour la reconstruction 3D effective d'un objet.

Table IV- 1. Restrictions de la zone de recherche (*en pixel*).

Zone de recherche originale	Résultat de la première optimisation	Résultat de la seconde optimisation <i>($d_{max_i} = 2\pi * p_i * \alpha / 360$)</i>
816	514	90
816	518	90
816	523	91
816	524	91
816	528	92
816	530	93
816	536	94
816	543	95
816	544	95
816	550	96

6. Avantages

La minimisation du temps des algorithmes stéréoscopiques pour la reconstruction 3D constitue l'axe principal de notre travail. Le principe est de restreindre la zone de recherche à de petits fragments des images ; optimisant ainsi l'algorithme de mise en correspondance de manière très importante.

Le regroupement des améliorations proposées permet d'effectuer des recherches sur de petites parties des images de référence. Nous pouvons ainsi résumer les avantages suivants :

- **Simplicité de réalisation du système de prise de vues,**
- **Utilisation de photographies réelles (*aucune nécessité d'utiliser des arrières plans de couleurs spécifiques ou de procéder à des segmentations*),**
- **Facilité de l'implémentation de l'approche proposée,**
- **Accélération de la mise en correspondance (*malgré le temps consacré pour la restriction du champ de recherche*),**
- **Augmentation de la fiabilité et de la qualité des résultats de la mise en correspondance (*réduction de la possibilité des mises en correspondance incorrectes*),**
- **Obtention de la plus grande précision géométrique possible,**
- **La mise en correspondance et la reconstruction sont réalisées en une seule passe ; ce qui évite l'utilisation de variables intermédiaires (*tableaux volumineux*),**

- **Utilisation des photos réelles pour la coloration de l'objet avec éventuelle possibilité de placage de textures selon choix,**
- **Possibilité de sauvegarder le modèle sous format approprié pour éventuelle manipulation par des outils spécialisés.**

7. Conclusion

Nous avons présenté une méthode qui combine la reconstruction à partir de la technique de l'enveloppe visuelle et d'images stéréo pour la reconstruction 3D d'objets.

D'abord, nous avons proposé un algorithme instantané utilisant les silhouettes de l'objet 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 verticalement le champ de recherche des correspondances.

Ensuite, nous avons proposé un deuxième algorithme (*aussi instantané*) pour déterminer les valeurs d_{max_i} utilisées par la suite pour restreindre la zone de recherche. Nous avons en même temps pu extraire les couleurs réelles à partir des photographies disponibles pour les appliquer au modèle 3D obtenu.

La limitation du champ de recherche améliore la vitesse et la qualité de la reconstruction stéréo. L'information stéréo peut compenser certains des inconvénients inhérents de la méthode de l'enveloppe visuelle, tels que l'incapacité de reconstruire les détails extérieurs et les régions concaves.

L'implémentation d'un outil basé sur l'approche que nous avons proposé nous a permis d'obtenir des résultats dans de très brèves durées (*et nous pensons même qu'elle peut être améliorée afin de la qualifier temps réel*).

Le modèle 3D obtenu est très précis et représente exactement l'objet à reconstruire (*chose évidente vu le nombre de photographies utilisées*).

La qualité visuelle du modèle rendu est très proche de l'objet réel, et une légère différence est aperçue (*due au décalage des photographies*). Cette qualité peut être améliorée [22] en utilisant plus de photographies ou par l'application de l'un des procédés de stéréoscopie disponibles.

Néanmoins, l'indisponibilité de plateforme programmée (*surtout algorithmes de stéréovision et mise en correspondance*) nous empêche pour l'instant d'aller plus loin dans ce travail. Cette tâche fait part major de nos futures perspectives, où nous tenterons d'une part de mettre en œuvre une plateforme dans ce sens, et d'autre part, d'appliquer notre algorithme général (*restriction de champ de recherche pour la reconstruction 3D*) en impliquant différents algorithmes de stéréovision.

Conclusion générale et perspectives

Dans ce mémoire, nous avons présenté 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, dont l'un des objectifs est la reconstruction de la structure tridimensionnelle d'objets à partir d'une ou plusieurs images.

Les approches traditionnelles consistent 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 images de synthèse ainsi générées n'étonnent plus grand monde, pourtant les chercheurs ne cessent d'améliorer la qualité graphique des images tout en réduisant leur temps de calcul.

Beaucoup de techniques d'IMBR ont été développées ces dernières années, et ont trouvé des applications passionnantes dans les domaines de la réalité augmentée, réalité virtuelle, jeux, cinématographie, etc.

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'IMBR dans le but d'améliorer la modélisation d'environnements 3D, tant au niveau de la rapidité de conception, qu'au niveau de la précision et du réalisme.

La reconstruction 3D vise essentiellement à fournir des images de synthèse au moindre coût et meilleure qualité. En effet, la reconstruction 3D 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 s'introduisant à ce domaine, nous avons constaté que plusieurs recherches sont menées dans ce contexte. Ainsi on peut trouver énormément de procédés et de techniques de reconstruction 3D, chacun avec ces propres caractéristiques et son propre champ d'application.

Nous avons alors donné deux propositions principales dans le cadre de la reconstruction 3D :

- Premièrement, on a tenté d'élaborer une classification multicritère des approches actuelles de reconstruction 3D.
- Ensuite, nous avons proposé une nouvelle méthode multi-vue pour la reconstruction 3D.

Puisque le rassemblement de toutes les techniques dans un simple état de l'art est très difficile (*voir même impossible*), nous avons tenté de les classer selon différents critères. Ainsi, nous avons consacré un chapitre entier pour cette classification en se basant sur neuf critères totalement distincts l'un de l'autre.

Nous avons ensuite proposé une approche qui combine (*dans une nouvelle optique*) deux méthodes d'IMBR pour la reconstruction réaliste d'objets 3D. La première est la technique de l'enveloppe visuelle, et la deuxième est la reconstruction à partir d'images stéréo.

Ces deux méthodes sont complémentaires en nature. Nous avons proposé un algorithme pour la construction instantanée de l'enveloppe visuelle d'un objet (*l'utilisateur ne détecte aucune lenteur*), et les résultats donnent des formes très proches de celles des objets à reconstruire. Cette enveloppe va servir par la suite à éviter des calculs inutiles pour des endroits en dehors du volume de l'objet lors de l'application des algorithmes stéréoscopique.

Les méthodes de la stéréovision raffinent le modèle construit de l'objet par la détection des régions concaves sur sa surface. Nous les utilisons ici non seulement pour reconstruire l'objet, mais aussi en adjonction avec l'enveloppe visuelle pour réduire les zones de recherche de correspondants à de petits fragments de l'image réelle. Nous utilisons pour cette fin la mesure de la disparité maximale pour tous points appartenant à la bordure de l'enveloppe visuelle.

Notre but principal était d'accélérer les calculs, mais nous avons aussi constaté que nos propositions donnent de bons résultats, puisque nous utilisons, d'une part, un grand nombre d'images réelles (*éclairage, couleur*) sans avoir besoin de segmentation, et d'autre part, puisque aucune correspondance en dehors de l'enveloppe n'a été calculée (*élimination antérieure d'éventuels bruits pouvant apparaître dans les résultats*).

Perspectives

D'après notre brève expérience dans le domaine de la reconstruction 3D, nous pensons qu'il est possible de prendre en considération des objets plus complexes et d'améliorer les résultats obtenus par l'application de notre approche. Nous trouvons évident d'énoncer les futures perspectives suivantes :

Nous comptons utiliser des vidéos pour l'extraction automatique des images et des silhouettes, ce qui va permettre d'améliorer l'exactitude du modèle 3D (*utilisation de plus de silhouettes décalées d'angle plus petits*) et de diminuer encore plus le temps nécessaire pour la reconstruction que nous proposons.

Nous envisageons de sauvegarder les résultats sous formats permettant de futures manipulations par des outils spécialisés de modélisation 3D.

Aussi, parmi les perspectives immédiates de notre travail, nous pouvons bien sûr citer la validation de l'approche proposée avec plus d'algorithmes de mise en correspondance. Nous avons déjà démontré la faisabilité d'une telle approche, néanmoins, la validation sur d'autres algorithmes de mise en correspondance nous permettrait d'en quantifier le temps et la qualité de la reconstruction ainsi obtenue.

Nous pensons aussi à l'adaptation de l'approche proposée pour la reconstruction 3D en temps réel et d'impliquer les cartes graphiques.

Le plus important en ce qui concerne nos perspectives, est de pouvoir regrouper dans une plateforme unique plusieurs méthodes de mise en correspondance (*qui est un problème très délicat*) et les mettre à la disposition des personnes travaillant dans le domaine de la reconstruction 3D par stéréovision (*l'une des difficultés que nous avons rencontré pour la validation de notre proposition est l'indisponibilité de techniques de mise en correspondance programmées et prêtes à l'emploi*).

Références

- [1] A. CRIMINISI, A. ZISSERMAN. Shape from texture: homogeneity revisite. University of Oxford, 2000.
- [2] A. HEYDEN, K. A. STROM. Euclidean Reconstruction from Constant Intrinsic Parameters. Proc. 13th International Conference on Pattern Recognition, IEEE Computer Soc. Press, 1996.
- [3] A. LAURENTINI. The visual hull concept for silhouette based image understanding. IEEE Trans. PAMI, 1994.
- [4] A. MEYER, H. M. BRICENO, S BOUAKAZ. User-guided Shape from Shading to Reconstruct Fine Details from a Single Photograph. Université de Lyon, LIRIS. France, 2007.
- [5] A. SHASHUA, L. WOLF. Homography tensors: On algebraic entities that represent three views of static or moving planar points. Proc of the 6th European Conference on Computer Vision. Ireland, Springer-Verlag, 2000.
- [6] A. V. DEN. VideoTrace: Rapid interactive scene modelling from video. University of Adelaide. AUSTRALIA, 2007.
- [7] B. BOUFAMA. Reconstruction tridimensionnelle en vision par ordinateur : Cas des caméras non étalonnées. Thèse PhD de l'institut National Polytechnique De Grenoble, 1994.
- [8] B. BOUFAMA. Recovering the Three-Dimensional Structure Using Two-Dimensional Images: Essentials and Review. The International Conference on Complex Systems. Cisc'4 University Of Jijel, 2004.
- [9] B. NAZARIAN. Imagerie Médicale 3D : Visualisations, segmentations et reconstructions. CNRS, France, 2002.
- [10] C. BAILLARD. An Automatic reconstruction of piecewise planar models from multiple views. Conference on Computer Vision and Pattern Recognition, 1999.
- [11] C. H. ESTEBAN, F. S. ENST. Reconstruction 3D d'objets par multi-stéréo. AFIG 2001 - 14èmes journées de l'Association Française d'Informatique Graphique, pp. 27-36, Limoges, Novembre 2001.
- [12] C. H. ESTEBAN, Y. YEMEZ, F. SCHMITT. 3D Reconstruction of Real Objects from silhouettes and stereo. Pattern Recognition and Computer Vision Colloquium Prague, Czech, 2001.
- [13] C. SHENCHANG. QuickTime ccVR : An image-based approach to virtual environment navigation. Proceedings of the Conference on Computer Graphics (SIGGRAPH-95), Los Angeles, USA. 1995.
- [14] C. ZHANG, T. CHEN. A Survey on Image-Based Rendering. Technical Report AMP 03-03 Advanced Multimedia Processing Lab Carnegie Mellon University Pittsburgh, 2003.
- [15] D. ABDERRAHIM. Synthèse de Nouvelles Vues pour les Applications en Réalité Augmentée. Mémoire magistère, Université Mentouri, Juin 2006.
- [16] D. BURSCSKA, D. COBZAS, Z. DODDS, G. HAGER, M. JAGERSAND, K. YEREX. Recent Methods for Image-based Modeling and Rendering. IEEE Virtual Reality, 2003.
- [17] D. LIEBOWITZ. Camera Calibration and Reconstruction of Geometry from Images. PhD thesis, University of Oxford, 2001.
- [18] D. MARR. A computational investigation into the human representation and processing of visual information. Freeman, 1982.
- [19] D. NISTÉR. Automatic dense reconstruction from uncalibrated video sequences. These , Royal Institute of Technology KTH. Sweden, 2001.
- [20] D. NISTÉR. Automatic passive recovery of 3D from images and video. 3DPVT, 2004.
- [21] D. PAUL. Modeling and Rendering Architecture from Photographs. University of Michigan at Ann Arbor, 1996.

- [22] D. ROUABHIA, B. M. CHAOUKI. "Combining visual hull and stereovision techniques for new 3D reconstruction method" CCCA'11 proceeding. Tunisie, mars 2011.
- [23] E. BOYER, J. FRANCO. A hybrid approach for computing visual hulls of complex objects. In proceeding of IEEE conference on computer vision and pattern recognition, USA. 2003.
- [24] F. CHAUMETTE, P. RIVES. Modélisation et calibration d'une caméra. AFCET RFIA, 1990.
- [25] F. Devernay, O. Faugeras. "From projective to euclidian reconstruction". Research report INRIA NUMÉRO2725, 1995.
- [26] F. DORNAIKA, C. GARCIA. Robust Camera Calibration using 2D to 3D Feature Correspondences. Technical Report. GMD, 1997.
- [27] F. JEAN-SEBASTIEN, B. EDMOND. Une Approche Hybride pour Calculer l'Enveloppe Visuelle d'Objets Complexes. GRAVIR, Equipe MOVI, INRIA Rhône-Alpes.
- [28] F. PIERRE-ALEXANDRE. Vision stéréoscopique : appariements. Université Laval Canada, 2004.
- [29] F. TAILLANDIER. Reconstruction du bâti en milieu urbain: une approche multi-vue. Thèse de doctorat de l'Ecole Polytechnique, 2004.
- [30] 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, 1998.
- [31] G. VAN. MEERBERGEN, M. VERGAUWEN, M. POLLEFEYS, L. VAN. GOOL. A Hierarchical Symmetric Stereo Algorithm Using Dynamic Programming. IJCV, 2002.
- [32] H. ISHIKAWA, D. GEIGER. Occlusions, discontinuities, and epipolar lines in stereo. In ECCV, 1998.
- [33] H. Yang, S. Park, S. Lee. "Reconstruction of 3D human body pose based on top-down learning". ICIC 2005, international conference on intelligent computing, Hefei , CHINE, 2005.
- [34] I. OSEBE, S. YOU, U. NEUMANN. Rapid Part-Based 3D Modeling. Integrated Media Systems Center. Los Angeles, 2006.
- [35] J. BLANC. Synthèse de nouvelles vues d'une scène 3D à partir d'images existantes. Thèse PhD de l'Institut National Polytechnique de Grenoble, 1998.
- [36] J. Koenderink, A. Van Doorn. "Affine structure from motion". Journal of the optical society of america(A) Vol. 8, No. 2, 1991.
- [37] J. R. BERGEN. The Plenoptic Function and the Elements of Early Vision. Computational Models of Visual Proc. The MIT Press. Cambridge, 1991.
- [38] J. S. KIM, I. S. KWEON. Semi-metric reconstruction from a single image using orthogonality and parallelism, Department of Electrical Engineering and Computer Science. Korea, 2005.
- [39] L. DE LUCA. Relevé et multi-présentation du patrimoine architectural. Thèse. Ecole supérieur d'Arts et Métiers. France, 2006.
- [40] L. MCMILLAN, G. BISHOP. Plenoptic Modeling: An Image-Based Rendering System. Proc of SIGGRAPH 95. California, 1995.
- [41] M. C. FRASSON. Reconstruction interactive de scènes tridimensionnelles à partir d'images. Université de Montréal, 1998.

- [42] M. Fahiem, S. Haq, F. Saleemi, "A review of 3D reconstruction techniques from 2D orthographic line drawings". Geometric Modeling and Imaging, Zurich, 2007.
- [43] M. HAN, T. KANADE. Multiple motion scene reconstruction from uncalibrated views. Proc of the 8th International Conference on Computer Vision. Canada, 2001.
- [44] M. HAN, T. KANADE. Reconstruction of a scene with multiple linearly moving objects. Proc of the Conference on Computer Vision and Pattern Recognition. USA, 2000.
- [45] M. J. HANNAH. Computer Matching of Areas in Stereo Images. PhD thesis. Stanford University, 1974.
- [46] M. LAURENT. Acquisition 3D au service de l'industrie. Magazine électronique CiMax, 1997.
- [47] M. LEVOY, P. HANRAHAN. Light Field Rendering . Proc. SIGGRAPH 96. NewOrleans, 1996.
- [48] M. LEVOY. The Digital Michelangelo Project: 3D Scanning of Large Statues. Proc. Siggraph, ACM, 2000.
- [49] M. OLIVEIRA. Image-Based Modeling and Rendering Techniques: A Survey. Instituto de Informática. RITA, 2002.
- [50] M. POLLEFEYS, L. VAN GOOL. Visual modeling: from images to images. Center for Processing of Speech and Images. K.U.Leuven, Belgium, 2002.
- [51] M. POLLEFEYS, R. KOCH, M. VERGAUWEN, L. VAN. GOOL. Automatic Generation of 3D Models from Photographs. Proc. Virtual Systems and MultiMedia, 1998.
- [52] M. POLLEFEYS. Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences. These ESAT-PSI. Leuven, 1999.
- [53] M. R. JENKIN, A. D. JEPSON, J. K. TSOTSOS. Techniques for disparity measurement. CVGIP: Image Understanding, 1991.
- [54] M. SEITZ, C. R. DYER. View morphing. Proc of the ACM Conference on Computer Graphics. USA, 1996.
- [55] O. FAUGERAS, Q. LUONG, S. J. MAYBANK. Camera Self-Calibration : Theory and Experiments. Proc. European Conference on Computer Vision. Santa- Margerita. Italy, 1992.
- [56] P. BEARDSLEY, A. ZISSERMAN, D. MURRAY. Sequential Updating of Projective and Affine Structure from Motion. International Journal of Computer Vision, 1997.
- [57] P. DEBEVEC, C. J. TAYLOR, J. MALIK. Modeling and rendering architecture from photographs: A hybrid geometry and image-based approach. SIGGRAPH, 1996.
- [58] P. JOUVELOT. La compression de données. Cour. Ecole des mines de Paris, France, 2006.
- [59] P. STURM, S. RAMALINGAM. A generic concept for camera calibration. ECCV, 2004.
- [60] P. STURM. Structure and motion for dynamic scenes: the case of points moving in planes. Proc of the 7th European Conference on Computer Vision. Denmark, 2002.
- [61] R. ARCILA, F. HETROY, F. Dupont. Etat de l'art des méthodes de segmentation de séquences de maillages et proposition d'une classification. LIRIS - CNRS, Université de Lyon, France, 2009.
- [62] R. CHARLES. Volumetric scene reconstruction from multiple views. Foundations of image understanding, Boston. 2001.
- [63] R. HARTLEY, A. ZISSERMAN. Multiple View Geometry in Computer Vision. Cambridge University Press, second edition. England, 2004.

- [64] R. HARTLEY. Euclidean reconstruction from uncalibrated views. Applications of Invariance in Computer Vision. LNCS, Springer-Verlag, 1994.
- [65] R. HORAUD, O. MONGA. Vision par ordinateur : outils fondamentaux. Hermès, 1995.
- [66] R. HORAUD. Vision 3D projective, affine et euclidienne. Technical report, INRIA Rhône-Alpes and GRAVIR-CNRS, 2000.
- [67] R. KERIVEN. A variational framework for shape from contours. Research Report Ecole Nationale des Ponts et Chaussées. CERMICS, France, 2002.
- [68] R. ZHANG, P. S. TSAI, J. E. CRYER, M. SHAH. Shape from shading: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1999.
- [69] S. AVIDAN, A. SHASHUA. Trajectory triangulation: 3D reconstruction of moving points from a monocular image sequence. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000.
- [70] S. CORNOU, M. DHOME, P. SAYD. Reconstruction tridimensionnelle de scènes structurées par des contraintes géométriques. Université Blaise Pascal- Clermont-Ferrand.
- [71] S. E. CHEN, L. WILLIAMS. View interpolation for image synthesis. Computer Graphics, 1993.
- [72] S. GORTLER. The Lumigraph. Proc. SIGGRAPH 96. New Orleans, 1996.
- [73] S. HEUNG-YEUNG SHUM, H. LI-WEI. Rendering with concentric mosaics. In Proceedings of the 26th annual conference on Computer graphics and interactive techniques, Addison-Wesley. 1999.
- [74] S. JIAO-YING. Image Based Rendering And Modeling Techniques And Their Applications. State Key laboratory of Computer Aided Design and Graphics. China, 2003.
- [75] S. ROY, I. J. COX. A Maximum-Flow Formulation of the N-camera Stereo Correspondence Problem. ICCV, 1998.
- [76] S. ROY. Stereo Without Epipolar Lines : A Maximum-Flow Formulation. IJCV, 1999.
- [77] S. T. BARNARD, M. A. FISCHLER. Computational stereo. ACM Comp. Surveys, 1982.
- [78] S. ZHENG, Z. ZHANA, Z. ZHANG. A flexible and automatic 3d reconstruction method. School of Remote Sensing Information Engineering. China.
- [79] T. BEIER, S. NEELY. Feature based image metamorphosis. Proc of the 19th Annual ACM Conference on Computer Graphics and Interactive Techniques. Chicago, 1992.
- [80] T. CLAUDE. Initiation à la modélisation et à la programmation 3D. Ellipses, 2007.
- [81] T. WERNER, A. ZISSERMAN. Model selection for automated reconstruction from multiple views. British Machine Vision Conference. UK, 2002.
- [82] U. R. DHOND, J. K. AGGARWAL. Structure from stereo: review. IEEE Trans, 1989.
- [83] X. Decoret, G. Schaufler, F. Sillion, J. Dorsey. Multi-layered imposter for accelerated rendering. Computer Graphics Forum (Eurographics'99), 1999.
- [84] Y. ALOIMONOS, I. WEISS, A. BANDYOPADHYAY. Active vision. Proc. ICCV, 1987.
- [85] Y. HERRY, K. ANJYO, K. ARAI. Tour into the picture: Using a spidery mesh interface to make animation from a single image. Proc of the ACM SIGGRAPH Conference. Los Angeles, 1997.
- [86] Y. MA, AL. An Invitation to 3D Vision: From Images to Geometric Models. Springer Verlag, 2003.

- [87] Y. MATSUMOTO, K. FUJIMURA, T. KITAMURA. Shape-from-silhouette/stereo and its application to 3-d digitizer. 1999.
- [88] Z. BROWN. Advances in Computational Stereo. IEEE transactions on pattern analysis and machine intelligence. Brown, 2003.