



République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université de Larbi Tébessi –Tébessa-
Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie
Département : Mathématiques et informatique



Mémoire

Présenté en vue de l'obtention du diplôme MASTER

Filière : Informatique
Option : Système et multimédia

Thème :

Une Méthode Hybride Pour La Reconstruction 3D à Partir d'une Vidéo

Présenté par :

HEDHOUD Yousra

TOUAHRIA Oussama

Devant le jury :

BENOUR Akram MCB Président

MENASSEL Rafik MAA Examineur

ROUABHIA Djaber MAB Encadreur

2016/2017

Remerciements

*Nous tenons tout d'abord à remercier Allah le tout puissant et miséricordieux,
qui nous a donné la force et la patience d'accomplir ce travail.*

En second lieu nous tenons à remercier notre encadreur Mr : Rouabhia Djaber,

Pour ces précieux conseils et son aide durant toute la période du travail.

*Nos vifs remerciements vont également aux membres du jury pour l'intérêt
qu'ils ont porté à notre travail en acceptant de l'examiner.*

*Enfin, nous tenons également à remercier toutes les personnes qui ont participé
de près ou de loin à la réalisation de ce travail.*

Dédicaces

Tous les mots ne sauraient exprimer la gratitude, l'amour, le respect, la reconnaissance, c'est tout simplement que : Je dédie ce modeste travail :

A Ma tendre Mère : Tu représentes pour moi la source de tendresse et l'exemple de dévouement qui n'a pas cessé de m'encourager. Tu as fait plus qu'une mère puisse faire pour que ses enfants suivent le bon chemin dans leur vie et leurs études.

A Mon très cher Père : Aucune dédicace ne saurait exprimer l'amour, l'estime, le dévouement et le respect que j'ai toujours pour vous. Rien au monde ne vaut les efforts fournis jour et nuit pour mon éducation et mon bien être. Ce travail est le fruit de tes sacrifices que tu as consentis pour mon éducation et ma formation le long de ces années.

A Mes frères et sœurs qui n'ont cessé d'être pour moi des exemples de persévérance, de courage et de générosité.

A tous mes chers amis proches, Leur générosité et leur soutien durant toute l'élaboration du travail m'oblige de leur témoigner.

A tous les membres de ma promotion.

A tous mes enseignants depuis mes premières années d'études.

A tous ceux qui me sens chers et que j'ai omis de citer.

Dussama

Dédicaces

Je dédie ce mémoire

A mes très chers parents,

Pour leurs sacrifices, leur amour, leurs prières et leurs soutiens

A tous mes amis,

Sans eux la vie me semblerait bien fade, je vous souhaite la prospérité et le succès.

A tous mes enseignants avec ma profonde considération, qui n'ont épargné aucun effort pour m'offrir un bon enseignement.

Et tous ceux qui nous ont assisté, dans la réalisation et le bon déroulement de ce travail

Yousra

Table des matières

TABLE DES MATIERES	I
LISTE DES TABLEAUX	II
LISTE DES FIGURES	III
RESUME	IV
ABSTRACT.....	V
المخلص.....	VI
INTRODUCTION GENERALE	1
PROBLEMATIQUE.....	2
MOTIVATION.....	3
ORGANISATION DU MEMOIRE.....	4
NOTIONS DE BASE	5
INTRODUCTION.....	5
1. SYSTEMES DE COORDONNEES.....	5
2. TRANSFORMATIONS USUELLES.....	6
3. MODELISATION DE LA CAMERA.....	7
4. FORMATION DE L'IMAGE.....	9
5. RELATIONS GEOMETRIQUES ENTRE LES IMAGES.....	10
6. RELATION ENTRE VIDEO ET IMAGES.....	12
CONCLUSION.....	12
PRELIMINAIRE.....	13
INTRODUCTION.....	13
1. METHODES DE MODELISATION ET RENDU A BASE D'IMAGE : IBMR.....	13
2. QUELQUES APPROCHES D'IBMR.....	15
3. CLASSIFICATION MULTICRITERES DES METHODES DE RECONSTRUCTION 3D.....	19
CONCLUSION.....	28
ENVELOPPE VISUELLE ET STEREOVISION	29
INTRODUCTION.....	29
1. L'ENVELOPPE VISUELLE.....	29
2. LA STEREOVISION.....	33
CONCLUSION.....	46
TRAVAUX SIMILAIRES	47
INTRODUCTION.....	47
2. TRAVAIL DE ERIC NOWARK ET AL.....	47
3. TRAVAIL DE MING LI ET AL.....	50
4. TRAVAIL DE MATSUMOTO ET AL.....	52
5. CONCLUSION.....	59
UNE METHODE HYBRIDE POUR LA RECONSTRUCTION 3D A PARTIR D'UNE VIDEO	60
1. INTRODUCTION.....	60
2. SYSTEME D'ACQUISITION DE LA VIDEO.....	60
3. MATERIELS ET LANGAGE DE PROGRAMMATION.....	61
4. DEMARCHE PROPOSEE DU PROCESSUS DE RECONSTRUCTION 3D.....	62
5. CONCLUSION.....	79
CONCLUSION GENERALE ET PERSPECTIVES	60
REFERENCES	82

Liste des Tableaux

Notions de base

Tableau [II.1] : Critères de classification des méthodes de reconstruction 3D. 19

Travaux Similaires

Tableau [IV.1] : Analyse de temps d'exécution de chaque étape du processus.....58

Une Méthode Hybride pour la Reconstruction 3D à partir d'une Vidéo

Tableau [V.1] : Différentes notions des fichiers obj..... 71

Tableau [V.2] : Tableau comparatif de temps de calcul de stéréo vision avec stéréovision limitée 76

Liste des figures

Notions de base

Figure [I.1] : Système de coordonnées.	5
Figure [I.2] : Les trois transformations élémentaires du modèle sténopé, et les repères associés.	8
Figure [I.3] : Modèle de la caméra.	8
Figure [I.4] : Formation de l'image.	10
Figure [I.5] : Géométrie épipolaire	10
Figure [I.6] Relation géométrique entre deux image : Géométrie épipolaire.	11
Figure [I.7] : Relation géométrique entre trois images : Trilinéarité.	11

Préliminaire

Figure [II.1] : Modélisation basée images.	14
Figure [II.2] : Rendu basée image.	14
Figure [II.3] : Résultat de <i>shape from shading</i>	15
Figure [II.4] : <i>Shape from silhouette</i> à partir 3 vues différentes.	16
Figure [II.5] : (a) Texture extraite d'une robe, (b) surface reconstruite de la texture.	16
Figure [II.6] : Forme à partir de focus/defocus.	17
Figure [II.7] : Panneau d'affichage rendu.	18
Figure [II.8] : Imposteur. A gauche, l'objet original, à droite, l'imposteur.	18
Figure [II.9] : Exemple d'une voiture reconstruite automatiquement.	20
Figure [II.10] : Reconstruction semi-automatique : Exemple de détection des vertices et arrêtes pour obtenir le modèle final.	20
Figure [II.11] : Exemple de reconstruction 3D d'une ville.	21
Figure [II.12] : Modèle 3D d'un bâtiment service publiques.	21
Figure [II.13] : Modèle de reconstruction 3D d'une petite scène.	22
Figure [II.14] : Télémètre Laser.	23
Figure [II.15] : Lumière structurée illuminant un visage.	24
Figure [II.16] : Simple exemple du système de stéréovision.	24
Figure [II.17] : Exemple de reconstruction mono-objet.	25
Figure [II.18] : Regroupement des éléments reconstruits.	25
Figure [II.19] : Résultat d'une reconstruction projective.	26
Figure [II.20] : Reconstruction euclidienne.	27
Figure [II.21] : Résultat d'une reconstruction affine.	27
Figure [II.22] : Résultat d'une reconstruction métrique.	28

Enveloppe visuelle et stéréovision

Figure [III.1] : Un objet 2D représenté avec une vue, sa silhouette et le cône correspondant.	29
Figure [III.2] : v_∞ et V_r pour un objet 2D : v_∞ est l'enveloppe convexe de l'objet.	30
Figure [III.3] : Exemple d'objet avec une zone concave.	30
Figure [III.4] : Exemple d'objet avec fantôme.	31
Figure [III.5] : Illustration d'une coupe de la géométrie estimée d'un objet O, en rouge enveloppe visuelle surfacique et bleue une enveloppe visuelle volumique.	32

Figure [III.6] : (a) Géométrie épipolaire quelconque. (b) Géométrie épipolaire rectifiée.	33
Figure [III.7] : Modèle géométrique de prise de vues.	34
Figure [III.8] : Vecteur de disparité.	36
Figure [III.9] : Exemple de carte de disparité.	36
Figure [III.10] : Zone occultée – Le pixel marqué d’un point rouge a son correspondant occulté dans l’image de droite-	37
Figure [III.11] : Zone de discontinuité de profondeur – Les deux pixels d’un point rouge se correspondent mais leurs voisinages diffèrent car ils sont proches d’une discontinuité de profondeur-.	38
Figure [III.12] : Zone homogène –Manque d’information qui caractérise le pixel-.	38
Figure [III.13] : Zone avec textures répétitives –Pour un pixel il existe plusieurs candidats dans l’autre image.	38
Figure [III.14] : Géométrie épipolaire.	39
Figure [III.15] : Exemple de la rectification épipolaire.	39
Figure [III.16] : Un point image correspond à un point objet est un seul.	40
Figure [III.17] : Contrainte d’ordre.	41
Figure [III.18] : Contrainte de continuité figurale.	41
Figure [III.19] : Mise en correspondance locale.	43
Figure [III.20] : Système de projection des scène –Représentation de triangle formé par le système.-.	45

Travaux similaires

Figure [IV.1] : Exemple des résultats obtenus par l’approche de Nowark et al.	49
Figure [IV.2] : Exemple de modélisation des zones concaves.	49
Figure [IV.3] : Trois paires de caméras sont disposées le long d’un arc.	50
Figure [IV.4] : Démarche de l’approche proposée. (a) : Extraction des silhouettes. (b) : Création d’enveloppe visuelle. (c) Estimation de profondeur à partir la stéréovision.	50
Figure [IV.5] : Résultat de modélisation de la méthode de Li et al (Exemple 1).	51
Figure [IV.6] : Résultat de modélisation de la méthode de Li et al (Exemple 2).	52
Figure [IV.7] : Shape from silhouette basé vote.	54
Figure [IV.8] : Vote sur l’espace voxel : la densité montre le score de chaque voxel.	54
Figure [IV.9] : (a) Estimation de surface partielle. (b) Génération totale de modèle.	55
Figure [IV.10] : (a) Estimation de surface partielle (b) Génération totale du modèle.	55
Figure [IV.11] : Présentation du système.	57
Figure [IV.12] : Exemples d’images originales et de résultats de modélisation (Ecureuil)	57
Figure [IV.13] : Exemples d’images originales et de résultats de modélisation (Château).	57

Une Méthode Hybride pour la Reconstruction 3D à partir d’une Vidéo

Figure [V.1] : Exemple d’un mini studio.	61
Figure [V.2] : Interface graphique du MATLAB.	62
Figure [V.3] : Quelques images extraites (Objet 1).	63
Figure [V.4] : Quelques images extraites (Objet 2).	63

Figure [V.5] : De gauche en droite : L'image originale, image prétraitée et la silhouette obtenue (En haut Objet 1 et en bas Objet 2).....	65
Figure [V.6] : Organigramme de l'algorithme extraction des silhouettes.	66
Figure [V.7] : Symétrie inversée de deux silhouettes.....	67
Figure [V.8] : Représentation des limites supérieure et inférieure.	67
Figure [V.9] : De gauche en droite la profondeur de la ligne (i) calculée à partir de la partie gauche et droite.	69
Figure [V.10] : Résultats d'enveloppe visuelle obtenus par MATLAB.....	71
Figure [V.11] : Importation d'objet reconstruit en Blender.	72
Figure [V.12] : Cartes de disparité résultantes des objets utilisées.....	76
Figure [V.13] Courbes comparatives des temps de calculs : (en rouge, le temps d'exécution avec limitation), (en bleu, le temps d'exécution sans limitation).	77
Figure [V.14] : Diagramme circulaire représente le pourcentage de minimisation de temps de calcul d'algorithme proposée.....	78
Figure [V.15] : Diagramme de secteur représente le temps de calcul écoulé par la stéréovision et la stéréovision limitée.	78

Résumé

Dans les dernières décennies, le domaine de la reconstruction 3D a évolué rapidement et de nombreuses techniques ont été introduites dans la communauté scientifique. La stéréovision et l'enveloppe visuelle sont quelques-unes de ces techniques et les deux sont considérées comme des méthodes IBMR.

L'enveloppe visuelle utilise des silhouettes d'objets extraits d'images prises de différents points de vue, pour obtenir un modèle 3D approximatif de cet objet. Et la stéréovision, est une technique qui estime la troisième dimension (la profondeur), en commençant par deux images de la même scène prises à partir de deux positions différentes, et en calculant les pixels correspondants dans les deux images pour obtenir la carte de disparité qui est utilisée pour calculer la profondeur de cette scène.

Ces deux méthodes ont des inconvénients lorsqu'elles sont appliquées seules, alors l'objectif de cette étude est de les combiner pour surmonter leur problèmes et de proposer une nouvelle méthode de reconstruction 3D robuste et performante.

Dans notre étude, nous avons proposé une nouvelle approche pour estimer l'enveloppe visuelle, nous avons travaillé avec des images extraites d'une vidéo capturée d'un objet réel, afin de travailler avec un grand nombre de silhouettes, ce qui nous a permis d'extraire une forme 3D très approximative, et à partir de ce résultat, nous avons généré un fichier "OBJ", qui peut être ouvert et manipulé avec de nombreux logiciels de modélisation.

Nous avons utilisé les informations géométriques obtenues à partir du résultat de l'enveloppe visuelle pour réduire l'espace de recherche dans la stéréovision. Nous avons, d'une part, minimisé son temps d'exécution plus que la moitié de la valeur original, et d'une autre part, la carte de disparité qui est le résultat de la stéréovision est alors appliquée à la forme 3D initiale obtenue par la technique de l'enveloppe visuelle pour corriger les mauvaises régions estimées.

Mots clés : Reconstruction 3D, IBMR, enveloppe visuelle, stéréovision et mise en correspondance.

Abstract

In the last decades, the field of 3D reconstruction has evolved rapidly and many techniques were introduced to the scientific community. Stereovision and Visual Hull are some of those techniques and both of them are considered as IBMR methods.

The visual hull uses silhouettes of objects extracted from images taken from different viewpoints to obtain an approximate 3D model of that object. And the stereovision, is a technique that estimates the third dimension (which is the depth), by starting with two images of the same scene taken from two different positions, and calculating the corresponding pixels in the two images to obtain the disparity map, which then will be used to calculate the depth of that scene.

These two methods have some disadvantages when applied alone, so the aim of this study is to combine them together to overcome their problems and propose a robust and performant new 3D reconstruction method.

In our study, we introduced a new approach to estimate the visual hull, and we worked with images extracted from a video taken of a real object, in order to work with a big number of silhouettes, which allowed us to extract a very approximate 3D form, and from this visual hull result we generated an “OBJ” file, which can be opened and manipulated with many modelling softwares.

We used geometric information obtained from the result of visual hull to narrow the space of research in the stereovision. We minimized its execution time to more than half the value, this from a part. And in another part, the disparity map which is the result of the stereovision is then applied on the initial 3D form obtained by the visual hull technic to correct the bad estimated regions.

Key words: 3D reconstruction, IBMR, visual hull, stereovision, bloc matching.

المُلخَص

في العقود الأخيرة، تطور مجال إعادة النمذجة ثلاثية الأبعاد بسرعة وتم إدخال العديد من التقنيات إلى المجتمع العلمي. "الرؤية البصرية المزدوجة" و"الحيز البصري" هي بعض تلك التقنيات وكلاهما يعتبر من طرق استخدام الصور للتصميم و التغليف والتولين.

الحيز البصري" هي طريقة تستخدم صور ظليلة للمجسم المراد إعادة نمذجته، مستخرجة من الصور المأخوذة من وجهات نظر مختلفة للحصول على نموذج ثلاثي الأبعاد تقريبي لهذا المجسم. و "الرؤية البصرية المزدوجة"، هي طريقة أخرى لتقدير البعد الثالث (العمق)، من خلال العمل على صورتين لنفس المشهد مأخوذة من منظورين مختلفين، وحساب النقاط الصورية المتماثلة في الصورتين للحصول على خارطة التباين، والتي ثم سيتم استخدامها لحساب عمق هذا المشهد.

وهاتان الطريقتان لهما بعض العيوب عند تطبيقهما منفردتين، لذلك فإن الهدف من هذه الدراسة هو الدمج بينهما معا للتغلب على هذه المشاكل واقتراح طريقة إعادة نمذجة جديدة وذات أداء أكثر فعالية.

في دراستنا، قدمنا طريقة جديدة لتقدير الحيز البصري، واستخدمنا صور مستخرجة من فيديو مأخوذ لمجسم حقيقي، من أجل العمل مع عدد كبير من الصور الظلية، مما سمح لنا باستخراج شكل ثلاثي الأبعاد تقريبي جدا، ومن نتيجة الحيز البصري أنشأنا ملف "obj" ، يمكن فتحه و معالجته بالعديد من برامج النمذجة ثلاثية الأبعاد.

استخدمنا المعلومات الهندسية التي تم الحصول عليها من نتيجة الحيز البصري لتضييق مساحة البحث في "الرؤية البصرية المزدوجة"، قللنا من وقت المعالجة إلى أكثر من نصف القيمة، هذا من جهة. و من جهة أخرى، استخدمنا خريطة التفاوت التي هي نتيجة "الرؤية البصرية المزدوجة" على الشكل الأولي الذي تحصلنا عليه من الحيز البصري لتصحيح المناطق الغير مقدره جيدا.

الكلمات المفتاحية: إعادة النمذجة ثلاثية الأبعاد، "الرؤية البصرية المزدوجة"، الحيز البصري.

Introduction générale

La vision artificielle a pour but de reproduire certaines fonctionnalités de la vision humaine à travers l'analyse des images. Pour un ordinateur c'est un problème assez délicat puisque l'information disponible dans ce cas est une représentation bidimensionnelle d'un environnement tridimensionnel - la troisième dimension est perdue au cours du processus de formation d'image par projection -.

Malgré le fait que les images sont utilisées dans différents domaines (conception, perception, aide à la décision, communication, robotique...etc.), comme étant un véritable support portant de multiples informations ; la projection 3D-2D entraîne en effet une perte importante d'informations. L'inférence de la troisième dimension est ardue et nécessite une compréhension approfondie du phénomène de vision. La littérature est riche en travaux distincts et méthodes quasiment différentes, et plusieurs projets de recherche sont en cours dans ce contexte.

L'analyse d'images regroupe plusieurs disciplines que l'on classe en deux catégories [1]:

- Les processus de bas-niveaux, qui nécessitent très peu d'informations sur le contenu des images. Il s'agit ici, des processus de filtrage, d'amélioration et de restauration d'images, il est regroupé sous le terme traitement d'images, ainsi que d'extraction d'indices.
- Les processus de haut-niveaux, qui fonctionnent en aval de ceux de bas niveaux, et qui peuvent nécessiter des informations sur le contenu des images. Il s'agit des processus cognitifs de façon générale, de la reconstruction tridimensionnelle, la reconnaissance de formes...

Notre mémoire intitulé « Une méthode hybride pour la reconstruction 3D à partir de vidéo » rentre dans le cadre de la deuxième catégorie. L'objectif est de surmonter les problèmes dus à la perte d'informations, et ce dans le but de reconstruire en 3D un objet réel capturé en vidéo dans des délais satisfaisants et sans affecter la qualité du résultat obtenu.

La reconstruction 3D représente un domaine actif depuis de nombreuses années, elle vise à générer un modèle 3D à partir d'informations issues de scènes réelles, les recherches s'appuient sur l'obtention d'une forme assez réaliste de l'objet. Le principe est de déterminer les coordonnées 3D de n'importe quel point d'un objet.

De multiples méthodes sont disponibles pour la reconstruction 3D, elles se différencient selon le type d'objet à reconstruire (forme, dimension, texture ...), la situation de l'environnement (éclairage, scène structurée...). Il existe multiples critères de classification de ces méthodes dont les détails seront vivement présentés dans la démarche de notre travail.

L'une des difficultés liées à la reconstruction 3D ; est quelle fait appel à plusieurs axes de recherches tels que : Le traitement d'images, l'isolation d'un objet de son environnement, le calibrage de caméras, la mise en correspondances, la triangulation, la vision par ordinateur... etc. A cet effet, une grande masse de connaissances théoriques doit être acquise et bien saisie afin d'aboutir à la résolution automatique d'un tel problème.

Dans le cadre de ce mémoire, nous visons à combiner deux méthodes très célèbres ; à savoir « l'enveloppe visuelle » et « la stéréovision ». Qui seront appliquées à un objet acquis par vidéo, dans le but d'obtenir son modèle 3D de qualité visuelle très proche de la réalité dans un délai acceptable et réduit.

L'idée de l'enveloppe visuelle [2] fut introduite pour la première fois en 1974 par *Baumgart* dans sa thèse de doctorat ; son principe est d'obtenir une représentation 3D approximative de l'objet à partir d'un ensemble de silhouettes (Shape from silhouette) prises par différentes caméras à partir de différents points de vue, l'intersection de ces silhouettes forme un cône visuel. Il existe de nombreuses méthodes pour calculer l'enveloppe visuelle d'un objet et elles seront survolées dans le cadre de notre travail.

La stéréovision [3] est le procédé d'estimation de la profondeur à partir de deux images de la même scène prises de deux positions différentes, en calculant la correspondance entre les pixels des deux images. Ces derniers subissent un petit décalage qui peut être utilisé pour calculer la « Disparité ». Si on arrive à donner une disparité à chaque pixel d'une image, on peut faire de même pour tous les points de l'autre image. Ces disparités seront utilisées pour la triangulation [4] qui permet de calculer les profondeurs dans l'espace de tous les pixels appariés, donc de reconstruire la forme 3D de l'objet photographié.

Problématique

La Modélisation 3D est l'une des approches de synthèse d'images dites classiques, le processus de création de la forme 3D dans ce cas suit deux phases essentielles :

- **Phase de modélisation 3D :**

Elle consiste à créer d'une façon manuelle et à l'aide d'un logiciel de modélisation 3D des objets simples ou complexes, qui permet de dessiner carrément un modèle 3D ou de le construire à partir de formes primitives simples telles que des sphères, des polygones, des cônes, des courbes ...etc. Elle peut aussi faire intervenir des outils plus complexes.

- **Phase de rendu :**

La phase de modélisation est suivie par une phase de rendu qui est un processus créatif qui ressemble à la photogrammétrie¹ ou la cinématographie². Les objets obtenus par modélisation sont recouverts par la suite avec des couleurs ou des textures [5]. Il est même possible d'ajouter des sources lumineuses, d'ombres, de l'atmosphère, ou des caméras virtuelles pour que les résultats semblent plus proches de la réalité. On peut aussi appliquer des effets optiques tels que la réfraction de la lumière ou du mouvement flou sur les objets en mouvement.

Cette phase implique beaucoup de calculs et prend beaucoup de temps, même sur des ordinateurs puissants, puisqu'on applique un effet particulier pour chaque pixel de l'objet.

Malgré que la modélisation 3D donne des résultats très satisfaisants pour les objets synthétiques, elle est moins robuste pour les scènes réelles qui contiennent beaucoup de détails à modéliser comme par exemple le cas d'une forêt ou d'une ville. Ces méthodes sont gourmandes en temps de calcul : Le calcul d'une seule image peut prendre un temps très important malgré que les machines d'aujourd'hui soient devenues de plus en plus puissantes.

¹ **Photogrammétrie** : Technique qui vise à effectuer des mesures dans une scène, en utilisant la parallaxe obtenue entre des images acquises selon des points de vue différents.

² **Cinématographie** : Ensemble des méthodes mis en œuvre pour obtenir la reproduction du mouvement par l'intermédiaire d'un film.

Cette modélisation est utilisée généralement pour mettre en situation des objets avant leur fabrication. Dans tous les cas, ce genre d'approches nécessite une phase d'apprentissage longue et fastidieuse et doit être réalisée par des experts et spécialistes qui maîtrisent fortement les logiciels de modélisation et fait intervenir plusieurs artistes dans ce domaine. De plus, les durées y afférentes sont trop grandes, sans parler de la qualité des objets obtenus qui sont synthétiques et clairement artificiels.

Les problèmes liés à la modélisation 3D limitent leurs utilisations, et l'alternative consiste à effectuer des recherches sur d'autres types de modélisation 3D comme les méthodes d'IBMR, dont on peut citer en particulier la reconstruction 3D à partir d'images réelles, qui englobe un vaste éventail de théories et de techniques.

Motivation

Plusieurs méthodes de reconstruction 3D existent et chacune d'entre elles est appliquée à un type particulier de scènes. L'une de ces méthodes combine le principe de la stéréovision avec celui de l'enveloppe visuelle. L'enveloppe visuelle est utile seulement pour les objets avec des zones convexes, elle ne peut pas traiter des zones concaves [6]. La stéréovision est instable est délicate pendant le calcul des points correspondants, surtout pour les surfaces non texturées et les zones occluses [7]. La combinaison de ces deux méthodes permet de limiter les inconvénients de chacune d'elles.

Dans ce mémoire, nous proposons une approche de reconstruction 3D qui se base sur le principe de l'hybridation de ces deux méthodes. Appliquée sur une vidéo capturant un objet en rotation, nous espérons que l'approche que nous développons permettra de reconstruire l'objet 3D dans délais acceptables et avec une qualité visuelle de haut niveau.

La méthode que nous proposons utilise uniquement les informations présentes dans l'ensemble des images extraites de la vidéo en vue d'obtenir un modèle 3D de l'objet d'une façon complètement automatique (sans aucune intervention humaine). L'idée générale est d'utiliser l'enveloppe visuelle pour -en premier lieu- limiter au maximum l'espace englobant l'objet, et de la faire suivre par une étape de stéréovision qui permettra de raffiner le modèle reconstruit par la détection des régions concaves sur la surface de l'objet.

Le résultat de l'enveloppe visuelle permet de réduire les probabilités de mises en correspondance incorrectes obtenues lors de la phase de la stéréovision. Cette combinaison permet aussi de réduire le temps de calcul nécessaire pour reconstruire un objet.

Nous proposons dans le cadre de notre travail une nouvelle technique pour l'estimation de l'enveloppe visuelle. Les résultats obtenus sont très satisfaisants pour des objets simples et elles donnent un maillage très proche de la forme réelle de l'objet reconstruit. Les temps de calcul sont très réduits et nous pensons qu'ils peuvent même être améliorés pour devenir temps réel.

Organisation du mémoire

La reconstruction 3D à partir d'images à base d'une hybridation de deux méthodes différentes de synthèse d'images est un sujet souvent étudié. Une large bibliographie qui traite les différentes notions de base reliées à ce domaine est accessible. Pour notre travail, nous nous intéressons en particulier à la reconstruction 3D à partir de vidéo.

Dans le reste de ce mémoire nous essayons d'introduire ce domaine de recherche. La description du contenu de chaque chapitre est comme suit :

Dans le chapitre I ; nous allons présenter les différentes notions de base nécessaires pour aborder notre travail.

Au cours du chapitre II ; nous passons en revue dans sa première partie les méthodes les plus connues de la modélisation 3D à partir d'images. Et dans sa seconde partie nous présentons une classification multi critères des différentes méthodes de reconstruction 3D.

Le chapitre III traite les notions préliminaires liées à l'enveloppe visuelle et la stéréovision, ainsi que quelques techniques existantes pour réaliser chacune de ces deux approches.

Dans le chapitre IV nous allons présenter quelques travaux similaires qui se basent sur cette combinaison.

Notre objectif dans ce travail est de combiner deux techniques d'IBMR : l'enveloppe visuelle et la stéréovision. Dans le dernier chapitre (chapitre V), nous proposons une nouvelle approche de reconstruction 3D à partir de vidéo et qui combine l'enveloppe visuelle et la stéréovision. Nous proposons un algorithme parallèle de calcul de l'enveloppe visuelle. Nous proposons, par la suite, de limiter la zone de recherche des correspondants (par l'enveloppe visuelle) avant d'appliquer une phase de stéréoscopie. Ceci doit permettre de minimiser considérablement le temps d'exécution et d'améliorer la qualité du résultat final.

Avant de passer aux résultats expérimentaux, nous expliquerons le système de prise de vues réalisé de notre part et utilisé pour la capture d'un objet réel en rotation par angle fixe.

Nous présentons, par la suite, les résultats expérimentaux obtenus par l'application de notre approche.

Nous terminons ce travail par une conclusion et un ensemble de perspectives qui seront nos buts de recherche au futur.

Chapitre I

Notions de base

Introduction

La vision par ordinateur sert à rendre une machine capable de réaliser des tâches semblables à la vision humaine. Le processus de la vision se base sur deux phases importantes, la première est la phase d'acquisition des images, qui sont généralement prises par des caméras et représentées par une matrice de deux dimensions où chaque élément de cette matrice est appelé pixel¹. La deuxième phase est le traitement de ces images.

Dans notre travail nous nous intéressons à un traitement d'images de haut niveau dont le but est de récupérer la troisième dimension perdue au cours de la capture des images. Il s'agit de la reconstruction 3D qui est un domaine très actif. De nombreuses méthodes sont proposées pour arriver à cet objectif, chacune est dédiée à un type particulier d'objet à reconstruire. Dans ce mémoire nous étudions la reconstruction 3D à partir d'images.

Ce chapitre est consacré pour détaillé quelques notions de base liées à ce domaine de recherche. Nous allons voir comment on peut passer d'une scène réelle à une scène représentée dans une image à l'aide d'une projection, cette projection est faite à l'aide de caméras. Nous allons présenter le modèle de caméra le plus connu ainsi que la formation d'images. Et comme le but de notre recherche est de reconstruire la forme 3D à partir d'une vidéo, nous allons présenter la relation entre la vidéo et les images.

1. Systèmes de coordonnées

Il existe plusieurs systèmes de coordonnées qui permettent de représenter la position de la caméra et de l'objet, cette position est précisée à l'aide d'un référentiel. Trois types de référentiels différents existent [8] :

Référentiel « Monde » lié à la scène $R_s (O_s, X, Y, Z)$.

Référentiel « Caméra » $R_c (O_c, x, y, z)$.

Référentiel « Image » $R_i (O_i, u, v)$.

Dans le cas le plus général la position de la caméra et de l'objet est obtenue dans le système R_c .

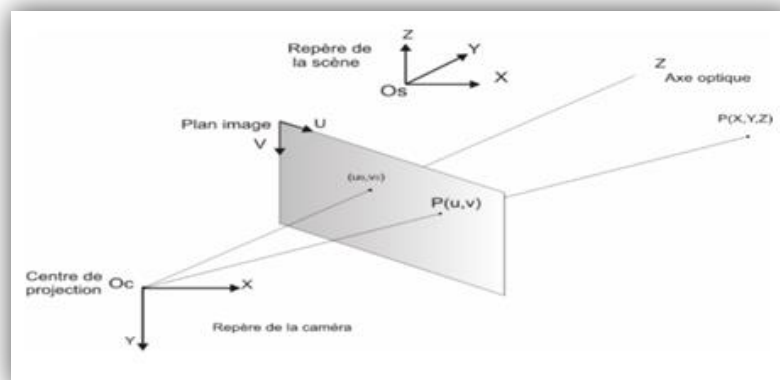


Figure [I.1] : Système de coordonnées.

¹ **Pixel** : « élément d'image », dont le nom provient de la locution anglaise picture element.

1.1 Coordonnées homogènes

Les systèmes de coordonnées homogènes [9] sont utilisés pour faire une transformation entre deux référentiels différents. Le principe de base est d'utiliser un vecteur de $n+1$ coordonnées dans un espace de représentation de dimension n ; n coordonnées plus une coordonnée supplémentaire initialisée à 1 utilisée pour représenter un vecteur de position ou bien initialisée à 0 pour représenter un vecteur de déplacement invariant pour la translation (Le zéro désigne le point à l'infini).

Les coordonnées homogènes sont représentées par : $[x, y, z, h]$.

Tels que : x, y, z représentent la position d'un point dans un espace 3D et h est la direction du point vers l'infini.

Le passage des coordonnées usuelles vers les coordonnées homogènes est réalisé par :

$[x, y, z, h] \rightarrow [x/h, y/h, z/h]$ Si $h \neq 0$.

$(x, y, z) \rightarrow [x, y, z, 1]$

2. Transformations usuelles

Les transformations usuelles sont généralement des translations, des rotations et des changements d'échelle, elles sont représentées sous forme de matrices.

2.1 Translation

La translation est une transformation géométrique qui permet de glisser et de décaler l'objet sans changer sans angle ou ses dimensions. Cette transformation est obtenue par la matrice suivante :

$$T(d_x, d_y, d_z) = \begin{pmatrix} 1 & 0 & 0 & d_x \\ 0 & 1 & 0 & d_y \\ 0 & 0 & 1 & d_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Tel que : d_x, d_y et d_z sont les coordonnées en axe x, y et z respectivement.

2.2 Changement d'échelle

C'est une transformation linéaire qui augmente ou réduit le volume d'un objet par un facteur. Le résultat d'une mise à l'échelle est similaire à la forme originale (Dans le sens géométrique). Cette transformation est obtenue par la matrice suivante :

h_x, h_y et h_z sont les facteurs de mouvement selon les axes x, y, z respectivement.

$$H(h_x, h_y, h_z) = \begin{pmatrix} h_x & 0 & 0 & 0 \\ 0 & h_y & 0 & 0 \\ 0 & 0 & h_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

2.3 Rotation

La rotation permet de tourner un objet dans l'espace sur l'un des axes. Les matrices suivantes représentent respectivement la formule de rotation sur les axes x, y et z avec un angle θ :

$$R_x(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) & 0 \\ 0 & \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$R_y(\theta) = \begin{pmatrix} \cos(\theta) & 0 & \sin(\theta) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$R_z(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 & 0 \\ \sin(\theta) & \cos(\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

3. Modélisation de la caméra

Le modèle de la caméra constitue plusieurs transformations (Figure I.3) : La transformation entre repère monde et repère caméra, la projection du repère caméra vers le repère rétine et la transformation entre le repère rétine et le repère image [10].

3.1 Transformation scène-Caméra

Le processus de transformation du repère monde et le repère de la caméra peut être décrit d'une manière très simple si on utilise les coordonnées homogènes. Dans ce cas les translations, les rotations et les changements d'échelle sont représentées par de simples produits matriciels.

Une transformation pour un point $M : P_M(x_M, y_M, z_M, 1)^T$ dans l'espace est décrite par :

$$P_c = T_M^c \cdot P_M$$

Tel que le point P_c de coordonnées $(x_c, y_c, z_c, 1)^T$ est exprimé dans le repère de la caméra.

La matrice T_M^c de transformation entre les deux repères a la forme suivante :

$$T_M^c = \begin{pmatrix} R_M^c & t_M^c \\ 0 & 1 \end{pmatrix}$$

3.2 Projection caméra-Rétine

La projection de la caméra vers la rétine est une projection perspective¹. Les algorithmes de vision s'appuient généralement sur le modèle sténopé qui est connu aussi par « modèle en trou », ce modèle transforme un point 3D de l'espace M en un point image x et peut se décomposer en trois transformations élémentaires successives (Voire figure I.2) :

- 1) Transformation entre repère monde et celui de la caméra.
- 2) Transformation entre repère de la caméra et le repère (Plan rétinien).
- 3) Transformation entre repère capteur et le repère d'image.

² **Projection perspective** : Transformation géométrique réalisée par un système optique qui forme l'image d'objets situés dans l'espace. Cette transformation est utilisée pour représenter des objets sur une image.

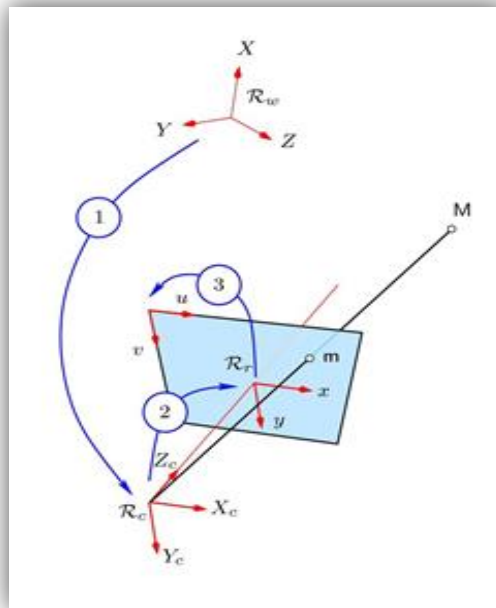


Figure [L.2] : Les trois transformations élémentaires du modèle sténopé, et les repères associés.

3.3 Transformation Rétine-image

L'image est une représentation bidimensionnelle (u, v) qui se forme d'un ensemble de pixels. La transformation du repère caméra au repère image s'écrit par :

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} k_u & 0 & 0 \\ 0 & k_v & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} + \begin{pmatrix} u_0 \\ v_0 \\ w_0 \end{pmatrix}$$

Tels que : u_0, v_0 et w_0 sont les coordonnées de F (centre de projection) et k_u est le facteur d'échelle vertical et k_v est le facteur d'échelle horizontal.

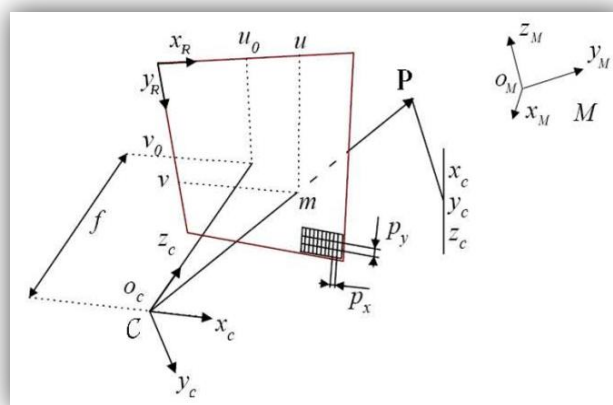


Figure [L.3] : Modèle de la caméra.

4. Formation de l'image

Une image obtenue avec une caméra est le résultat d'une transformation géométrique, cette transformation fait passer d'une représentation tridimensionnelle à une représentation bidimensionnelle.

La caméra est associée à un repère appelé centre projectif $O_c, (O_c, x)$ où (O_c, y) sont les axes parallèles au plan image et (O_c, z) est appelé l'axe optique et est orienté vers la scène, il est perpendiculaire au plan image.

La formation de l'image passe par trois phases [11] :

- 1) Un déplacement tridimensionnel : Pour passer du repère de la caméra, il faut effectuer un changement aux points tridimensionnelles exprimés dans un repère de la scène, ce changement comporte des paramètres intrinsèques¹ (6 paramètres : 3 pour la rotation et 3 pour la translation).
- 2) Projection 3D-2D : Après le changement de repère, les points tridimensionnels sont projetés sur le plan image, dans ce cas de nouvelles coordonnées sont obtenues appelées coordonnées normalisées.
- 3) Changement de coordonnées : A cette étape, il faut transformer les coordonnées normalisées pour passer aux coordonnées pixel, cette transformation est appelée transformation affine² du plan et comporte 5 paramètres appelés paramètres intrinsèques³ de la caméra. La formule de cette transformation peut être écrite par :

$$A = \begin{pmatrix} \alpha_u & -\alpha_u \cos \theta & u_0 \\ 0 & \alpha_v \sin \theta & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

Tel que α_u et α_v sont deux facteurs d'échelles le long des axes plan image. u_0, v_0 sont les coordonnées en pixel du centre d'image ; c'est à dire les coordonnées du point d'intersection de l'axe optique avec le plan image. Et θ représente l'angle entre les deux axes du repère image, et comme cet angle est très proche de la valeur $\frac{\pi}{2}$ la forme de la matrice de transformation devient :

$$A = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

¹ **Paramètres intrinsèques** : Ce sont les paramètres liés à la caméra (Plus de détails dans le chapitre III).

² **Transformation affine** : Utilisée pour placer un objet dans l'espace ou bien pour faire des changements entre les repères.

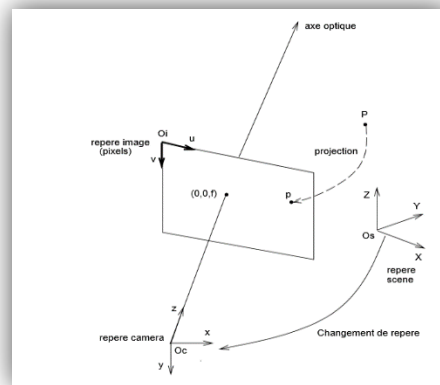


Figure [I.4] : Formation de l'image.

5. Relations géométriques entre les images.

L'ensemble des images prises par les caméras ne sont pas sans relation. La relation de transformation de la scène en une image est appelée homographie de plans, il est montré que deux images d'une scène plane sont liées par une homographie de plans [12].

Considérons deux images (Généralement une image gauche et une image droite) d'une même scène. Dans ce cas un point p d'une image gauche représente la projection de tous les points de l'espace appartenant à une droite OP (Voir la figure I.5), donc les correspondants de P dans l'image de droite sont tous trouvés sur cette droite. Cette géométrie est appelée géométrie épipolaire.

5.1 Géométrie épipolaire.

La géométrie épipolaire est un modèle mathématique qui vise à trouver une relation entre un ensemble d'images d'une même scène prises sous différents angles de vues, le principe général est de trouver pour chaque pixel d'une image son correspondant dans l'autre image, la relation obtenue par cette géométrie permet de réduire la zone de recherche des points correspondants qui seront utilisés pour la reconstruction 3D de la scène.

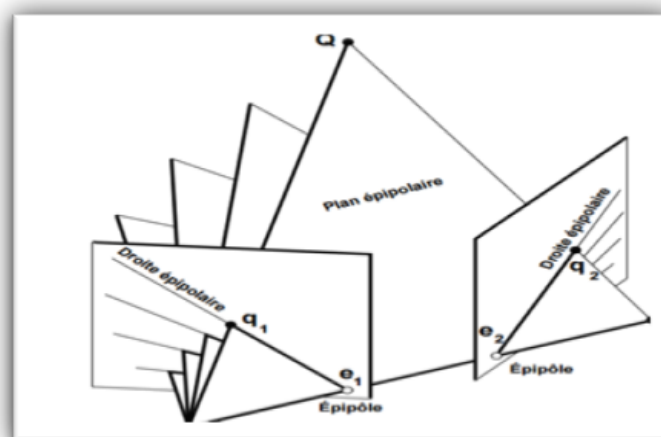


Figure [I.5] : Géométrie épipolaire

Considérons maintenant trois images d'une même scène, ces images sont traitées deux à deux, donc on dispose de trois paires d'images. On considère aussi trois matrices fondamentales¹ F_{11} , F_{13} et F_{23} (Voir figure I.6). Il est supposé qu'un point P de la scène qui se projette en p , p' et p'' dans les images 1, 2, 3 respectivement. Une géométrie épipolaire lie la première image avec la troisième image, on déduit donc que p'' se trouve sur la droite $(F_{13}p)$. De la même façon si on lie la deuxième image avec la troisième, dans ce cas p'' se trouve sur la droite $(F_{23}p')$. Ce dernier est donc trouvé par l'intersection de ces deux droites épipolaires :

$$p'' = (F_{13}p) \wedge (F_{23}p')$$

Il est déduit donc que les points p , p' et p'' sont reliés par une relation trilineaire qui dépend des matrices fondamentales. Cette relation est très utile pour la poursuite des points dans une séquence d'images, elle a un grand intérêt aussi dans la stéréovision trinoculaire.

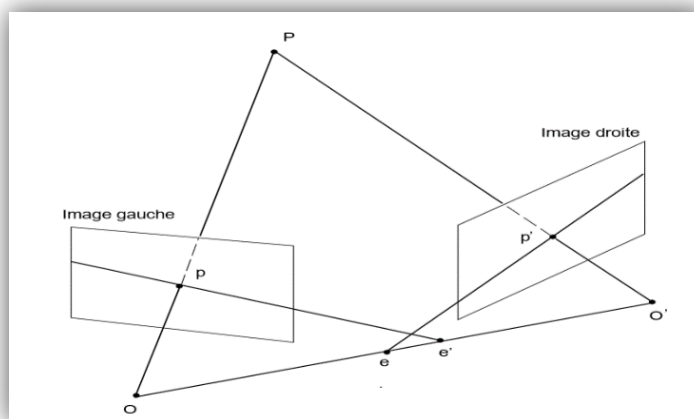


Figure [I.6] Relation géométrique entre deux images : Géométrie épipolaire.

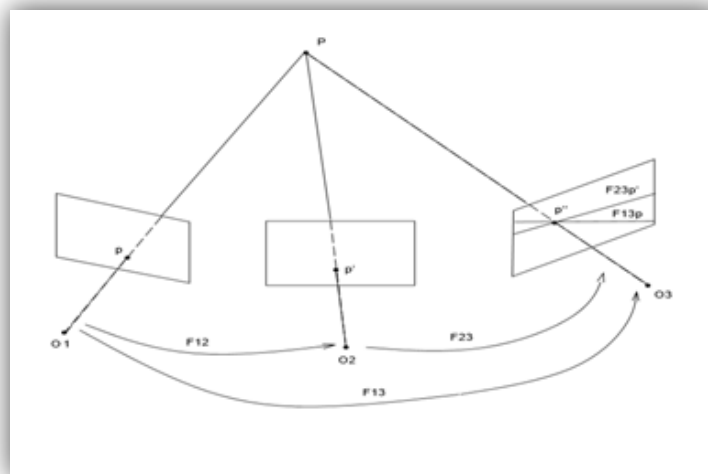


Figure [I.7] : Relation géométrique entre trois images : Trilinearité.

¹ **Matrices fondamentales** : C'est une matrice qui décrit la relation entre les coordonnées images et les coordonnées caméra.

6. Relation entre vidéo et images

La vidéo est une succession rapide d'images qui sont mises en œuvre à une certaine cadence. Pour obtenir une vidéo une caméra est utilisée pour prendre un ensemble d'images dans une unité de temps (seconde).

Le nombre d'images pris par seconde représente une unité de mesure qui correspond au nombre d'images affichées par un dispositif, plus ce nombre est élevé plus l'animation semble fluide. Cette mesure est notée par *FPS* (*Frames per second*) c'est-à-dire le nombre des images par seconds en français.

Pour le codage d'une vidéo, le terme *GOP* (Group of image) est utilisé, ce dernier signifie l'ordre dans lequel les images internes sont disposées. Les images qui appartiennent aux *GOP* permettent de générer l'ensemble d'images visibles dans la vidéo.

La reconstruction 3D à partir d'images peut être utilisée aussi dans le cas de vidéo. Le principe est très simple, c'est l'opération réciproque de prendre une vidéo : c'est-à-dire il suffit de prendre cette vidéo et d'y extraire l'ensemble d'images qui la compose. La méthode choisie pour la reconstruction sera appliquée sur la séquence d'images obtenues.

Conclusion

Dans ce premier chapitre nous avons présenté quelques notions de base en vue d'initiation à ce domaine, et d'enlever certaines ambiguïtés pour décrire quelques aspects utilisés dans la suite du mémoire. Parmi ces notions nous avons présenté la caméra et sa modélisation qui permet de décrire un objet photographié dans l'espace et projeté sur une image, donc nous avons détaillé le processus de formation de l'image.

Nous avons aussi présenté la relation entre les images prises par la caméra, cette relation est utilisée par la suite dans le traitement de ces images. Par exemple pour extraire quelques informations. Dans notre cas les informations qui seront utilisées pour le calcul de la troisième dimension par stéréoscopie.

L'objectif de notre travail est d'appliquer une reconstruction 3D sur la base d'une méthode qui utilise les images extraites d'une vidéo. Nous avons ainsi présenté la relation qui existe entre une vidéo et une image.

Après avoir présenté les généralités de la reconstruction 3D, nous allons présenter dans le chapitre suivant une classification générale des différentes méthodes de la reconstruction 3D. Avant d'entamer notre approche de reconstruction 3D à partir d'images qui fait le sujet de notre travail.

Chapitre II

Préliminaire

Introduction

La modélisation 3D classique consiste à créer des objets 3D en commençant par un ensemble de primitives de base (cubes, cylindres, courbes ...), elle nécessite des utilisateurs experts et un temps de traitement assez important (Pouvant être chiffré en Homme-Année¹). Cette modélisation n'est pas utile pour la réalisation de scènes complexes, malgré le grand progrès dans ce domaine.

Les résultats finaux sont, dans la plupart des cas, artificiels et les objets obtenus sont synthétiques. Les problèmes détectés dans la modélisation 3D classique ne résident pas seulement dans la complexité du travail, on trouve aussi que le matériel nécessaire pour la modélisation est très coûteux : Les logiciels de modélisation 3D sont assez cher et nécessitent des machines très puissantes.

Contrairement à la synthèse d'images classique, les méthodes IBMR (*Image Based Modeling and Rendering*) permettent de reconstruire la forme 3D à partir d'images d'une scène réelle. Ces approches ont attiré beaucoup d'attention dans la communauté graphique au cours de ces dernières années. Le principe général des méthodes IBMR est d'utiliser l'information présente dans les photographies pour la reconstruction 3D.

Dans ce chapitre, nous allons présenter en première partie plus de détails sur ces approches : Le fonctionnement, les méthodes les plus connues dans la littérature ...etc. Dans la deuxième partie nous allons présenter une classification des méthodes de reconstruction 3D qui regroupe toute les méthodes existantes.

1. Méthodes de modélisation et rendu à base d'image : IBMR

Il est connu que l'image est une représentation bidimensionnelle d'un monde tridimensionnel, la profondeur est perdue au cours de la projection de ces images. Le but des méthodes de modélisation et de rendu à partir d'images (*IBMR*) est de récupérer cette dimension perdue, en se basant sur les informations photométriques (couleur et intensité de chaque pixel) et géométriques (paramètres des caméras et disparité de chaque pixel) disponibles dans les images.

Certains auteurs font la séparation entre les méthodes de modélisation et les méthodes de rendu, dans ce cas les termes *IBR* et *IBM* sont utilisés respectivement pour (*Image Based Rendering* et *Image Based Modeling*). La technique *IBM* consiste à générer le modèle 3D à partir d'un ensemble de photographies d'une scène (Voir figure II.1), par contre le rendu basé sur l'image (*IBR*) utilise des images pour colorer ou texturer un modèle 3D (Voir figure II.2).

Les techniques d'IBMR se différencient selon la façon dont les informations disponibles dans les images sont utilisées. Certaines utilisent, interpolent², et déforment l'ensemble d'images d'une façon directe [13]. D'autres méthodes utilisent la géométrie de l'image pour générer des vues cohérentes de la scène, le résultat de cette méthode est une reconstruction partielle [14]. Il existe aussi des méthodes hybrides qui utilisent les vues de la scène réelle combinées avec une modélisation classique pour améliorer les résultats et accélérer le rendu [15].

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

² **Interpolation** : Opération mathématique permettant de construire une courbe à partir des données d'un nombre fini de points, ou une fonction à partir de la donnée d'un nombre fini de valeurs.



Figure [II.1] : Modélisation basée images. (Florian Moreno - Avril 2009 - Portal of Architectural Image-Based Modeling)



Figure [II.2] : Rendu basée image. (Fujitaand, Masahiro, Takashi, Kanai - hardware assisted relief texture mapping)

Les méthodes *IBMR* deviennent très utilisées en raison de leurs avantages par rapport aux autres techniques. Parmi ces avantages [16] :

- Les caméras et les webcams ne sont pas coûteuses. La possibilité de capturer des effets subtils des modèles 3D à partir de caméras ouvrirait la possibilité aux consommateurs de passer de la 2D vers la 3D.
- Les méthodes *IBMR* permettent d'éviter de nombreuses étapes difficiles suivies dans d'autres méthodes de modélisation 3D.
- Les performances des cartes graphiques ont augmenté de façon spectaculaire au cours de la dernière décennie. Néanmoins, le désir de réalisme et de grand détail limite souvent le taux de rendu de l'image au-dessus du temps réel. En appliquant des techniques *IBMR*, des images de certaines vues canoniques peuvent être pré-calculées, puis déformées en temps réel et interpolées dans la séquence finale souhaitée de la vitesse de la vidéo.

2. Quelques approches d'IBMR

De multiples méthodes de reconstruction 3D basées purement images sont existantes ; dans cette section nous nous intéressons aux approches les plus utilisées.

2.1 Forme à partir X (*Shape from X*)

2.1.1 Forme à partir d'ombres (*Shape from shading*)

L'approche *shape from shading* [17] est une technique de reconstruction 3D établie en premier temps par Horn au début des années 1970, ce dernier fut le premier qui a posé ce nouveau concept sous la forme d'une UDP voir [18]. Cette UDP appelée « L'équation d'irradiance », dont le principe est d'exploiter un ombre pour trouver le relief d'une scène à partir d'une seule photographie, puis d'estimer la forme d'un objet éclairé à partir des variations d'éclairage observé.

Pour cela, il est nécessaire d'avoir une image en niveau de gris et de connaître la direction exacte de la source lumineuse, dans ce cas il devient possible de déterminer la position 3D.

- Si la surface de la forme est décrite par l'information de la normale, cela revient à résoudre pour chaque pixel une équation linéaire à trois inconnus.
- Et si la surface est décrite en termes de gradient¹, cela revient à résoudre une équation linéaire à deux inconnus.

Cette approche donne des modèles 3D robustes, mais un problème majeur qui se pose est que les algorithmes *shape from shading* donnent des résultats dans des temps loin du temps réel. La figure suivante représente le résultat d'une telle approche [19] :



Figure [II.3] : Résultat de *shape from shading*. A gauche l'image source, au centre l'image en niveau de gris et le résultat dans l'image de droite.

¹ **Gradient** : Est un vecteur représentant la variation d'une fonction par rapport à la variation de ses différents paramètres.

2.1.2 Forme à partir silhouette (*Shape from silhouette*)

Cette méthode utilise les silhouettes de l'objet pour obtenir la forme 3D [20]. Un ensemble d'images prises par différents points de vue qui seront traitées pour obtenir le fond de la forme de l'objet.

Les silhouettes sont généralement représentées par un masque binaire, toute erreur de l'extraction n des silhouettes influe directement sur la qualité de la reconstruction 3D. Ces méthodes permettent d'exprimer la forme 3D, mais elles sont limitées uniquement aux objets entièrement visibles pour toutes les caméras.

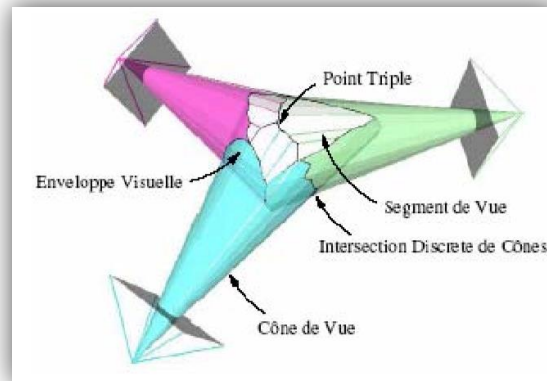


Figure [II.4] : *Shape from silhouette* à partir 3 vues différentes.

2.1.3 Forme à partir de textures (*Shape from texture*)

L'approche shape from texture [21] permet d'estimer la forme 3D de l'objet en fonction de variations observées sur la texture, c'est-à-dire le changement des matériaux et des couleurs.

Une texture est une caractéristique matérielle d'une surface constituée d'un motif plus ou moins régulier (grain d'un bois, poli d'un métal ...etc.). Les variations de texture dans l'image peuvent donner une bonne indication du relief¹ mais seulement dans le cas où cette texture est homogène.

L'avantage de ces méthodes et la simplicité des calculs et le texturage immédiat du modèle 3D. Mais l'inconvénient réside dans la sensibilité aux changements d'éclairage et l'augmentation de la complexité en temps de calcul pour obtenir une bonne définition.

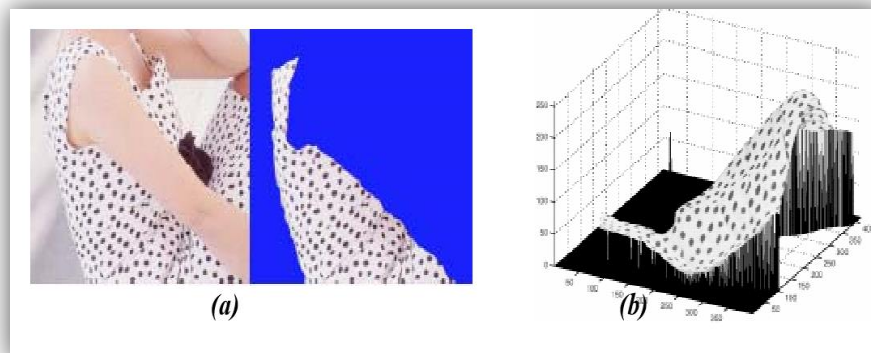


Figure [II.5] : (a) Texture extraite d'une robe, (b) surface reconstruite de la texture.

¹ **Relief** : Désigne une image (dessin, film, etc.) donnant l'illusion d'une troisième dimension, visible en stéréoscopie.

2.1.4 Forme à partir du mouvement (*Shape from motion*)

La reconstruction 3D à partir de mouvement [22] consiste à capturer une séquence d'images d'un objet en mouvement. La forme 3D est calculée en estimant pour chaque image le mouvement visible d'un ensemble de points ou bien de tous les points de l'image, la vitesse calculée permet de retrouver la forme 3D. Ceci est effectué en supposant que le mouvement est très faible entre chaque image, et la scène est statique afin que le déplacement des points dans l'image ne soit dû qu'au mouvement de la caméra. Cette approche est caractérisée par la facilité de mettre en correspondance les pixels entre les images par des méthodes de suivi visuel et ainsi d'exploiter les images. L'inconvénient réside dans la sensibilité au bruit par rapport à une méthode stéréo, dans ce cas le calcul de disparité est instable car le déplacement entre les deux images est très faible.

2.1.5 Forme à partir Focus/ defocus (*Shape from focus/defocus*)

C'est une technique qui consiste à reconstruire la forme 3D à partir de plusieurs images prises par une caméra fixe, avec différentes valeurs de mise au point. Cette approche utilise le fait que l'image de la scène dépend de la radiance de la région au lieu de la forme de cette région. On trouve donc qu'il existe plusieurs approches qui permettent de résoudre ce problème en se basant sur l'hypothèse que la surface de la scène est localement plane et parallèle au plan de la caméra (Hypothèse de l'équifocal) [23]. Ces méthodes restent sensibles au manque de texture des objets à reconstruire et sont moins robustes au bruit.

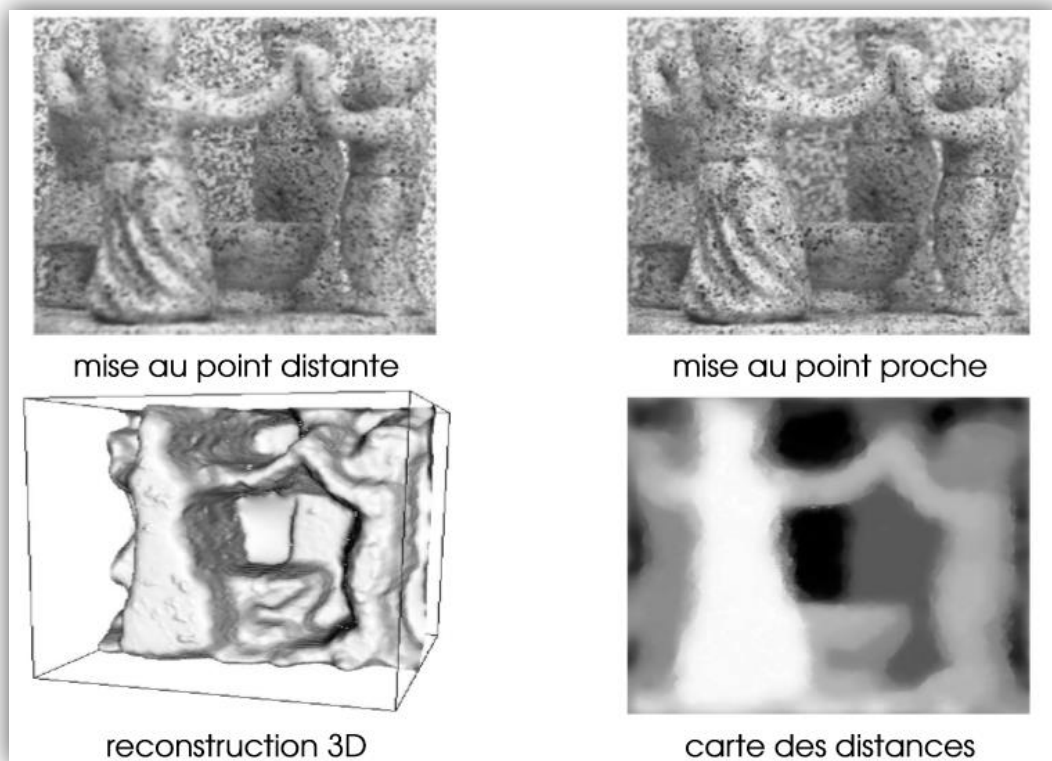


Figure [II.6] : Forme à partir de focus/defocus.

2.2 Les imposteurs

Les imposteurs [24] représentent une technique qui réalise un pré-rendu hors ligne d'une scène réelle, l'image créée est utilisée par la suite comme une texture plaquée sur un rectangle placé à l'intérieur de la scène (Précisément à la place de l'objet complexe de la scène). Le but de cette technique est d'obtenir une complexité visuelle au moindre coût.

Les imposteurs les plus connus sont les Panneau d'affichage (Dits *billboards* en anglais) et les *sprites* (Dits *Lutin*). Les *billboards* sont utilisés pour représenter des objets complexes et difficiles à modéliser, comme par exemples des arbres ou bien des nuages. Les *sprites* se différencient des *billboards*, un *sprite* est en mouvement et est appliqué directement à l'écran, il est généralement petit. On distingue deux types différents d'imposteurs : Les imposteurs statiques et les imposteurs dynamiques.

2.2.1 Imposteurs statiques

Les imposteurs statiques permettent une visualisation, ils visent à remplacer les objets complexes de la scène par des modèles simples est texturés dont le but est d'obtenir un taux d'affichage élevé et constant.

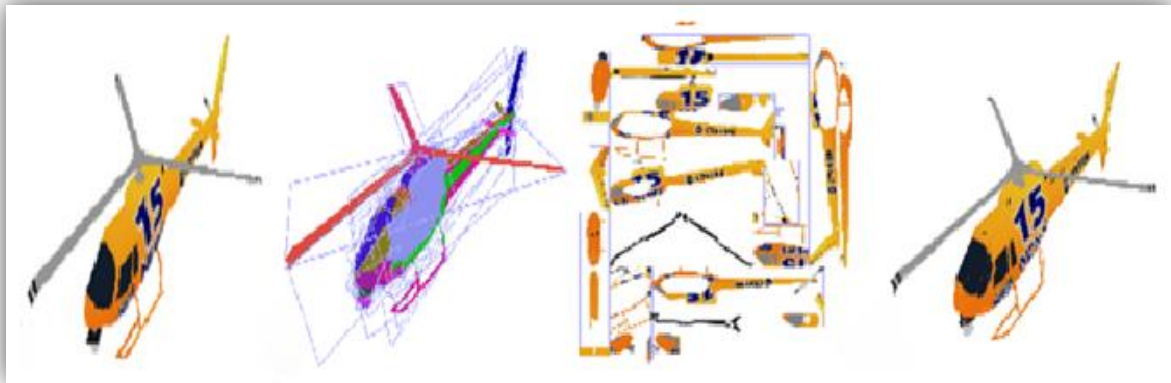


Figure [II.7] : De gauche en droite Modèle original, ensemble optimal de rectangles, plaquage de texture, Panneau d'affichage rendu.

2.2.2 Imposteurs dynamiques

Au contraire aux imposteurs statiques, les imposteurs dynamiques sont pré-générés. Dans cette méthode, l'imposteur est représenté par une image qui contient un polygone transparent où l'objet sera plaqué (Voir figure II.8).

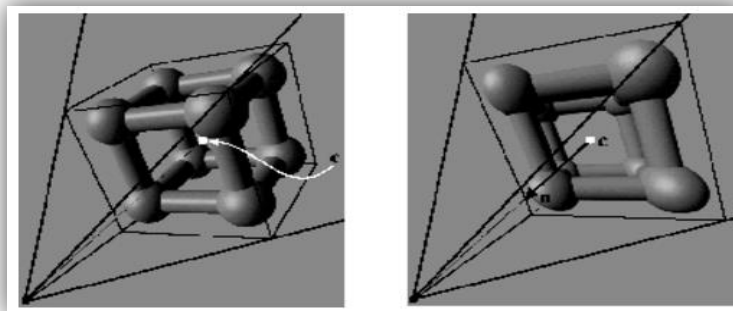


Figure [II.8] : Imposteur. A gauche, l'objet original, à droite, l'imposteur.

3.Classification multicritères des méthodes de reconstruction 3D

On trouve qu'il y a de multiples méthodes de reconstruction 3D, mais elles sont souvent limitées à un domaine spécifique d'application. Dans cette section nous allons aborder une classification multicritère [25] qui regroupe les différentes approches existantes. Cette classification permet d'aider les personnes au choix de la méthode adéquate avec la scène étudiée ou à la qualité envisagée. Pour éviter toute confusion possible, le tableau suivant résume les différentes méthodes présentées dans cette classification selon des critères bien définis :

Tableau [II.1] : Critères de classification des méthodes de reconstruction 3D.

Critère de classification	Méthodes existants
Niveau d'automatisme	Automatique
	Semi-automatique
	Manuelle
Niveau de complexité de scène	Grande scène
	Scène moyenne
	Petite scène
Mouvement de la scène	Scène statique
	Scène dynamique
Type d'approche utilisée	Approche ascendante
	Approche descendante
	Approche hybride
Passive/active	Méthode passive
	Méthode active
Multiplicité d'objet	Mono-objet
	Multi-objet
Degré de précision	Exigences de description
	Apparence visuelle
	Exactitude du modèle géométrique
	Plusieurs aspects à la fois
Degré de généralité de modélisation 3D.	Projective
	Euclidienne
	Raffine
	Métrique

3.1 Niveau d'automatisme

3.1.1 Reconstruction automatique

Le but de ces approches est de reconstruire une scène tridimensionnelle purement avec machine [26]. Le seul travail de l'utilisateur dans ce cas est de préparer la scène à reconstruire et les matériels à utiliser pour l'acquisition... etc.

Ce type d'approche est dédié pour la reconstruction d'objets simples. On trouve que la reconstruction purement automatique n'est pas encore efficace pour modéliser des scènes grandes et complexes, elle est aussi délicate pour le cas d'images éloignées. Par contre la reconstruction complètement automatique de scènes à partir des images proches est plus réussite, il devient possible d'avoir des formes 3D exacte.



Figure [II.9] : Exemple d'une voiture reconstruite automatiquement.

3.1.2 Reconstruction semi-automatique

Ces méthodes nécessitent des interventions spécifiques des utilisateurs. Le rôle de l'utilisateur est de réaliser l'ensemble des tâches difficiles à automatiser, il peut par exemple désigner les points ou bien les droites sur une ou plusieurs images (Voir figure II.10) pour approximer le modèle 3D à reconstruire, il peut même avoir des modèles génériques disponibles à augmenter jusqu'à obtenir une bonne estimation du modèle 3D. Le reste du travail sera effectué par la suite par le logiciel.



Figure [II.10] : Reconstruction semi-automatique : Exemple de détection des vertices et arrêtes pour obtenir le modèle final.

3.1.3 Reconstruction manuelle

Cette technique est presque pareil à la reconstruction 3D classique, la différence est que dans ce cas les objets à modéliser sont des objets réels au contraire à la reconstruction classique (objets synthétisés). Ce genre de reconstruction utilisé dans le cas d'indisponibilité des informations pertinentes à la reconstruction envisagée.

3.2 Niveau de complexité de la scène

3.2.1 Grandes scènes

La reconstruction 3D de scènes larges (comme par exemple la reconstruction des villes) nécessitent des techniques sophistiquées, elles ont des algorithmes spécifiques pour le plaquage de textures et pour la détection de formes, ces approches font intervenir les satellites, les données GPS, les avions ...etc.



Figure [II.11] : Exemple de reconstruction 3D d'une ville.

3.2.2 Scènes moyenne

La reconstruction de scènes d'ordres moyens (Des immeubles ou bien des bâtiments par exemple), utilise des techniques qui permettent l'acquisition des données nécessaires pour le processus de la reconstruction. Dans ce genre d'approches il est impossible par exemple de faire tourner l'objet considéré et de le photographier par une caméra fixe.

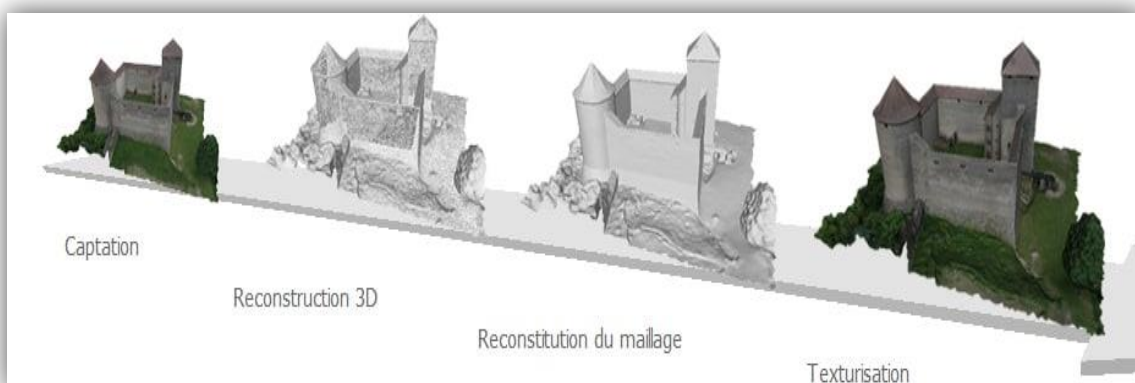


Figure [II.12] : Modèle 3D d'un bâtiment service publiques.

3.2.3 Petites scènes

Dans ce cas, l'acquisition de projections en 2D de l'objet à étudier n'est pas difficile et le temps requis pour collecter des données critiques pour la reconstruction est relativement petit.

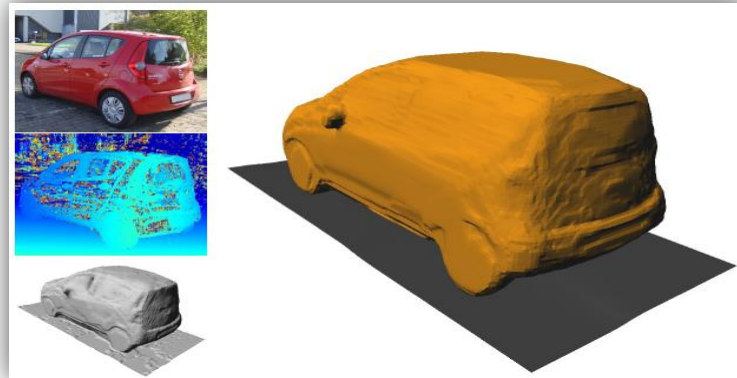


Figure [II.13] : Modèle de reconstruction 3D d'une petite scène.

3.3 Mouvement de la scène

3.3.1 Scènes statiques

La reconstruction de scènes immobiles 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 possible.

3.3.2 Scènes dynamiques

Pour reconstruire des objets mobiles il est nécessaire d'estimer les paramètres du modèle en mouvement, dans ce cas le problème porte sur la détermination de la zone de recherche. Lorsque le modèle en mouvement est inconnu, il doit être nécessaire d'estimer le vecteur de vitesse de chaque point. Les algorithmes de ce type de reconstruction manquent de robustesse.

3.4 Types d'approches utilisées

3.4.1 Approche ascendante

Elles sont connues par les approches (Bottom-up), dans ce cas la forme 3D est reconstruite à partir de primitives (Points 3D, courbes, segments...) sans prendre en compte d'éventuelles connaissances à priori sur les scènes que l'on cherche à reconstruire. Ces approches visent à reconstruire la surface par un maillage s'appuyant sur les primitives extraites, elles sont bien adaptées pour caractériser les structures au sein des objets et les limites saillantes de surface des objets.

Le manque d'informations qui concernent la scène génère parfois des résultats peu réalistes à cause d'une combinatoire importante, de plus l'extraction de primitives est sujette aux problèmes de sur-détection et de sous-détection qui influence le résultat final. Cette approche est utilisée par exemple pour la détection et la reconstruction des facettes 3D [27].

3.4.2 Approche descendante

Dans les approches descendantes (Top-down), une connaissance concernant les scènes est exigée, on commence par un ensemble de modèles bien connus pour choisir celui qui est le plus cohérent avec l'objet à reconstruire. Ces approches sont limitées par la bibliothèque des modèles précédemment définis et elles sont plus robustes que les approches ascendantes.

3.4.3 Approche hybride

Ces approches utilisent les approches (Bottom-up) et (Top-down) à la fois. Elles commencent par une formulation des hypothèses par une approche ascendante puis une vérification de ces hypothèses est effectuée par une approche descendante et à la fin les hypothèses de modèles géométriques et morphologiques élémentaires sont ajustées. Cela offre une très bonne connaissance sur la scène ce qui permet de la décrire par une grammaire de formes.

Le problème de ces approches est qu'il existe des cas où certains objets de la scène (surfaces et limites de surfaces courbe) ne peuvent être décrites par la grammaire de formes basiques, ce qui se reflète sur la forme 3D finale de cette scène.

3.5 Passive ou active

3.5.1 Les méthodes actives

Les méthodes actives [28] sont des méthodes qui utilisent des sources lumineuses avec des capteurs qui ont une action directe sur l'objet à étudié. Si la position de la source lumineuse et le modèle de la caméra sont connus, il devient possible de calculer les coordonnées tridimensionnelles des points de la scène par une simple triangulation. Ces méthodes sont utilisables dans différentes applications, particulièrement dans le domaine de la robotique.

Parmi les méthodes actives, on trouve la méthode de télémétrie Laser [29], ces méthodes utilisent en général le temps que met une impulsion à faire un aller-retour pour calculer la distance du point visé.

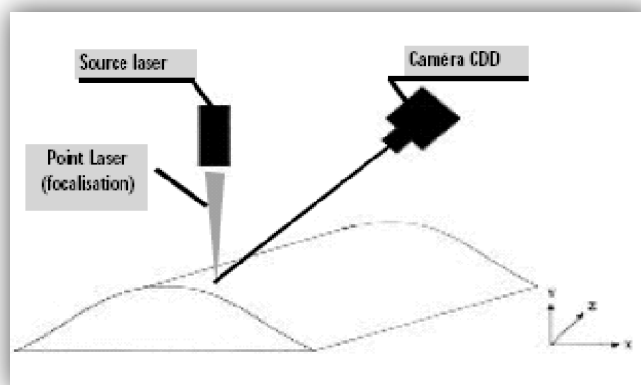


Figure [II.14] : Télémètre Laser.

Un autre type de méthodes actives est la reconstruction basée sur la lumière structurée [30]. Cette dernière se réalise à base d'un éclairage actif projeté sur la scène et qui permet d'obtenir un nuage de points 3D dense. Le principe de cette méthode sert à observer les déformations de cette scène à l'aide d'une caméra décalée légèrement à un projecteur (La source lumineuse), le résultat de l'éclairage permet de calculer les profondeurs par une triangulation.

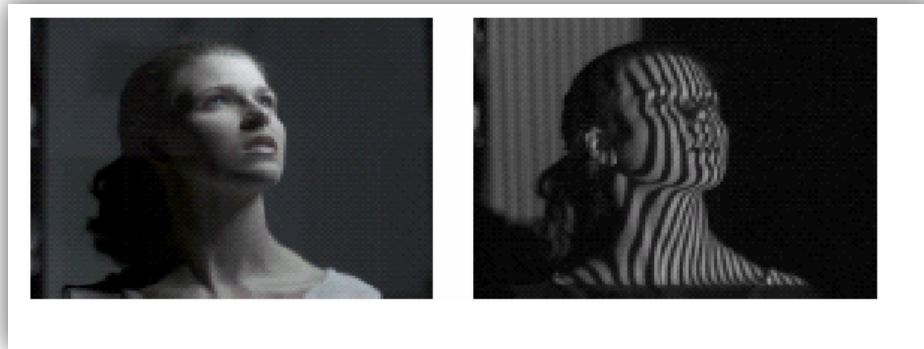


Figure [II.15] : Lumière structurée illuminant un visage.

3.5.2 Les méthodes passives

Le principe de la reconstruction 3D avec les méthodes passives [31] est d'imager une scène à reconstruire à partir de plusieurs points de vue différents, les capteurs utilisés dans ce cas sont généralement des caméras. Ces méthodes n'ont aucune action sur l'environnement, le terme « passive » vient du fait qu'on ne considère que la perception des radiations visibles dans la scène. La stéréovision est l'une des méthodes passives les plus connues et les plus utilisées, cette dernière se base sur l'utilisation d'au moins de deux caméras pour prendre les captures de la scène depuis des angles de vues différents. Cette méthode dépend de la calibration des caméras, de la mise en correspondance entre les différents points qui composent l'image qui seront utilisés par la suite dans une phase de triangulation qui permet de calculer les profondeurs de la scène. La phase la plus difficile dans la méthode de la stéréovision est la mise en correspondance, ce qui la rend inefficace pour reconstruire des scènes complexes.

Cette méthode fait partie intégrale de notre travail, et nous allons détailler son principe de fonctionnement dans le chapitre suivant.

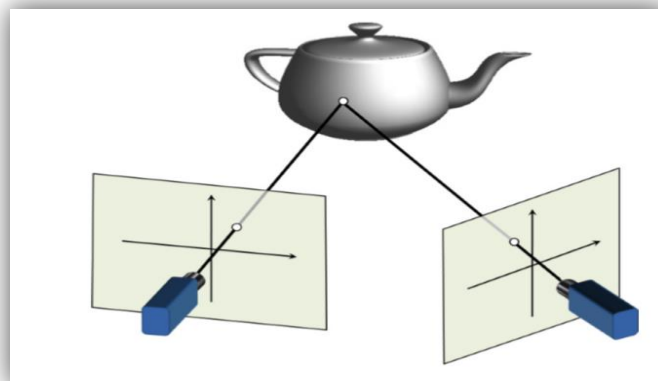


Figure [II.16] : Simple exemple du système de stéréovision.

3.6 Multiplicité des objets

3.6.1 Reconstruction mono-objet

Elle vise à traiter la scène comme un élément non décomposable, ceci est le cas d'un seul objet tout simplement ou bien de reconstruire la scène une fois pour toute si elle contient un ensemble d'objets indissociables. Cette méthode n'est pas fiable pour les objets complexes.

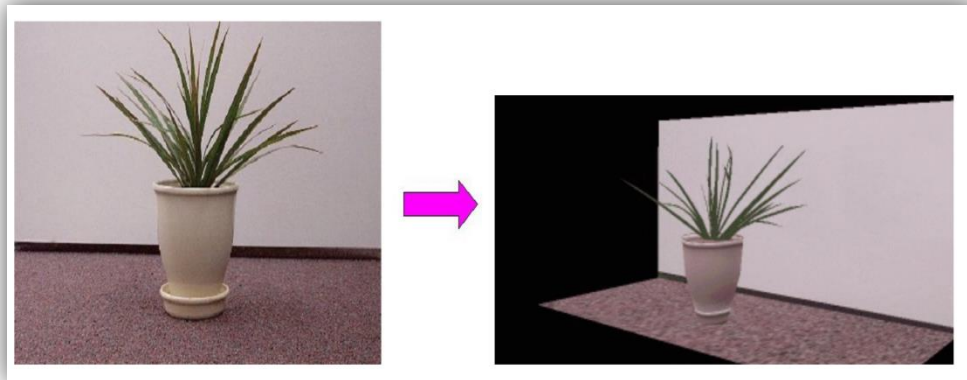


Figure [II.17] : Exemple de reconstruction mono-objet.

3.6.2 Reconstruction multi-objet

Pour aboutir à la reconstruction d'une scène complexe, cette dernière doit être décomposée en séparant les objets qui la compose et en procédant par la reconstruction de chaque objet indépendamment aux autres. Les modèles générés à la fin sont associés dans un seul modèle globale.

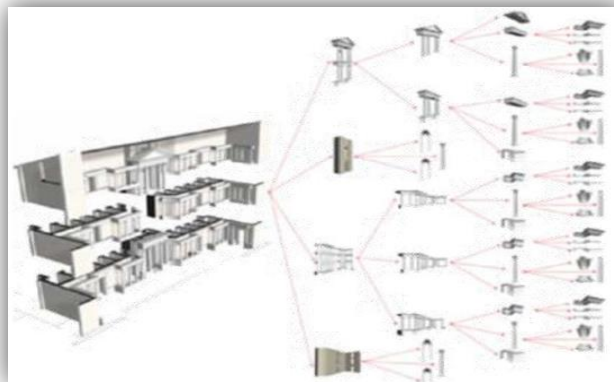


Figure [II.18] : Regroupement des éléments reconstruits.

3.7 Degré de précision

3.7.1 L'apparence visuelle

Ces méthodes se basent sur l'apparence visuelle des surfaces qui composent l'objet étudié. Cette apparence est estimée en étudiant les caractéristiques des matériaux et de la source lumineuse ainsi que la géométrie des objets.

3.7.2 Exigences de description

Cette approche consiste à regrouper un ensemble de points pertinents pour restituer les seules informations nécessaires à la description d'un aspect spécifique tels que la lecture de la forme, la récupération des caractères constructifs... etc.

Ce type d'approches est caractérisé par le traitement conçu en cohérence avec un objectif de représentation.

3.7.3 Exactitude du modèle géométrique

Ces méthodes s'appuient sur un balayage laser 3D pour réaliser la reconstruction 3D. Un problème lié à ces méthodes est que les surfaces composant l'objet entier sont difficilement calculées. La qualité des surfaces résultantes représente un autre défaut de ces approches, les modèles obtenus contiennent des artefacts de la géométrie (Des trous, une discontinuité... etc). De plus le type de matériaux impactés par le faisceau laser détermine l'intensité du signal de retour, ce qui cause une perte d'informations, donc le maillage généré ne représente pas la forme réelle de l'objet.

3.7.4 Plusieurs aspects à la fois

La représentation de l'objet étudié dans ce cas nécessite l'intervention de plusieurs méthodes qui sont utilisées pour la restitution des aspects reliés par cet objet. Dans l'étape d'acquisition, la capture d'informations métriques et photométriques est prise en compte d'après les différents niveaux de complexité géométrique que l'objet examiné exhibe et par rapport aux contraintes distinctives du système de représentation choisi.

3.8 Le degré de généralité de la modélisation 3D

3.8.1 Reconstruction projective

La reconstruction projective est réalisée à partir de plusieurs images d'une scène et des coordonnées des points correspondants dans les différentes images, ces coordonnées doivent être représentées dans un repère projectif. Cette reconstruction est réalisée sans aucune connaissance sur la structure de la scène ou sur l'orientation des caméras.

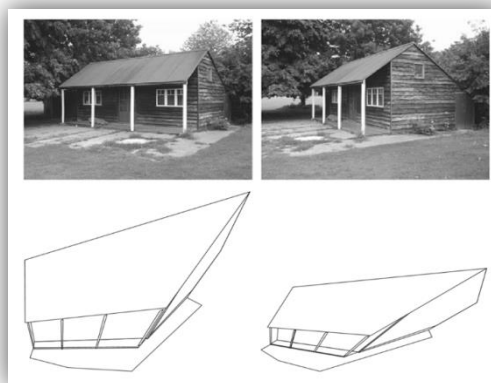


Figure [II.19] : Résultat d'une reconstruction projective. (Kassem Al Ismaeil - Structure from Motion & Camera Self-Calibration, 2011)

3.8.2 Reconstruction euclidienne

Contrairement à la reconstruction projective, la reconstruction euclidienne nécessite une connaissance à priori de la calibration des caméras. Les paramètres de la caméra qui sont estimés à l'aide de la géométrie de la scène, permettent de donner une correspondance entre les dimensions réelles et les dimensions de la caméra. Ce type de reconstruction respecte les angles et les distances et est aussi la plus complète.

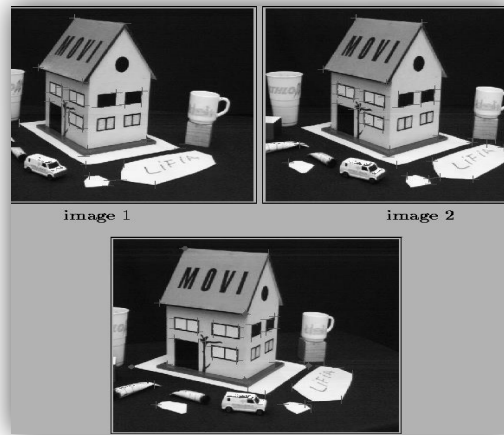


Figure [II.20] : Reconstruction euclidienne. (Bill Triggs, 1998)

3.8.3 Reconstruction affine

Cette méthode peut être appliquée quand les points sont définis dans un repère affine. La difficulté réside dans l'estimation de ce repère. Pour réaliser cette reconstruction il suffit alors de prendre une quelconque paire de points à l'infini dans les deux images et de reconstruire le point 3D pour obtenir le troisième point qui permet de déterminer le plan à l'infini.

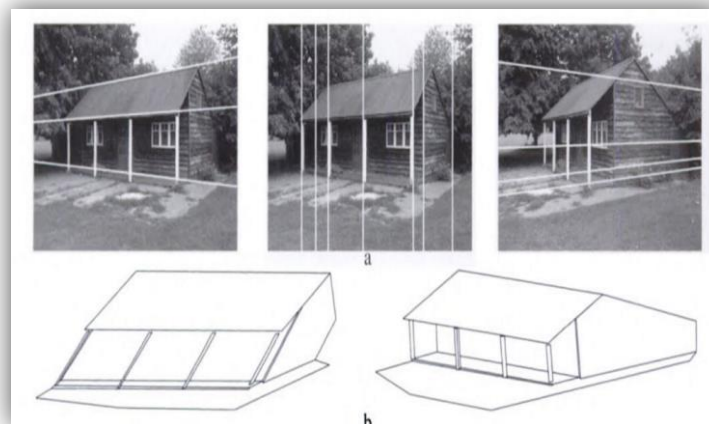


Figure [II.21] : Résultat d'une reconstruction affine. (Kassem Al Ismaeil - Structure from Motion & Camera Self-Calibration, 2011)

3.8.4 Reconstruction métrique

Cette approche est utilisée lorsque les coordonnées des points sont définies dans un repère métrique. La reconstruction est alors définie à une transformation métrique près. Une telle reconstruction contient des informations métriques (Changement d'échelle).

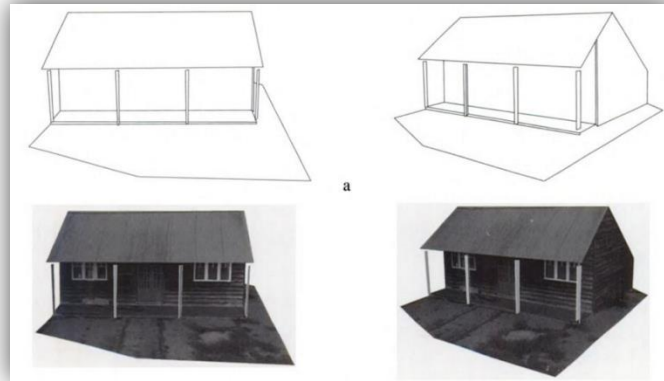


Figure [II.22] : Résultat d'une reconstruction métrique. (Kassem Al Ismaeil - Structure from Motion & Camera Self-Calibration, 2011)

Conclusion

Dans ce chapitre nous avons passé en revue les méthodes de reconstruction 3D à partir d'images qui sont connues par le terme *IBMR*. Ce type de reconstruction utilise les informations présente dans les photographies pour le calcul de la troisième dimension, nous avons présenté aussi comment ces approches sont assez utiles par rapport à une modélisation 3D classique.

Il existe plusieurs méthodes appartenant à la famille *IBMR* qui sont très connues dans la littérature. Parmi ces méthodes on trouve celles qui se basent sur une seule image comme le cas de la méthode shape from texture, ou bien sur plusieurs images comme le cas de shape from silhouettes. Certaines d'autres, utilisent d'autres principes totalement différents comme par exemple les imposteurs. De plus il existe autres approches qui appliquent des sources lumineuses sur les images, la projection de cette source donne des informations supplémentaires qui aident au calcul de la profondeur. Pour toutes ces méthodes nous avons présenté une brève description sur le fonctionnement de chacune d'entre-elles.

Nous avons aussi présenté au cours de ce chapitre une classification générale des différentes techniques de reconstruction 3D en se basant sur de multiples critères. Cette classification a pour but de simplifier le choix de la méthode à utiliser, ce qui permet d'éviter une mauvaise reconstruction 3D. Il n'est pas possible par exemple d'excepter de reconstruire une grande ville en temps réel, par contre cette méthode est utile pour le reconstruction d'un seul objet fixé.

Dans le prochain chapitre nous restons dans le cadre des méthodes *IBMR*, et nous allons détailler comment les deux méthodes enveloppe visuelle et stéréovision fonctionnent. Ces approches constituent la base de notre travail par la suite.

Chapitre III

Enveloppe visuelle et stéréovision

Introduction

Nous avons vu au chapitre précédent que les méthodes *IBMR* utilisent les informations disponibles dans les photographies pour calculer la troisième dimension perdue au cours du processus de projection. Cette reconstruction est effectuée à base des points visibles dans l'ensemble d'images disposées. Ce problème est assez délicat pour le résoudre par une machine, et d'obtenir un résultat proche de la réalité. L'enveloppe visuelle et la stéréovision sont classées parmi les méthodes *IBMR* les plus connues.

Dans cette partie nous allons présenter en premier lieu la méthode d'enveloppe visuelle, en passant par son fonctionnement, arrivant aux approches existantes pour la réaliser ainsi que les problèmes liés à cette approche.

Par la suite nous allons exposer la méthode de stéréovision en expliquant son but et ces différentes étapes.

Ces méthodes constituent le vif de notre travail et seront utilisées par la suite dans l'implémentation de notre méthode.

1. L'enveloppe visuelle

L'enveloppe visuelle est une méthode assez populaire et utile pour la reconstruction 3D à cause de sa simplicité et efficacité. Elle est introduite en premier temps par *Baumgart* dans sa thèse de doctorat en 1979 [2]. Le principe de base est de réaliser une représentation 3D d'un objet à partir d'un ensemble de silhouettes prises sous différents points de vues d'où vient le terme (Shape from silhouette).

L'enveloppe visuelle se définit généralement comme l'intersection de tous les cônes¹ de vue associés aux points de vue considérés, et constitue donc une région fermée de l'espace dont les points ne se projettent qu'à l'intérieur de tous les contours occultant. *Laurentini* [32] est considéré le premier qui a donné un sens à la notion de l'enveloppe visuelle, sa définition est comme suit :

« L'enveloppe visuelle $VH(S, R)$ d'un objet S par rapport à une région d'observation R est une région de E^3 telle que, pour chaque point $P \in VH(S, R)$, et chaque point de vue $V \in R$, la demi ligne partant de V et passant par P contient au moins un point de S ».

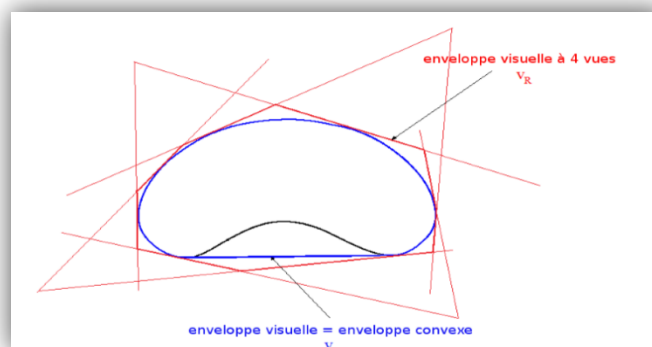


Figure [III.1] : Un objet 2D représenté avec une vue, sa silhouette et le cône correspondant.

¹ **Cône** : Un cône est défini de manière unique par son vertex et une section plane quelconque ne contenant pas le vertex.

Comme il est illustré dans la figure (III.1) on voit que pour chaque vue r l'objet est toujours contenu dans la cône V_r . Lorsque la taille de R tend vers l'infini et en même temps elle inclut toutes les vues possibles, on trouve que V_r converge vers la forme de l'enveloppe visuelle.

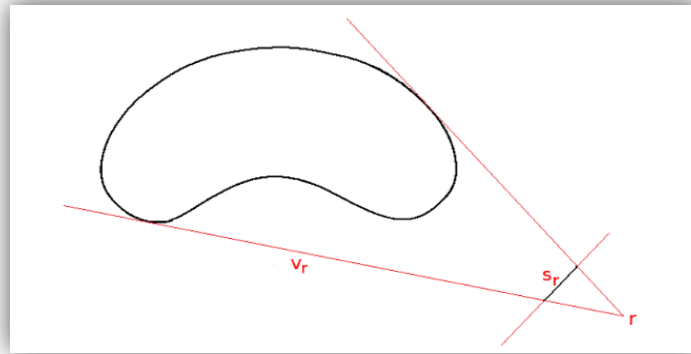


Figure [III.2] : v^∞ et V_r pour un objet 2D : v^∞ est l'enveloppe convexe de l'objet.

1.1 Problèmes liés à l'enveloppe visuelle

Il est impossible d'avoir une enveloppe visuelle complètement identique à la forme réelle de l'objet à reconstruire [33].

- La méthode de l'enveloppe visuelle n'est pas utile dans le cas où l'objet contient des zones concaves (Voir figure III.3). Ceci est dû au fait que l'enveloppe visuelle ne peut représenter que la forme de la surface 3D des pièces qui apparaissent sous forme de contours de silhouette 2D.
- L'enveloppe visuelle peut inclure certaines portions faussement positives appelées volumes fantômes comme illustré dans la (figure III.4). Les deux zones B et C peuvent coexister avec les zones d'objet réel A et D en termes de consistance de silhouette.
- Les faux négatifs¹ dans les silhouettes 2D de n'importe quel point de vue peut gâcher le résultat obtenu. Ce problème s'augmente si un grand nombre de caméras qui observent l'objet est utilisé.

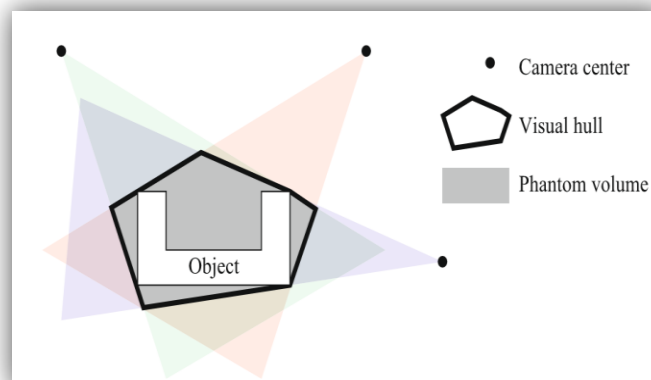


Figure [III.3] : Exemple d'objet avec une zone concave.

¹ **Faux négatifs** : ce sont des zones existantes dans l'objet réel mais ne sont pas disponibles dans la silhouette.

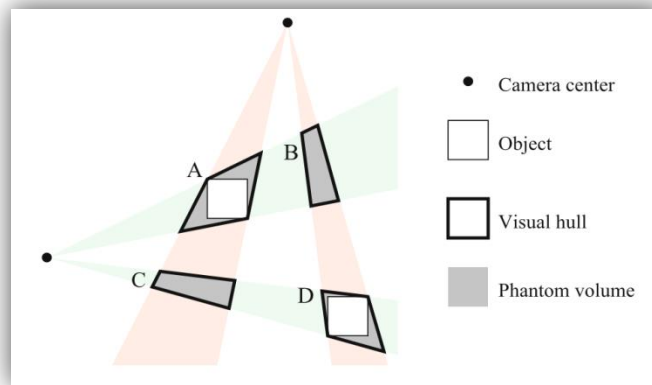


Figure [III.4] : Exemple d'objet avec fantôme.

1.2 Extraction des silhouettes

Comme il est mentionné ci-dessus, l'enveloppe visuelle peut être estimée à partir d'un ensemble de silhouettes prises par différents angles de vues. Dans chaque image il est nécessaire de déterminer l'ensemble des pixels correspondants à l'objet, ces pixels forment une silhouette. Les pixels qui n'appartiennent pas à la silhouette représentent la scène.

L'extraction de ces silhouettes doit être faite d'une façon exacte, pour ne pas avoir des conflits dans la forme 3D par la suite, elle représente un problème difficile qui continue à se développer et fait le sujet de plusieurs travaux de recherche. Il existe de nombreuses méthodes pour le calcul des silhouettes qui sont classées globalement comme suit :

- Méthodes basées différence de pixels.
- Méthodes basées régions.
- Méthodes basées contours.

Les différentes méthodes d'estimation des silhouettes fait partie de notre recherche, et nous citons brièvement dans la section suivante le principe de chaque méthode.

1.2.1 Méthodes basées différence de pixels

Dans ce cas l'objet est filmé sur un fond de couleur unique afin de détecter les pixels qui séparent l'objet de son arrière-plan. Les couleurs clés les plus répandues sont le vert et le bleu. La silhouette est estimée à partir des pixels qui donnent une différence par rapport à la couleur du fond. Il existe plusieurs méthodes [34] qui utilisent cette approche, car elle donne des résultats robustes mais elle est assez sensible au bruit issu par le capteur, ce qui nécessite l'utilisation des filtres pour l'amélioration des résultats.

1.2.2 Méthodes basées régions

Cette approche tente de traiter ces limitations à l'aide de remplissages des régions dans une image d'une façon heuristique. Ceci est fait par une propagation des critères locaux tels que la présence d'un faible gradient à l'intérieur de la région et d'un fort gradient aux bords de la région. Cette méthode est coûteuse en temps de calcul, surtout si un grand nombre d'images est utilisé.

1.2.3 Méthodes basées contours

Ce type d'approche consiste à traiter le problème d'extraction de silhouette comme un problème de minimisation d'une fonction d'énergie calculée sur un contour. Cette méthode prend en compte la présence d'un pic de gradient, il existe différentes méthodes de détection des contours. Le principal inconvénient lié aux méthodes basées contour réside dans la phase

d'initialisation du système qui est généralement faite d'une façon manuelle, la mal initialisation donne des silhouettes manquantes.

1.3 Méthodes de calcul d'enveloppe visuelle

Les approches de la reconstruction 3D à partir de l'enveloppe visuelle sont classées en deux grandes catégories : L'approche surfacique [35] et l'approche volumique [36]. La première catégorie est utilisée lorsqu'il s'agit de reconstruire la surface d'enveloppe visuelle en fournissant des points isolés ou un maillage. Mais on parle de la deuxième approche si on s'intéresse au volume délimité par l'enveloppe visuelle, et qu'on se base sur des discrétisations de l'espace.

1.3.1 Les approches surfaciques

Ces méthodes se basent sur le calcul d'une représentation de la surface explicite de l'enveloppe visuelle. Les points ou bien les facettes, sont estimés par intersection des cônes de silhouettes. Ces approches sont particulièrement utilisées dans les contextes de reconstruction 3D de la géométrie pour le rendu depuis n'importe quel point de vue. Il devient possible de calculer la forme 3D à partir d'enveloppe visuelle surfacique en temps réel. Ces méthodes sont assez sensibles à l'extraction des silhouettes et aux imprécisions de calibrages des caméras, mais elles sont très utiles dans le contexte de modélisation pour le rendu.

1.3.2 Les approches volumiques

Au contraire aux méthodes surfaciques, les approches volumiques estiment le volume de l'enveloppe visuelle par un ensemble de primitives élémentaires telles que : Les voxels¹, les arbres octaux². Les approches volumiques sont généralement plus rapides que les approches surfaciques, elles résistent aussi mieux que ces dernières aux silhouettes bruitées, ainsi que la reconstruction obtenue par les approches volumiques peut être raffinée d'une manière rapide.

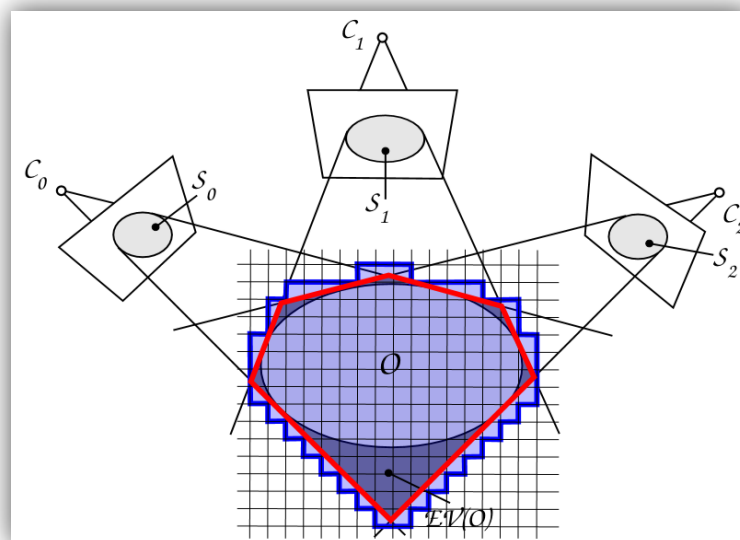


Figure [III.5] : Illustration d'une coupe de la géométrie estimée d'un objet O , en rouge enveloppe visuelle surfacique et bleue une enveloppe visuelle volumique.

¹ **Voxel** : C'est un pixel en 3D.

² **Arbre octal** : SDD de type arbre utilisée pour partitionner un espace 3D en le subdivisant sur 8 octants.

2. La stéréovision

La stéréovision vise à réaliser une tâche similaire à la vision humaine [3], cette méthode est apparue presque en même temps que la photographie, elle a connu un développement considérable au cours des dernières décennies.

Le but de la stéréovision est d'estimer la troisième dimension d'une scène à partir de deux images prises de deux positions différentes avec des caméras CCD¹ ou bien des caméras vidéo. Les caméras étant modélisées par une projection centrale, tous les points situés sur un rayon lumineux se projettent en même pixel p de l'image de gauche. Le centre de projection de la caméra de gauche dans le plan image de la caméra droite est appelé un épipôle. Cet épipôle avec un point de l'image forme une ligne épipolaire qui sera utilisée pour rechercher le correspondant p' de p , afin de pouvoir ensuite calculer la position de P dans l'espace par une triangulation en connaissant la position et l'orientation relatives des caméras.

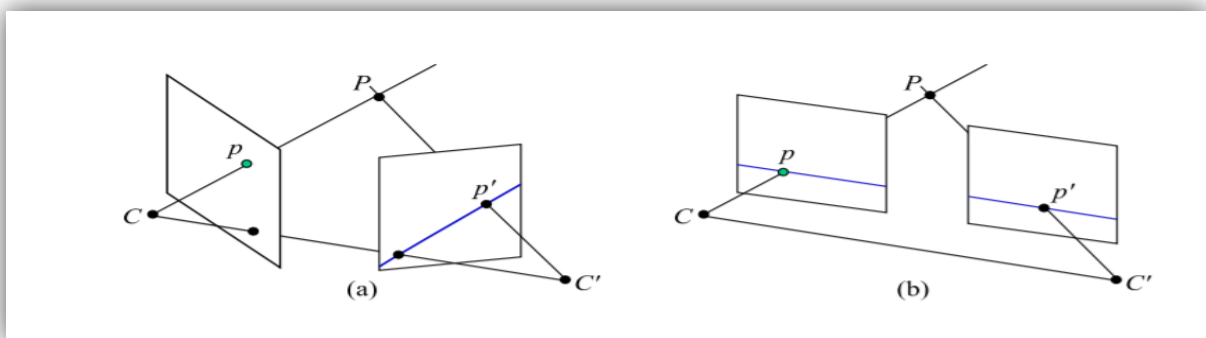


Figure [III.6] : (a) Géométrie épipolaire quelconque. (b) Géométrie épipolaire rectifiée.

On dit qu'une géométrie épipolaire est rectifiée lorsque tous les droites épipolaires sont parallèles et alignées.

La stéréovision s'effectue en trois étapes successives : Calibration, mise en correspondance et enfin la reconstruction 3D. Il existe un ensemble de problèmes liés à ce processus pendant leur automatisation [37] :

- Problème d'estimation du modèle de projection et d'estimation des relations spatiales entre deux caméras.
- Problème de mise en correspondance qui vise à trouver pour chaque pixel d'une image, le pixel correspondant dans l'autre image.
- Le problème de la triangulation ou bien la reconstruction 3D [4].

2.1 Calibration

La calibration de la caméra est une tâche importante qui fait l'objet de nombreux travaux [38], c'est une étape cruciale pour la réussite du processus de la reconstruction tridimensionnelle. Elle a pour but d'estimer les valeurs des paramètres du modèle géométrique de la prise de vues, ces paramètres sont de deux types : Paramètres intrinsèques et paramètre extrinsèque.

¹ CCD : (Charge Coupled Device) : soit « Appareil à Transfert de Charges ». Désigne un capteur qui convertit la lumière en charges électriques.

2.1.1 Paramètres intrinsèques

Ces paramètres sont internes à la caméra, on note comme exemples à ces paramètres :

- f : Distance focale.
- u_0 et v_0 : Les coordonnées de la projection du centre optique de la caméra sur le plan image.
- k_u et k_v : Les facteurs d'agrandissement de l'image.

2.1.2 Paramètres extrinsèques

Ce sont les paramètres géométriques d'orientation de position des caméras dans un repère attaché à la scène (le repère du monde). Parmi ces paramètres on a :

- t_x, t_y, t_z : Qui sont les composantes du vecteur de translation permettant de passer du repère lié à l'espace de travail au repère lié à la caméra.
- R_{3*3} : Qui est la matrice de rotation permettant de passer du repère lié à l'espace de travail au repère lié à la caméra.

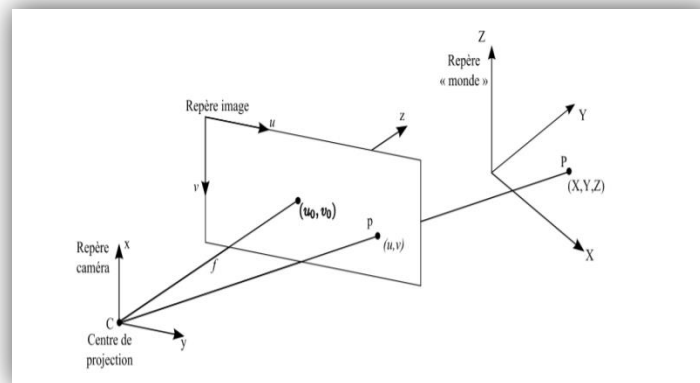


Figure [III.7] : Modèle géométrique de prise de vues.

Le passage entre les différents repères (repère monde et repère de la caméra) permet de retrouver les coordonnées de chaque point de la scène observée à partir de ces projections. Ce passage est caractérisé par deux matrices :

2.1.3 Matrice essentielle

Cette matrice permet de décrire les paramètres de passage du repère objet vers le repère caméra à base des paramètres extrinsèques qui sont la position et l'orientation des capteurs l'un par rapport à l'autre selon le repère de référence.

2.1.4 Matrice fondamentale

Cette matrice décrit les paramètres qui permettent le passage des coordonnées des points exprimés selon le repère caméra vers ceux exprimés selon le repère image, cela dépend des paramètres intrinsèques qui sont reliés par les caractéristiques de chaque caméra.

Avec la calibration, il devient possible d'estimer l'angle obtenu par l'intersection des deux rayons lumineux émis par le point correspondant, ceci est à partir de deux points homologues, il devient par conséquent possible d'estimer la position de ce point selon le repère objet (repère de référence). La calibration est classée en deux catégories globales.

2.1.4.1 Calibrage classique (hors ligne)

Cette calibration est faite par l'analyse d'images d'une mire dont la géométrie est connue précisément. Le principe de base est de photographier cette mire dont on connaît la forme, ces formes sont détectées par la suite sur les images afin de calculer les paramètres intrinsèques.

2.1.4.2 Auto-calibration (en ligne)

C'est une calibration automatique qui ne nécessite pas une connaissance des composants de l'image. Ce type de calibration est surtout utilisé dans les recherches en robotique. Dans ce cas on n'a pas besoin de photographier la mire, les paramètres intrinsèques sont calculés une fois pour toute, par contre les paramètres extrinsèques doivent être recalculés à chaque mouvement à l'aide de la géométrie épipolaire [15].

2.2 La mise en correspondance

La mise en correspondance ou bien connue aussi par l'appariement [39] est un problème largement étudié et toujours posé en vision stéréoscopique, elle représente l'étape la plus délicate dans la reconstruction 3D. De nombreuses techniques ont été proposées pour résoudre ce problème, seront détaillé ci-dessous.

La mise en correspondance de deux images stéréoscopiques consiste à établir une relation point à point entre celle-ci, elle s'agit de retrouver sur deux images les paires de pixels qui correspondent à la projection d'un même élément de la scène.

2.2.1 Primitives stéréoscopiques

Avant de passer à l'étape de la mise en correspondance, il faut d'abord connaître ce que l'on doit chercher à apparier, on s'intéresse aux types de primitives suivantes :

- Point : (pixels, points d'intérêt, éléments de contour, points caractéristiques le long d'un contour, jonctions etc...).
- Segment : (segments de droite, arcs de cercle, portions de conique etc...).
- Région : Les régions sont mal adaptées dans le processus de la stéréovision sauf quelques cas particuliers.

2.2.2 Notion de la disparité

La notion de la disparité est bien introduite dans le domaine de la vision humaine, elle représente la différence de position de la projection d'un même objet sur la rétine de l'œil. La disparité dans le domaine de la mise en correspondance est définie par :

La disparité notée par $disp$ est la différence de position entre la projection d'un même point de la scène observée sur les plans images des caméras gauche et droite.

$$d: \mathbb{N} \rightarrow \mathbb{R}^+$$

$$I_g(u, v), I_d(u, v') \mapsto disp(I_g(u, v)) = |v - v'|.$$

2.2.2.1 Vecteur de disparité

Le résultat, est stocké dans un vecteur de disparité noté par $d_{i,j}^{g \rightarrow d}$, ce vecteur décrit la position de deux pixels correspondants entre deux images gauche g et droite d . On a donc : $d_{i,j}^{g \rightarrow d} = d_{i,j}$

$$zd_{i,j} \in ([di_{min}; di_{max}] \times [dj_{min}; dj_{max}]) \cup \{OCC\}. \text{ Si } d_{i,j} \neq OCC \text{ alors : } \begin{bmatrix} i' \\ j' \end{bmatrix} = \begin{bmatrix} i \\ j \end{bmatrix} + d_{i,j}$$

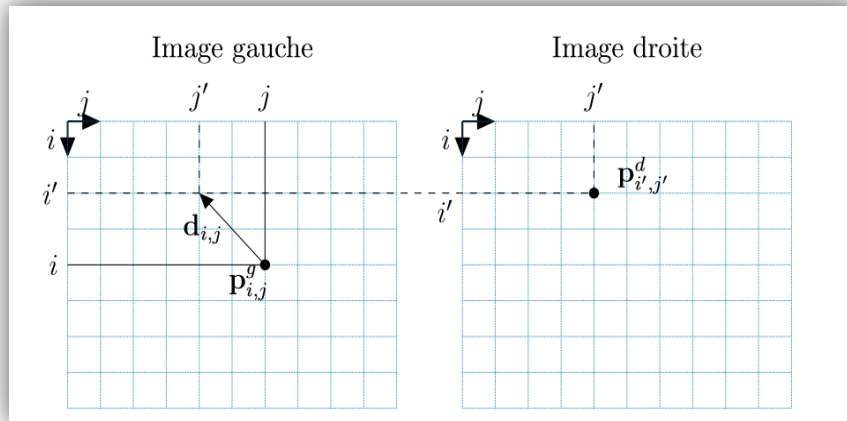


Figure [III.8] : Vecteur de disparité.

2.2.2.2 Carte de disparité

Le résultat final de la mise en correspondance est visualisé par une image dont le niveau de gris de chaque pixel représente sa disparité dans l'image de référence, cette image est appelée une carte de disparité, cette carte contient seulement les disparités des zones qui sont visibles dans les deux vues.



image de gauche de *Teddy*

image de droite de *Teddy*

carte de disparité
de référence

Figure [III.9] : Exemple de carte de disparité. (D. Scharstein and R. Szeliski - Middlebury Stereo Vision dataset).

2.2.3 Problème lié à la mise en correspondance

Les principales difficultés liées au processus de la mise en correspondance sont dues essentiellement à deux raisons [39] :

- **Information manquante** : Des fois on trouve que l'information nécessaire pour la mise en correspondance n'est pas disponible dans l'image, c'est le cas qu'on peut trouver dans :
 - **Les zones occultées** : Il s'agit des zones où un élément de la scène visible dans une image et n'apparaît pas dans l'autre image. (Figure III.10)
 - **Les zones de raccourcissement** : Différents points de la scène se projettent en plusieurs pixels d'une image mais sur un même pixel de l'autre image. Ce cas peut être détecté dans les surfaces très inclinées par rapport au plan image.
 - **Les zones de discontinuité de profondeur** : Cela peut se traduire par des voisinages de deux correspondants qui ne se ressemblent pas nécessairement (figure III.11).

- **Informations ambiguës** : Une information ambiguë est une information qui est présente dans une image mais il n'est pas possible de distinguer les correspondants.
 - **Régions homogènes** : Il est difficile de distinguer les points correspondants puisque tous les pixels de cette zone se ressemblent (figure III.12).
 - **Régions avec textures répétitives** : Il existe une ambiguïté lorsqu'un même motif se répète plusieurs fois (voir figure III.13).

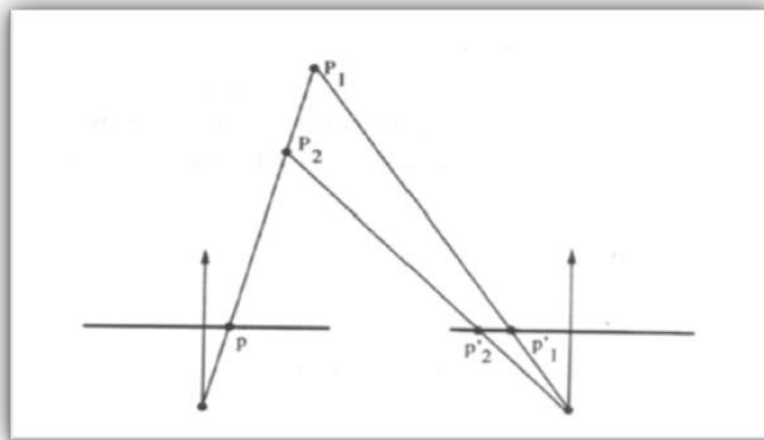


Figure [III.10] : Zone occultée – Le pixel marqué d'un point rouge a son correspondant occulté dans l'image de droite-

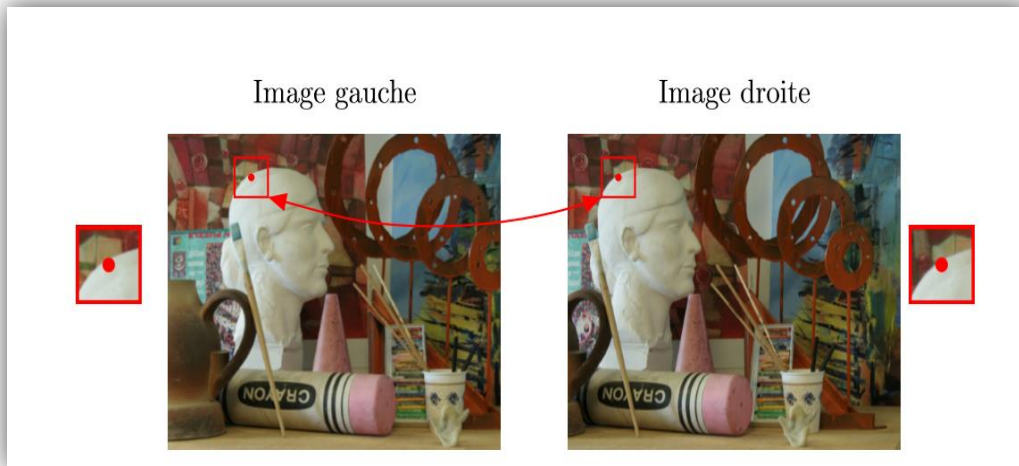


Figure [III.11] : Zone de discontinuité de profondeur – Les deux pixels d’un point rouge se correspondent mais leurs voisinages diffèrent car ils sont proches d’une discontinuité de profondeur-.

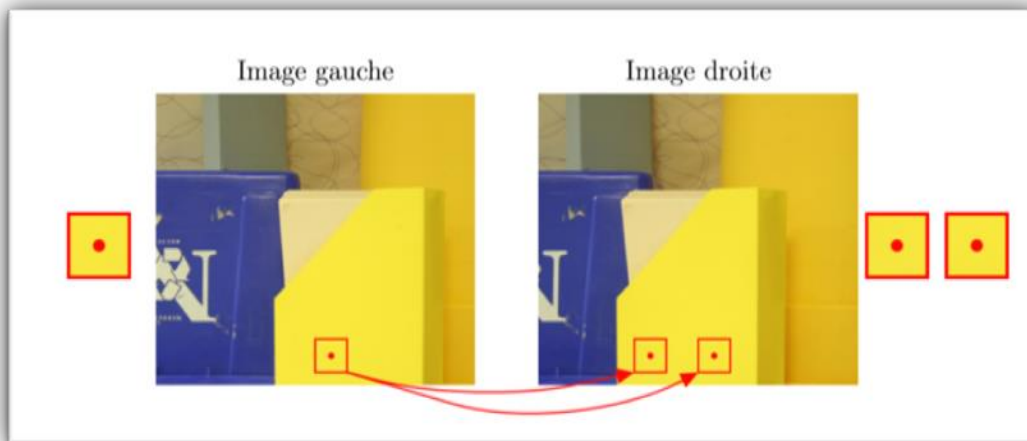


Figure [III.12] : Zone homogène –Manque d’information qui caractérise le pixel-.

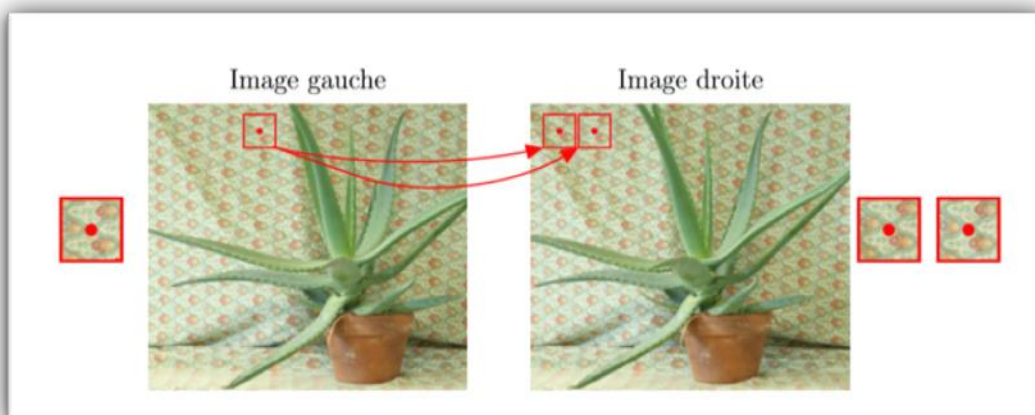


Figure [III.13] : Zone avec textures répétitives –Pour un pixel il existe plusieurs candidats dans l’autre image. (D. Scharstein and R. Szeliski - Middlebury Stereo Vision dataset)

2.2.4 Les contraintes de l'appariement

Un autre aspect assez important est la notion des contraintes de la mise en correspondance, on distingue deux types différents de ces contraintes : Contraintes géométriques et contraintes figurales. Ces contraintes sont utilisées dans le but de minimiser le champ de recherche de correspondants, ou bien pour éliminer certaines correspondances déjà établies.

2.2.4.1 Contraintes géométriques

Contrainte épipolaire

Il est possible de réduire la zone de recherche à une seule droite épipolaire en utilisant des propriétés de la géométrie épipolaire comme illustré dans la figure III.14.

P est un point de la scène, le point qui passe par P et les centres optiques O^g et O^d des deux caméras est appelé plan épipolaire. Ce plan coupe les images gauche et droite suivant deux droites appelées droites épipolaires. Pour renforcer cette contrainte, il faut que les deux images stéréo soient rectifiées. La rectification consiste à recalculer la position générale de deux images pour que la géométrie épipolaire de ces dernières soit simple, c'est-à-dire que les droites épipolaires soient horizontales.

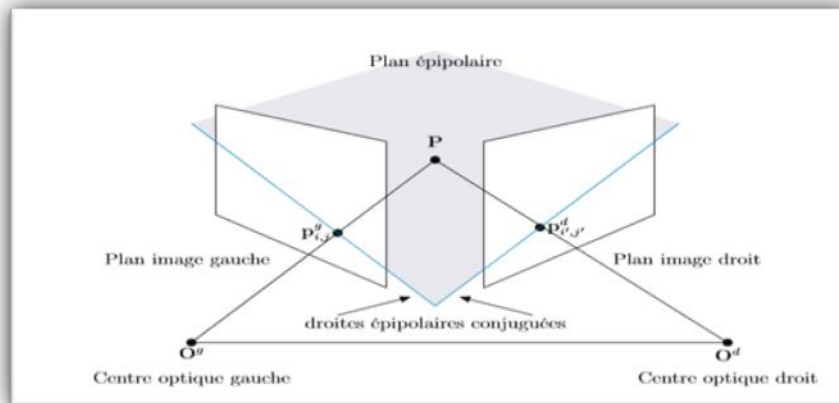


Figure [III.14] : Géométrie épipolaire.

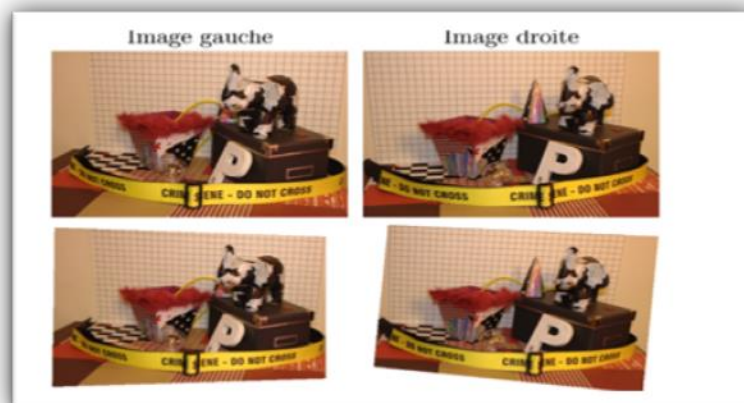


Figure [III.15] : Exemple de la rectification épipolaire. (G.Galles-Pixel matching for binocular stereovision by propagation of feature points matches and region-based randomized voting scheme)

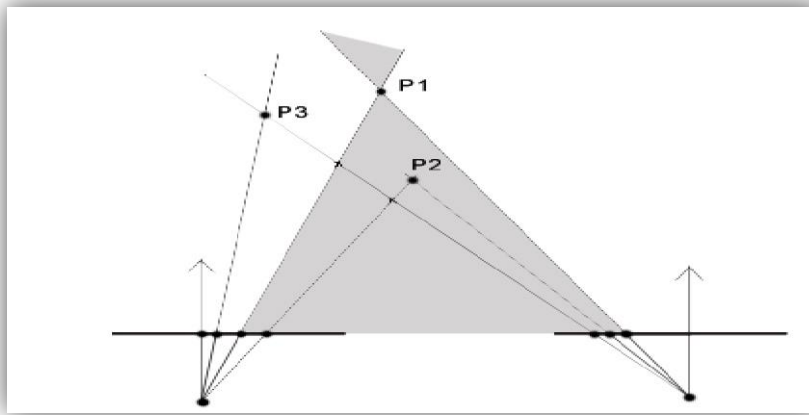


Figure [III.16] : Un point image correspond à un point objet est un seul.

Contrainte d'unicité

Le principe de cette contrainte est qu'un pixel d'une image gauche ne peut avoir qu'un seul pixel correspondant sur l'image de droite. S'il existe pour un pixel de l'image de droite, plus d'un pixel homologue dans l'image gauche, cela peut être dû à la présence des régions occultées, c'est le cas d'un pixel de l'image gauche qui n'a pas d'homologue dans l'image de droite.

Cette contrainte peut être formulée par :

$$\forall I_g(u, j) \in P_d, \exists ! I_g(u, v) \in P_g / (disp(I_g(u, v)) = |v - j|)$$

Si cette contrainte n'est pas respectée, la formulation devient :

$$\begin{aligned} \forall I_g(u, j) \in P_d, \exists ! I_g(u, v), I_g(u, v') \in P_g / (disp(I_g(u, v)) = |v - j|; \\ (disp(I_g(u, v')) = |v' - j| \text{ et } v \neq v') \end{aligned}$$

Contrainte de continuité de disparité

Cette contrainte est issue de la calibration de la caméra, elle implique que les primitives voisines spatialement ont des valeurs de disparité proches.

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. Maintenant le principe de la contrainte d'ordre est le suivant :

L'ordre des pixels le long de la droite épipolaire doit être le même que celui de leurs correspondants le long de la droite épipolaire de l'autre image. Voir la figure III.17

Cette contrainte peut être formulée par :

$$\begin{aligned} \forall I_g(u, v) I_g(u, v') \in P_g \text{ et } v > v', \exists I_d(u, j) I_d(u, j') \in P_d / (disp(I_g(u, v)) \\ = |v - j|, (disp(I_g(u, v')) = |v' - j'|, \text{ et } j > j'). \end{aligned}$$

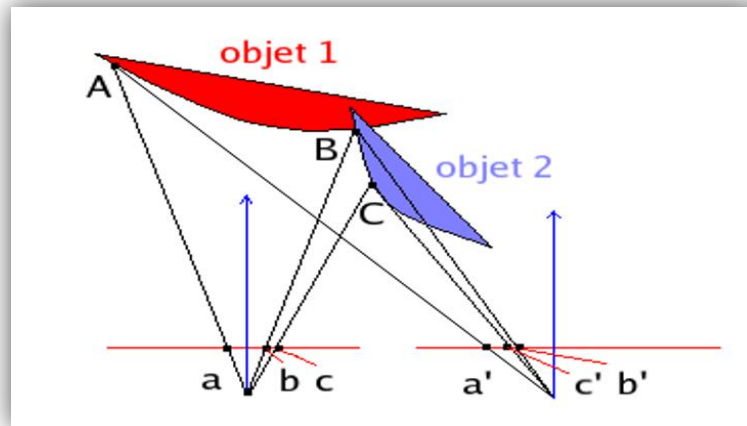


Figure [III.17] : Contrainte d'ordre.

2.2.4.2 Contraintes figurales

Les contraintes figurales peuvent imposer une contrainte d'ordre géométrique, dans cette catégorie on trouve toutes les contraintes spécifiques à une primitive particulière.

Disparité locale constante

Les contraintes d'ordre et d'unicité ne fonctionnent pas dans les cas où la scène contient des objets avec des surfaces fortement inclinées. Ce problème peut être évité si on impose une limite au gradient de disparité entre deux paires de points appariés consécutifs. Ce faisant on limite, en fait, l'inclinaison des surfaces des objets présents dans la scène, et de fait, on limite par là même la variété des objets à reconstruire.

Continuité figurale

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

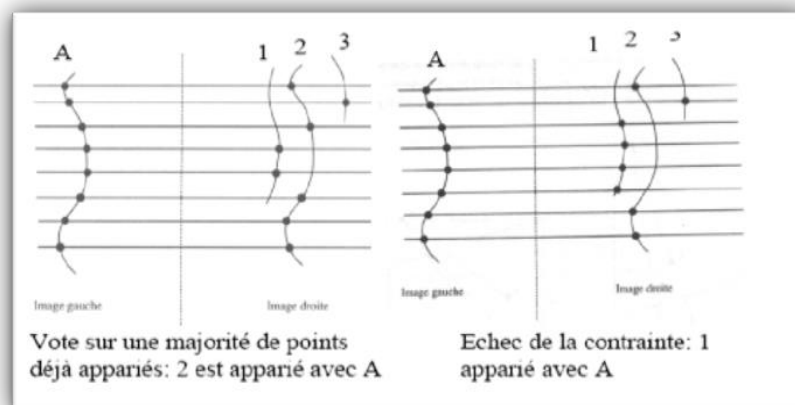


Figure [III.18] : Contrainte de continuité figurale.

2.2.5 Techniques de mise en correspondance

Il existe une diversité de méthodes [40] de mise en correspondance, les problèmes liés à cette dernière, vus ci-dessus, provoquent une instabilité de ce processus. La classification des méthodes d'appariement dépend de la complexité algorithmique, de la nature de la carte de disparité obtenue telles que les méthodes denses et éparses. Une autre classification qui est largement étudiée et qui classe l'appariement en méthodes locales et méthodes globales. Ces méthodes seront détaillées dans la section suivante.

2.2.5.1 Méthodes d'appariement denses

Ce type de méthode, prend en considération l'ensemble des pixels composant l'image. La mise en correspondance exploite l'information contenue dans le voisinage d'un pixel. La vraisemblance de la mise en correspondance de deux pixels est mesurée avec la corrélation¹ existante entre les voisins de chacun de ces deux pixels. Cette vraisemblance peut être estimée par l'écart des valeurs photométrique ou colorimétriques des pixels voisinages correspondants. Cette technique suppose la continuité des surfaces dans les scènes réelles, traduite par une continuité de disparité.

2.2.5.2 Méthodes d'appariement éparses

Le changement d'échelle dans les images causé par les transformations géométriques provoque une invariance dans les points d'intérêts². L'inconvénient majeur de ce type de méthode réside dans l'impossibilité de reconstruire une scène 3D complète. L'appariement de certains points d'intérêts considérés comme non ambigus, est adopté dans plusieurs algorithmes de mise en correspondance. Ces points correspondent aux primitives qui peuvent avoir un certain nombre d'attributs. Plus les primitives sont discriminantes, plus l'appariement est précis, robuste et fiable.

2.2.5.3 Méthodes d'appariement locales

Les méthodes locales mesurent la similarité des voisinages de deux pixels, Ce voisinage est appelé zone d'agrégation ou bien fenêtre de corrélation, pour plus de simplicité il est supposé que cette fenêtre carré de corrélation est obtenue par la formule :

$$(n * n) = (2r + 1) * (2r + 1), \text{ où } r \in \mathbb{N}^*$$

Cette fenêtre est centrée sur le pixel considéré. Il est noté aussi que \mathbf{a} est le vecteur de dimension n^2 qui contient les valeurs des pixels voisins du pixel considéré sur l'image de référence, et \mathbf{b} celui de l'autre image. Les méthodes locales supposent que les voisinages de deux correspondants se ressemblent, c'est-à-dire :

$$\mathbf{a} \approx \mathbf{b}$$

Dans la figure III.19 le pixel considéré pour la recherche de correspondant est marqué en noir, son pixel homologue est recherché dans la zone de corrélation. Dans cet exemple, la zone de recherche considérée est monodimensionnelle. Le but est de calculer pour chaque pixel candidat un score de corrélation est calculé. A la fin le candidat qui donne le meilleur score est retenu comme correspondant.

¹Corrélation : Utilisée pour calculer la similarité entre deux images.

² Points d'intérêts : Ensemble de points spécifiques tels que : Des coins, des courbes, des points de contours... qui sont utilisés comme des échantillons dans la phase de la mise en correspondance.

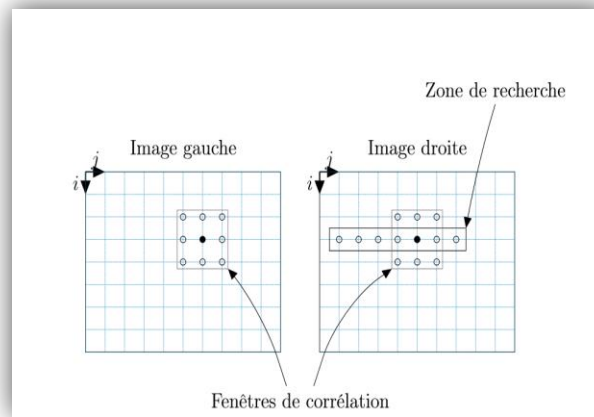


Figure [III.19] : Mise en correspondance locale.

Parmi les méthodes locales les plus connues on trouve les suivantes :

NCC: La méthode de corrélation croisée normalisée est une méthode couramment utilisée, elle est employée pour la mesure de similarité. C'est la somme des produits entre les termes correspondants des fenêtres de corrélation normalisée par le produit des moyennes quadratiques calculées pour chacune des fenêtres. L'avantage de cette méthode est la capacité de corriger les variations de l'intensité et de la couleur qui change d'une caméra à une autre. Par contre le problème de la corrélation normalisée est la lenteur des calculs.

Pour un pixel (u, v) est une disparité d on a :

$$\text{NCC}(u, v, d) = \frac{\sum_{i,j} (I_1(u+i, v+j) - \bar{I}_1) \cdot (I_2(u+d+i, v+j) - \bar{I}_2)}{\sqrt{\sum_{i,j} (I_1(u+i, v+j) - \bar{I}_1)^2 \cdot (I_2(u+d+i, v+j) - \bar{I}_2)^2}}$$

Tels que : I_1 et I_2 sont les deux images de la mise en correspondance, \bar{I}_1 et \bar{I}_2 correspondent aux moyennes de l'intensité des pixels sur la fenêtre de corrélation.

SSD : C'est la somme des différences carrées, utilisée pour la mise en correspondance en bloc.

$$\text{SSD}(u, v, d) = \sum_{i,j} (I_1(u+i, v+j) - I_2(u+d+i, v+j))^2$$

SAD: C'est la somme des valeurs absolues entre termes correspondants des deux fenêtres.

$$\text{SAD}(u, v, d) = \sum_{i,j} |I_1(u+i, v+j) - I_2(u+d+i, v+j)|$$

ZSAD: C'est une méthode analogue à la méthode SAD, mais on fait la différence entre écarts à la moyenne, elle représente un critère invariant aux variations uniformes de luminance dans une des images.

2.2.5.3 Méthodes d'appariement globales

Ces méthodes prennent en compte la totalité de l'image et cherchent à minimiser une fonction de coût qui exprime les erreurs de mise en correspondance. Le principe est d'estimer une fonction de disparité d permettant d'optimiser une énergie globale E , les deux étapes majeures de cette technique sont la définition d'une fonction d'énergie et l'application d'une technique d'optimisation. Le grand inconvénient de ce type de méthodes est la grande complexité des calculs.

Les recherches effectuées sur ces méthodes se basent sur la minimisation d'une fonction de coût globale. Cette technique regroupe les méthodes suivantes :

Programmation dynamique

Cette méthode consiste en la recherche d'une solution optimale par la minimisation de la complexité, en décomposant le problème général en plusieurs sous problèmes de complexité similaires. La programmation dynamique est équivalente à la recherche d'un chemin optimal dans un graphe. La première étape effectuée dans ce cas est la même dans les méthodes locales, la différence se situe au niveau de contraintes : Dans ce cas une contrainte de continuité de disparité est imposée, avec cette contrainte, on peut favoriser (défavoriser) la présence de discontinuités des profondeurs dans les appariements retenus. De cette façon l'approche assure une valeur de disparité en tous point de l'image de référence.

Relaxation

L'implémentation de cette méthode est très simple et facile, elle se base sur un algorithme itératif et localement parallèle et qui améliore une solution en minimisant localement l'énergie [41].

Flot de graphe (Graph-cut)

Le but de cette approche est de trouver une coupure de graphe qui possède un coût minimum parmi toutes les coupures possibles en se basant sur la minimisation d'une fonction d'énergie. Cette approche est abordée par Roy et Cox [42]. Les terminaux du graphe représentent les disparités pouvant être associées à un pixel de l'image, et les différents sommets correspondent aux pixels de l'image. Cette méthode permet d'obtenir une carte de disparité avec des zones homogènes en gardant les contours et son inconvénient fondamental est l'aplatissement des objets.

2.3 Reconstruction 3D

La reconstruction tridimensionnelle à partir d'images stéréo est basée sur les correspondants entre ces deux images. En lançant pour chaque caméra un rayon passant par le centre de projection et par le plan image à la position pixel donnée. Il suffit de trouver le point d'intersection de ces deux rayons qui correspond à la position 3D du point observé, cette position signifie que la profondeur est calculée à base d'une simple triangulation. La profondeur donc n'est qu'une fonction de disparité : Pour chaque point de l'image de référence si on peut trouver son point homologue dans la deuxième image, il devient possible de retrouver sa profondeur, si la correspondance entre tous les pixels des deux images est connue on peut avoir une reconstruction 3D complète de la scène photographiée.

On peut distinguer différents cas au cours du processus de reconstruction tridimensionnelle [37] :

- Les caméras sont calibrées donc les paramètres internes sont connus ainsi que la transformation rigide entre les deux caméras : Il s'agit donc d'une reconstruction euclidienne dans le repère de calibrage.
- Les paramètres internes sont connus mais la transformation rigide n'est pas connue : Dans ce cas il est nécessaire d'estimer la matrice essentielle qui permet d'extraire la transformation rigide des deux caméras à un facteur d'échelle pré. C'est une reconstruction euclidienne dans le repère de l'une des deux caméras.

Si les capteurs ne sont pas calibrés, c'est-à-dire ni les paramètres internes ni la transformation rigide sont connus. On a besoin donc, d'estimer la matrice fondamentale, à partir de cette dernière on peut obtenir une reconstruction projective tridimensionnelle.

2.3.1 La triangulation

La triangulation permet de calculer la position 3D (X, Y, Z) d'un point P à partir d'une projection en perspective de P sur les plans d'image des caméras une fois que la position relative et l'orientation des deux caméras sont connues.

Pour obtenir une précision des mesures exactes, la position entre les capteurs doit être importante et les capteurs doivent être parallèles et orientés vers l'objet à analyser. Soit \vec{P} un vecteur représentatif d'un point source dans le repère (O, X, Y, Z) . L'orientation du repère caméra est définie par la direction de la normale \vec{N} au plan du capteur. Et soient (\vec{u}, \vec{v}) les vecteurs des directions principales du capteur.

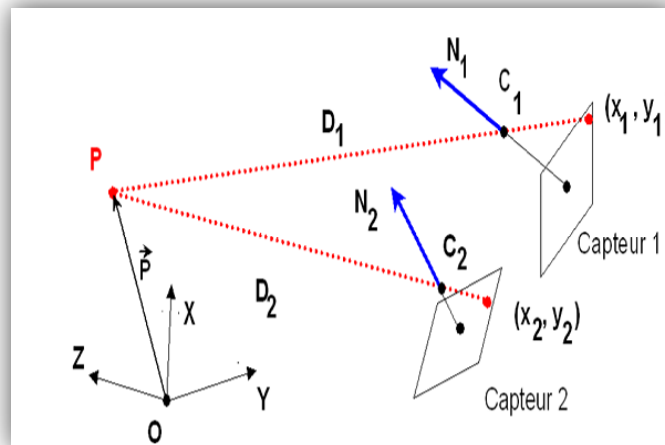


Figure [III.20] : Système de projection des scène –Représentation de triangle formé par le système.-.

La projection de P dans le capteur1 est donnée par :

$$x_1 = f \frac{\vec{D}_1 \cdot \vec{u}_1}{\vec{D}_1 \cdot \vec{N}_1} \quad y_1 = f \frac{\vec{D}_1 \cdot \vec{v}_1}{\vec{D}_1 \cdot \vec{N}_1} \quad \text{avec : } \vec{D}_1 = \vec{P} - \vec{C}_1$$

Et la projection de P sur le capteur 2 est aussi donnée par une relation identique :

$$x_2 = f \frac{\vec{D}_2 \cdot \vec{u}_2}{\vec{D}_2 \cdot \vec{N}_2} \quad y_2 = f \frac{\vec{D}_2 \cdot \vec{v}_2}{\vec{D}_2 \cdot \vec{N}_2} \quad \text{avec : } \vec{D}_2 = \vec{P} - \vec{C}_2$$

Afin de déterminer les trois composantes (x, y, z) du point P , il faut résoudre cet ensemble d'équations au sens des moindres-carrés¹.

¹ **Moindres-carrés** : C'est-à-dire résoudre l'ensemble d'équations d'une façon expérimentale

Conclusion

Dans ce chapitre nous avons traité deux méthodes différentes de reconstruction 3D basées images : L'enveloppe visuelle et la stéréovision. Nous avons détaillé la démarche de chaque méthode, son utilité et les problèmes rencontrés au cours de leurs automatisations.

L'enveloppe visuelle permet de générer la forme tridimensionnelle par une intersection des cônes des silhouettes. La forme produite fournit un volume englobant l'objet d'intérêt.

Par contre la stéréovision permet, quand-à-elle, l'extraction de la structure 3D d'une scène à partir de deux images prises sous différents points de vue.

Comme on a vu dans ce qui précède, chacune de ces deux méthodes possède des limitations qui influencent sur le résultat final de la reconstruction.

Si on combine les résultats de ces deux méthodes, on peut avoir des formes 3D plus robustes et très proches de la forme réelle de l'objet. Ceci sera le sujet du chapitre suivant, où on va voir quelques travaux qui se basent sur cette combinaison.

Chapitre IV

Travaux Similaires

Introduction

Dans les chapitres précédents nous avons vu deux méthodes différentes de reconstruction 3D à partir d'images : L'enveloppe visuelle et la stéréovision et, nous avons expliqué leur principe de fonctionnement, les avantages et les difficultés liés à chaque méthode.

L'enveloppe visuelle ne peut pas traiter les zones concaves même si un grand nombre de silhouettes est utilisé [9]. La stéréovision est non utilisable pour les cas des objets faiblement texturés, pour les régions occluses, aussi pour le cas où une partie de l'objet est sphérique ou curvilignes [10]. Ces difficultés causent un appariement difficile (voir impossible), et donnent des résultats non fiables et instables.

Par contre on trouve que ces deux méthodes sont bien complémentaires en nature et peuvent être combinées pour surmonter leurs inconvénients dans le but d'optimiser la qualité de la reconstruction 3D.

L'idée de l'hybridation de la méthode d'enveloppe visuelle avec la stéréovision est déjà abordée. Dans ce chapitre nous allons présenter quelques travaux de combinaison, et nous détaillerons la démarche de ces travaux, les objectifs, les inventions et les limitations des résultats de chaque travail.

2. Travail de Eric Nowark et al

Nowark et al [43] ont proposé une méthode hybride pour la reconstruction 3D d'un joueur de Golf. L'utilisation de la méthode d'enveloppe visuelle seulement induit des erreurs de modélisations des zones concaves, ces erreurs sont réduites par la technique de la stéréovision qui permet de positionner précisément des points dans l'espace.

Les joueurs sont filmés par un ensemble de caméras synchronisées ; qui capturent les mouvements 3D. Il s'agit en premier lieu de modéliser l'enveloppe visuelle du joueur à partir des silhouettes. La méthode de création d'enveloppe visuelle choisie est la méthode volumique. Le processus de sa création est le suivant :

▪ **Calibration :**

Une grille de calibration est utilisée avec une géométrie bien connue et un emplacement visible par toutes les caméras, il est supposé que le centre optique se situe au centre de la caméra. L'orientation des caméras, les positions et leurs distances focales sont donc retrouvées grâce à des outils de calibration avec correspondance 3D-2D, ces outils permettent de fournir :

- Distance focale en cm : f_c .
- Position de la caméra en cm dans le repère monde : T_c .
- Matrice de rotation permettant de passer du repère monde au repère caméra : R .

▪ **Segmentation :**

Cette étape est réalisée manuellement dans le but de traiter la suite du problème d'une façon rapide.

▪ Création de l'enveloppe visuelle :

Pour induire l'enveloppe visuelle, une création par arbres octaux est utilisée, il s'agit d'une méthode volumique. Cette dernière fait appel à des théories issues de divers domaines, le principe est de tester si un cube se projette intégralement à l'intérieur de toutes les silhouettes de l'objet à modéliser.

Si cette condition est vérifiée le résultat d'enveloppe sera correct, si ce n'est pas le cas, ce résultat sera manquant. Cette méthode n'a d'intérêts que si la position d'un cube par rapport à la silhouette peut être déterminée rapidement, sinon il est plus simple de travailler directement avec les voxels (cubes élémentaires).

Cette phase est suivie par un calcul de la résolution maximale de l'enveloppe visuelle à partir de la formule suivante :

$$\delta_{\text{théorique}} = \min_{i=1..n_cameras} \left\{ \frac{L_c d_i}{L_p f_i} \right\}$$

La résolution maximale de l'enveloppe visuelle est la distance $\delta = \delta_{\text{théorique}}$ tels que :

L_c est la largeur en cm du plan film théorique, L_p la largeur en pixels des images, C_i le centre optique de la caméra i , d_i le centre du plan focal physique de cette dernière et f_i sa distance focale. Enfin d_i le point 3D de l'objet à modéliser le plus éloigné de la caméra i .

Remarque : Cette valeur théorique est calculée à l'aide d'une configuration de théorème de Thalès.¹

▪ Transformation de l'enveloppe visuelle en maillage :

L'algorithme de transformation en maillage parcourt tous les voxels composant le volume, et à chaque fois qu'un voisin est absent, une face carrée est créée, identique à la face frontière entre le voxel et l'emplacement de son voisin absent. Cette face est ensuite ajoutée au maillage.

A cette étape, la phase de création d'enveloppe visuelle est terminée, le résultat à ce moment est combiné avec la méthode stéréovision. Les étapes de ce processus est comme suit :

▪ Détection des points caractéristiques :

Pour effectuer la mise en correspondance, des points caractéristiques sont automatiquement extraits des images. Ces points doivent se différencier suffisamment de leurs voisins pour faciliter les appariements. Un pixel (i, j) est défini comme un point caractéristique si et seulement si :

$$D_x = \max \{ |NG(i, j) - NG(i-1, j)|, |NG(i, j) - NG(i+1, j)| \} \geq f$$

Différenciation sur l'axe des abscisses.

Et:

$$D_y = \max \{ |NG(i, j) - NG(i, j-1)|, |NG(i, j) - NG(i, j+1)| \} \geq f$$

Différenciation sur l'axe des ordonnées.

Tels que $NG(i, j)$ est le niveau de gris des pixels (i, j) qui sont compris entre 0 et 1. Le seuil de différenciation minimal est f choisi arbitrairement tel que $0 \leq f \leq 1$.

Enfin pour ne pas avoir plusieurs points caractéristiques, un pixel qui a la plus grande valeur de différenciation est choisi : $\min(D_x, D_y)$.

¹ **Théorème de Thalès :** théorème de la géométrie qui consiste à calculer la longueur des droites.

- **Calcul de la position 3D avec stéréovision :**

Le but de l'algorithme est de calculer la position 3D des points caractéristiques vus précédemment, le principe est de calculer la disparité entre ces points à partir de la méthode SAD (sum of absolute difference) et SSD (sum of square difference). On constate que l'application de cet algorithme permet de placer tous les points à la bonne profondeur sur l'objet 3D.

- **Déformation du maillage.**

Pour obtenir la structure 3D, deux informations différentes sont disponibles. Premièrement, l'enveloppe visuelle définit un nuage de points très dense, dont les positions sont correctes dans les zones convexes, et incorrectes dans les zones concaves, et deuxièmement l'algorithme de stéréovision qui permet de positionner précisément des points dans l'espace. Le maillage est donc déformé selon cet ensemble de points corrects.

Avantages et inconvénients

Ce travail est appliqué sur un type particulier d'objet « Golfeur ». En ce qui concerne l'enveloppe visuelle, si le point de vue du joueur et en face de la caméra, la forme finale de cette enveloppe est parfaitement présentée. Mais si ce n'est pas le cas, si le point de vue créé par une seconde caméra de synthèse où les concavités sont vues de profil on trouve que l'objet 3D résultant est mal modélisé.

La combinaison du résultat d'enveloppe visuelle avec la stéréovision permet d'améliorer quelques parties concaves, mais pas d'une façon précise. Après la déformation du maillage certaines parties deviennent plus visibles, mais on trouve que quelques parties du corps sont mal reconstruites ; comme par exemple la représentation d'un bras plié.

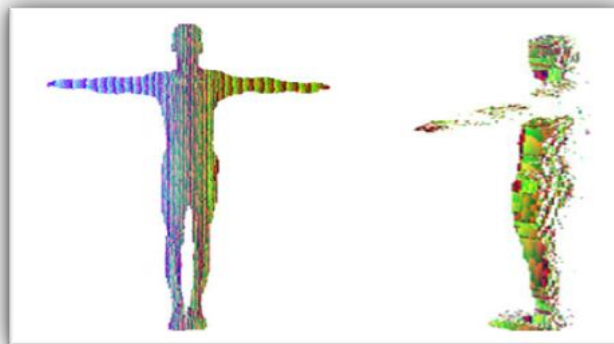


Figure [IV.1] : Exemple des résultats obtenus par l'approche de Nowark et al.

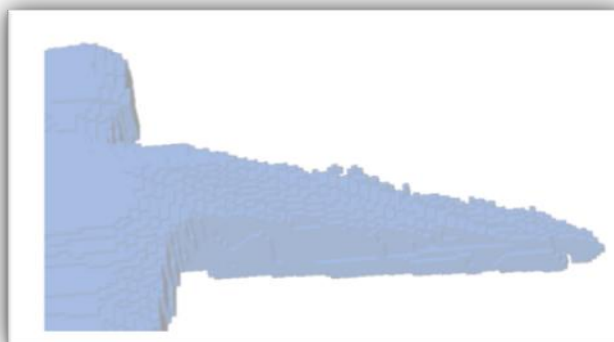


Figure [IV.2] : Exemple de modélisation des zones concaves.

3. Travail de Ming Li et al

Li et al [44] ont adopté le principe de combinaison de la méthode de l'enveloppe visuelle avec celle de la stéréovision pour l'acquisition de scènes réelles dynamiques aux taux interactifs. Tout d'abord ils ont utilisé des silhouettes de plusieurs vues pour construire une enveloppe visuelle qui représente une estimation initiale de l'objet dans la scène. Cette dernière est utilisée par la suite pour limiter la plage de disparité au cours du calcul de la profondeur à partir de la stéréovision, ce qui permet d'améliorer la qualité de la reconstruction stéréo ainsi que la vitesse. L'un des inconvénients de la méthode d'enveloppe visuelle est l'impossibilité de reconstruction des zones concaves et les détails de la surface, c'est pour cette raison que le travail consiste à la combiner avec la stéréoscopie.

Le système implémenté se compose de plusieurs caméras observant la même scène à partir de différentes directions de visualisation, chaque paire de caméras est connectée à un ordinateur client qui communique avec un autre ordinateur serveur via un réseau TCP/IP standard, à savoir que toutes les caméras sont calibrées à l'avance comme il est présenté par :

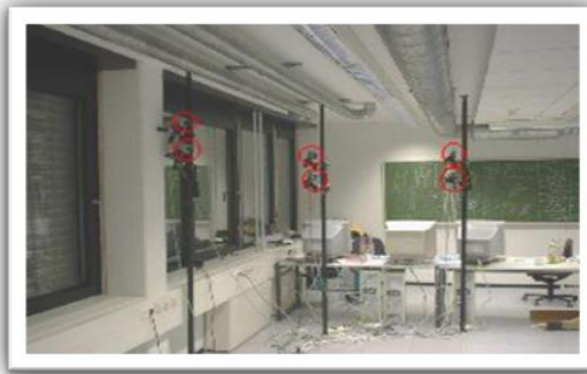


Figure [IV.3] : Trois paires de caméras sont disposées le long d'un arc.

L'initialisation du système comprend l'enregistrement d'une image d'arrière-plan pour chaque caméra et l'envoi d'informations d'étalonnage de chaque client vers le serveur, ensuite le système commence le cycle de traitement qui est défini comme le temps de traitement d'un ensemble d'images synchronisées collectées par toutes les caméras. Selon la direction du transfert de réseau, le processus de ce travail se compose de trois étapes :

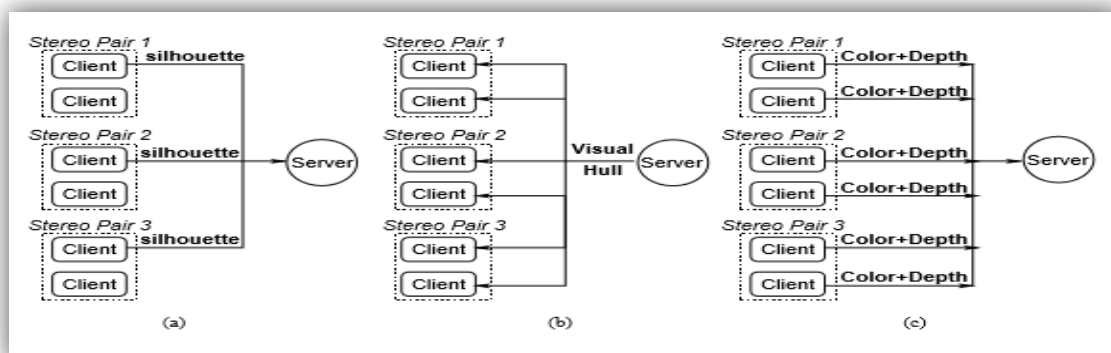


Figure [IV.4] : Démarche de l'approche proposée. (a) : Extraction des silhouettes. (b) : Création d'enveloppe visuelle. (c) Estimation de profondeur à partir la stéréovision.

- **Extraction des silhouettes :**

Tout d'abord, les paires d'images stéréo sont rectifiées individuellement pour s'aligner le long des lignes de balayage. Pour chaque image, l'objet en mouvement du premier plan est segmenté du fond acquis précédemment. Comme les caméras stéréo sont très proches l'une de l'autre, l'utilisation des deux silhouettes n'améliore pas de façon significative la reconstruction de l'enveloppe visuelle. Par conséquent une silhouette est extraite pour une caméra de la paire stéréo, cette silhouette est envoyée par la suite au serveur.

- **Calcul d'enveloppe visuelle :**

Lorsque toutes les silhouettes sont disponibles, le serveur commence le calcul de l'enveloppe visuelle à base d'une méthode surfacique car elle requiert moins de mémoire et convient pour un rendu direct rapide, cette enveloppe est diffusée par la suite à toutes les machines clientes.

- **Calcul de profondeur à partir de la stéréovision :**

Tous les clients utilisent l'enveloppe visuelle pour guider le calcul de la profondeur. A savoir que les informations sur les silhouettes sont disponibles, le calcul de la stéréoscopie ne doit être effectué que sur le masque d'objet au premier plan au lieu que sur l'image entière. Tant que les images sont rectifiées au début, alors le serveur commence directement la phase de la mise en correspondance, à ce moment la méthode SAD est choisie pour le calcul des correspondants.

Une fois que la correspondance entre les pixels est établie, il devient possible de générer la carte de disparité qui à partir de laquelle la carte de profondeur peut être obtenue d'une manière simple à l'aide de l'information de caméra, les cartes de profondeur¹, ainsi que les images en couleur obtenues, sont ensuite renvoyées au serveur pour le rendu.

Avantages et inconvénients

L'architecture du système a été conçue pour distribuer la charge de calcul entre le serveur et les clients. On trouve que le temps nécessaire pour l'acquisition des images, la rectification, la récupération des silhouettes et la reconstruction 3D est indépendant du nombre de de paires stéréo, ce qui offre une bonne évolutivité du travail dans des petites durées de traitement.

Le résultat obtenu dans ce travail est présenté dans la figure suivante :



Figure [IV.5] : Résultat de modélisation de la méthode de Li et al (Exemple 1).

¹ **Carte de profondeur** : Image ou un canal d'image qui contient des informations relatives à la distance entre les surfaces des objets des objets de scène de caméra.

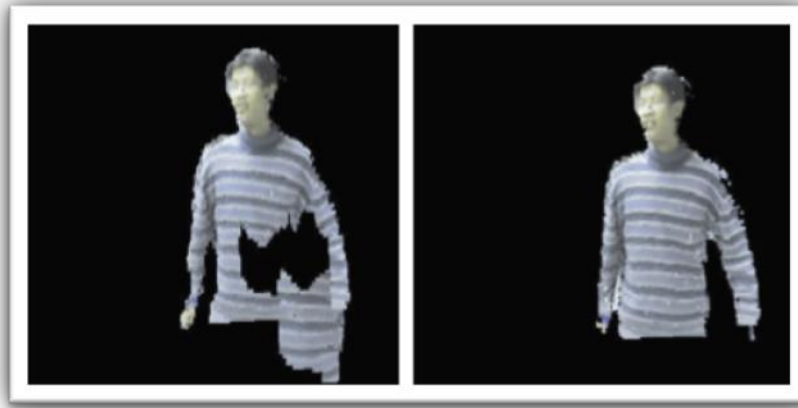


Figure [IV.6] : Résultat de modélisation de la méthode de Li et al (Exemple 2).

4. Travail de Matsumoto et al

Un autre algorithme de reconstruction tridimensionnelle proposé par Matsumoto et al [45], qui se base aussi sur la combinaison des deux méthodes vues précédemment. Dans cet algorithme l'enveloppe visuelle permet d'estimer une forme approximative de l'objet cible, puis la stéréovision est appliquée dans la zone estimée pour redéfinir la forme. Ce dernier remédie aux lacunes de chaque méthode et fournit une forme plus précise et un traitement rapide.

Formalisation du problème :

- Une série d'images de différentes vues tout autour d'un objet cible est en entrée.
- Les positions de la caméra sont connues pour toutes les images, ainsi que les paramètres intrinsèques et extrinsèques.
- La zone probable de la présence d'objet peut être définie dans un espace 3D.
- La description est basée sur les notations suivantes :
 - V : Ensemble de voxels.
 - V_s : Un ensemble de voxels qui correspondent à la surface de l'objet cible dans l'espace voxel ; $V_s \subseteq V$ (Surface d'objet).
 - I^i ; $i^{\text{ème}}$ plan d'image d'entrée, I_o^i et I_s^i signifie les i -ème images d'entrée originales (niveau de gris ou bien couleur) et sa silhouette (B/W), de plus I_{ss}^i et I_{sb}^i sont la silhouette et la zone de l'arrière-plan respectivement.
 - $P(v, i)$: Est la projection du point v dans I^i .

Les images originales et les images de silhouette donnent des indices et des contraintes importantes concernant la zone de présence de l'objet dans l'espace de voxel. Un voxel de V_s est projeté sur la partie silhouette des images d'entrée (contrainte de silhouette). D'autre part, les pixels d'images d'entrée sur lesquels un voxel de V_s est projeté devraient avoir la même couleur ou intensité (contrainte de couleur). En supposant :

- 1) Environnement d'illumination homogène.
- 2) Images sans bruit.
- 3) Surface Lambertian¹.

¹ **Surface Lambertian** : C'est une surface qui reflète la lumière également dans toutes les directions. Les surfaces Lambertiennes sont également appelées réflecteurs isotropes / Uniformes / Parfaitement diffus.

Contrainte de silhouette :

La contrainte de silhouette locale pour $v \in V_s$ dans I^i est définie par :

$$C_s(v, i) = \begin{cases} 1 & \text{si } p(v, i) \in I_{ss}^i \\ 0 & \text{sinon} \end{cases}$$

La contrainte totale de silhouette est :

$$\prod_i \prod_{v \in V_a} C_s(v, i) = 1. \quad (1)$$

Contrainte de couleur :

$$C_o(v, i) = \begin{cases} 1 & \text{si } v \text{ est projeté sur } I^i \text{ sans occlusion} \\ 0 & \text{sinon} \end{cases}$$

$$\sum_{v \in V_a} \sum_{i \in O^v} D(v, i) = 0, \quad (2)$$

Où $D(v, i)$ représente une évaluation de dispersion de couleur pour v . un exemple typique pour $D(v, i)$ c'est :

$$\sum_{j=i-m}^{i+m} \sum_{\text{blockara}} \{v(p(v, i)) - v(p(v, j))\}^2$$

Où $v(p(v, j))$ est la valeur de couleur dans $p(v, j)$.

Le problème cible est de déterminer l'ensemble V_s en utilisant ces contraintes, le but principal de l'approche proposée est de déterminer l'ensemble de voxels localisés V_s à partir de V .

L'enveloppe visuelle permet de produire un ensemble de voxels V_b qui donne une estimation approximative de la zone de présence d'objet en fonction de la condition de contrainte de silhouette.

La stéréovision basée sur les voxels permet de localiser $V_s (\subseteq V_b)$ avec la contrainte de couleur.

Déroulement du processus de l'approche :

- **Enveloppe visuelle basé sur vote :**

Comme il est mentionné ci-dessus, la silhouette d'un objet cible donne des contraintes importantes de la forme de l'objet. Plus précisément, un point dans l'espace 3D projeté dans l'intérieur d'une silhouette représentant l'objet, alors qu'un point qui est projeté sur l'extérieur de la silhouette est certainement libre. L'espace 3D occupé par l'objet peut être estimé plus précisément en utilisant plusieurs images de silhouette prises à différents points d'observation.

La méthode d'enveloppe visuelle est simple et conventionnelle, elle considère les silhouettes comme des contraintes fortes. Cela signifie que la méthode donne des silhouettes avec des erreurs. Par contre dans la méthode utilisée dans ce cas, les erreurs de silhouette sont sérieusement résolues où elles sont traitées comme des contraintes faibles. Dans l'algorithme étendu, l'estimation de forme est réalisée avec un système de localisation de vote (Figure IV.7). Dans ce cas seulement les voxels pleinement marqués sont jugés dans la zone de l'objet.

```

For all voxels:           — (initialization)
   $score(v) := 0.$ 
   $V_b := \phi.$ 
For all voxels:           — (voting)
  For all silhouettes:
    If  $C_s(v, i) \neq 0,$ 
      Then  $score(v) := score(v) + 1.$ 
For all voxels:           — (localization)
  If  $score(v) \geq Threshold,$ 
    Then  $V_b := V_b \cup \{v\}.$ 

```

Figure [IV.7] : Shape from silhouette basé vote.

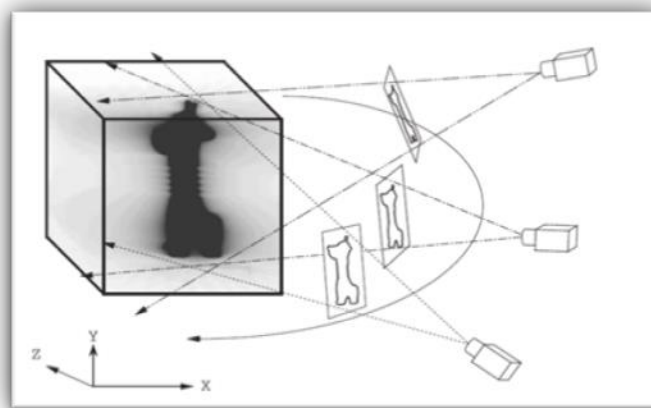


Figure [IV.8] : Vote sur l'espace voxel : la densité montre le score de chaque voxel.

▪ **Stéréovision basée vote :**

Si l'objet cible contient une partie concave, l'enveloppe visuelle ne reconstruit pas correctement la forme 3D. L'utilisation de la stéréovision permet de récupérer la connectivité ainsi que la surface de l'objet et des informations de textures. Cette étape comprend les sous étapes suivantes :

○ **Estimation partielle de surface :**

L'idée de base dans cette phase est : Si un voxel v correspond à la surface de l'objet, la couleur de son point de projection sur chaque plan d'image est similaire à l'autre. Les erreurs de ces couleurs sont accumulées sur chaque voxel (erreur de vote). Ensuite, les voxels avec l'erreur minimale accumulée sont choisis, cet ensemble de voxels peut être déterminé pour chaque point d'observation et est considéré comme des surfaces partielles pour le point d'observation. A ce moment le i -ème point d'observation est référencié à la surface partielle comme V_{ss}^i .

Il est supposé aussi que l'ensemble d'images observables O^v inclut au moins n membres consécutifs pour tout $v \in V_{sp}^i$, ceci est pour tout point d'observation i . Cela signifie que le problème d'occlusion peut être évité pour n images d'entrée consécutives pour tous les voxels de toutes les surfaces partielles.

Soit V_{rv}^i un ensemble de voxels dans le long d'un rayon du i -ème point d'observation à travers v . Pour le i -ième point d'observation, la surface partielle V_{sp}^i est identifiée avec

la procédure représentée dans (a) sur la figure (IV.4), les étapes sont appliquées de façon répétée à tous les points d'observation.

A la fin, un ensemble de surfaces partielles est obtenu comme des images originales.

o **Génération totale de modèle :**

Ce niveau génère un modèle 3D total à partir des surfaces partielles. Bien qu'une surface partielle soit calculée à chaque point d'observation, chacune peut avoir des erreurs. En intégrant les surfaces partielles avec le programme de localisation de vote, les erreurs sont arrondies pour obtenir le modèle total de l'objet avec une plus grande précision.

Soit $L(v, i)$ la distance euclidienne¹ entre v et la i ème point d'observation. V_s est un ensemble de voxels qui donne la représentation volumique de l'objet. Le processus d'intégration est détaillé dans la partie (b) de la figure (IV.4). Cette procédure ne fournit pas la solution exacte qui satisfait à la constante de couleur détendue mais elle donne une bonne approximation.

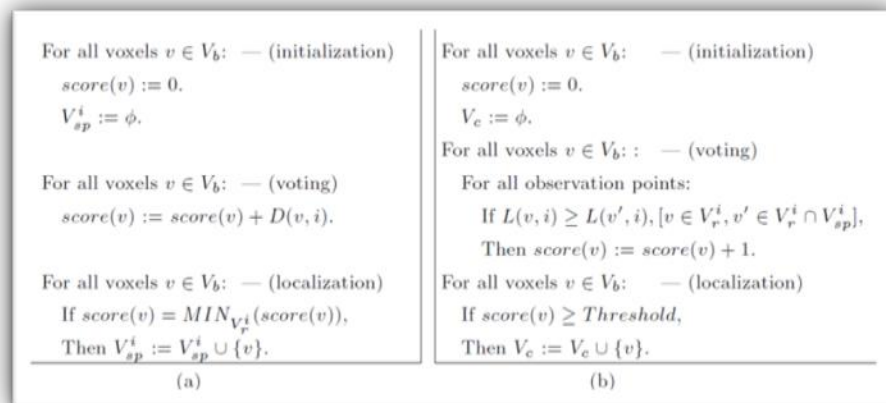


Figure [IV.9] : (a) Estimation de surface partielle. (b) Génération totale de modèle.

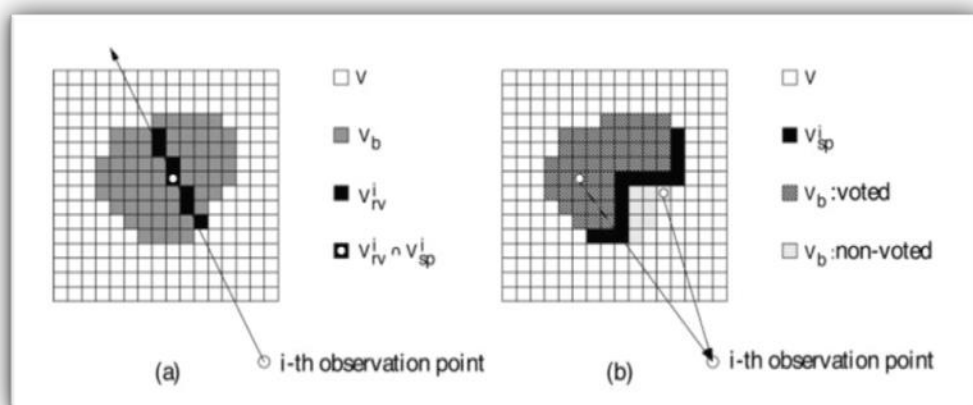


Figure [IV.10] : (a) Estimation de surface partielle (b) Génération totale du modèle.

¹ **Distance Euclidienne** : C'est une distance ordinaire c'est-à-dire entre deux points dans l'espace euclidienne.

▪ Combinaison d'enveloppe visuelle avec stéréovision :

Pour tester l'utilité de l'algorithme adopté, un numériseur 3D est construit. Le numériseur comprend un plateau tournant, une caméra monoculaire et ordinateur personnel. Ce numériseur est capable de générer automatiquement le modèle 3D texturé de l'objet cible.

Ce processus se compose de cinq étapes majeures :

○ Calibration :

Dans ce processus, la position de la caméra par rapport à la table tournante est automatiquement calculée en utilisant un objet de forme connue placé sur le plateau tournant, l'objet est capturé à partir de divers points d'observation (images originales). Dans la mise en œuvre actuelle, un panneau d'étalonnage avec motifs de cercle est utilisé. Après détection de chaque position de cercle, la transformation de Hough [5] est appliquée pour calculer la position de caméra pour renforcer la robustesse.

○ Capture d'images :

Dans cette étape, une série d'images est capturée pour couvrir toutes les vues de l'objet. En outre, l'image d'arrière-plan est capturée aussi. Cette dernière est presque identique aux images d'origine, sauf qu'elle exclut l'objet cible et est utilisée pour extraire des silhouettes.

○ Extraction des silhouettes :

Dans cette étape, l'extraction de silhouette naïve, ou l'extraction au niveau de pixel, est effectuée en premier. Ensuite, extraction de silhouette supplémentaire, extraction de région-niveau, est appliquée, ce qui améliore la précision de la silhouette obtenue. Dans l'extraction au niveau de la région, la moyenne de la soustraction absolue à chaque pixel dans une région est calculée pour juger si la région correspond à l'objet ou non.

○ Modélisation de formes :

L'étape de modélisation de la forme se compose de deux processus : la modélisation volumétrique et la conversion de la représentation de surface. Dans la conversion de représentation de surface, de nombreux polygones initiaux sont générés d'abord en connectant des polygones de surface adjacents. Ensuite, une opération de réduction de polygone fusionne des polygones pour réduire le nombre total de polygones.

○ Acquisition de texture :

La texture d'une pièce peut être dérivée à partir de plusieurs images d'origine, car une pièce peut être vue à partir de divers points de vue. Par conséquent, une image originale doit être identifiée pour chaque polygone pour en dériver sa texture. Afin de déterminer l'image source de manière appropriée, les points suivants doivent être considérés :

- 1) Une plus grande quantité d'informations de texture dans l'image source conduit à une meilleure dérivation de la texture.
- 2) Le lissage de texture sur la limite des pièces est nécessaire pour masquer les bords.



Figure [IV.11] : Présentation du système.

L'algorithme est testé sur des objets en céramique (un château miniature et un écureuil). Dans les expériences, l'étape de rotation de la table était de dix degrés, donc 36 images originales ont été prises. La taille de l'image était de 640 à 480. La résolution de l'espace de voxel est de $320 \times 320 \times 240$.

Les figures ci-dessous montrent des exemples d'images originales de chaque objet, des modèles en fil de fer¹ 3D obtenus et des images 3D texturées en résultats. Les résultats de modélisation montrent que cette approche reconstruit la forme avec une précision pratique. Le temps d'exécution de chaque étape de cette approche est présenté dans le tableau de la figure (IV.10).

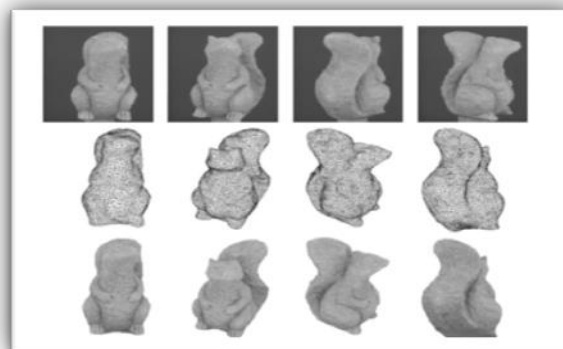


Figure [IV.12] : Exemples d'images originales et de résultats de modélisation (Ecureuil)

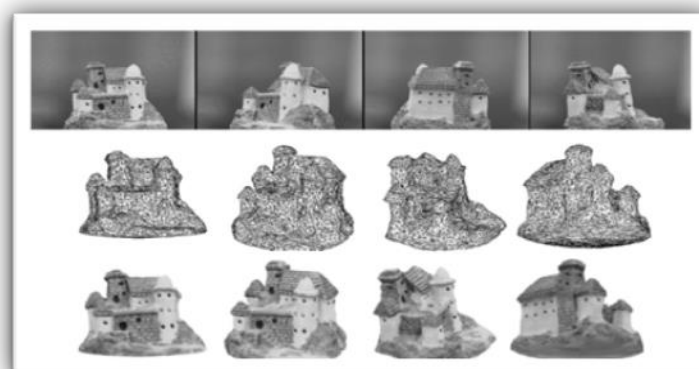


Figure [IV.13] : Exemples d'images originales et de résultats de modélisation (Château).

¹ **Image fil de fer** : C'est un modèle de modélisation, où un objet est représenté par un ensemble de segments, de courbes ou droites.

Tableau [IV.1] : Analyse de temps d'exécution de chaque étape du processus

Etape		Temps d'exécution (sec)	
		Château	Ecureuil
Extraction des silhouettes		53	31
Shape Modeling	Shape-from-silhouette	93	36
	Voting-based stereo	1230	1870
	Surface representation	106	38
Acquisition de texture		43	43

Avantages et inconvénients :

- Les principaux avantages de cet algorithme sont les suivants :
 - Traitement plus rapide que la simple reconstruction à partir de stéréovision.
 - Reconstruction de forme plus précise que la forme simple de l'enveloppe visuelle.
 - Meilleure réduction du bruit que la simple stéréovision.

- L'un des problèmes actuels de l'algorithme est que parfois la surface d'une partie mince peut être manquante en raison d'erreurs dans les surfaces partielles. Cette erreur peut être détectée en vérifiant la cohérence entre deux silhouettes : une générée à partir du modèle électronique et l'autre de la zone d'objet probable estimée par Shape-from-Silhouette.

5. Conclusion

Dans ce chapitre nous avons présenté quelques travaux qui utilisent l'idée de l'hybridation de la méthode de la stéréovision avec l'enveloppe visuelle. Pour chaque travail nous avons expliqué la démarche suivie pour réaliser chaque méthode et le principe de combinaison, nous avons présenté aussi les résultats obtenus par chaque travail ainsi que les avantages et les lacunes de chacun.

Pour le travail *Eric Nowark et al* on trouve que malgré que l'approche soit limitée à type particulier d'objet, mais nous constatons qu'il manque d'optimisation et ne donne pas des formes 3D robustes. De plus la méthode d'extraction des silhouettes utilisée dans cette approche est assez compliquée, ainsi que la forme de silhouette obtenue qui contient d'énormes trous et de bruit, ce qui aboutit par la suite à une enveloppe visuelle non complète. Plusieurs parties sont mal présentées comme par exemple la partie du bras plié, ceci est dû à la mauvaise méthode choisie pour le calcul des silhouettes, ce qui se reflète directement sur le résultat final obtenu. Un autre problème est signalé ; c'est la déformation du maillage d'objet 3D à partir des résultats obtenus par les deux méthodes. Donc la technique utilisée pour la combinaison des deux méthodes reste encore moins efficace.

Pour le travail de *Ming Li et al* une architecture client/serveur est utilisée dont le but de diviser les différents calculs sur un ensemble de machines, ceci permet d'améliorer le temps d'exécution nécessaire pour la réalisation de travail. Mais comme un inconvénient de cette approche, malgré l'utilisation des bibliothèques optimisées et spécifiées pour le traitement d'images, on constate que les résultats obtenus sont inexacts et elles semblent un peu loin de la réalité. En plus on constate que le coût de traitement très élevé malgré la diffusion de travail sur plusieurs machines clients.

Enfin l'approche proposée par *Matsumoto et al* est appliquée seulement sur des images synthétisées. L'algorithme n'est pas vérifié sur des images réelles qui sont généralement plus complexes et s'influencent directement par le monde extérieur. Mais d'une façon générale les structures 3D résultantes sont presque identiques à l'objet d'intérêt.

Dans le chapitre suivant nous allons présenter notre proposition qui traite la même idée de l'hybridation enveloppe visuelle/Stéréovision, mais d'une façon complètement différente à celles présentées dans ce chapitre.

Chapitre V

Une Méthode Hybride pour la Reconstruction 3D à partir d'une Vidéo

1. Introduction

La reconstruction 3D à partir d'images représente un sujet de recherche très actif dans les dernières années, les algorithmes et les méthodes dans ce domaine sont toujours en progression dont le but d'obtenir des résultats plus proches de la réalité dans un minimum de temps. Dans ce qui précède dans ce manuscrit, nous avons présenté les différentes méthodes de reconstruction 3D, ainsi que leur classification multicritère. Notre étude est basée sur deux méthodes spécifiques qui sont : L'enveloppe visuelle et la stéréovision.

L'enveloppe visuelle est une intersection des cônes des silhouettes d'entrée, qui donne des résultats exacts si un grand nombre de silhouettes est utilisé, mais l'enveloppe visuelle reste limitée pour les zones convexes et ne permet pas de traiter les zones concaves.

La stéréovision est une technique qui permet de reconstruire partiellement la structure 3D d'un objet ou d'une scène, à partir de deux images décalées prises de deux points de vue différents. Le but de cette méthode est de récupérer la troisième dimension perdue au cours de la formation des images en calculant l'ensemble des points correspondants. Ce processus est assez délicat surtout dans le cas où la scène contient des parties occluses ou bien des zones avec des textures homogènes. De plus le calcul des couples correspondants est très compliqué et nécessite un temps de traitement très élevé.

La combinaison de ces deux méthodes est un sujet déjà abordé, le but est de surmonter les limites de chaque approche : L'enveloppe visuelle permet de limiter la zone qui englobe l'objet d'intérêt, ce qui est bénéfique pour minimiser la zone de recherche lors de la phase de mise en correspondance de la stéréoscopie qui peut être aussi utilisée pour détecter les régions concaves, et permettre de raffiner le modèle 3D reconstruit et de minimiser les temps de calcul.

Dans ce qui suit nous allons traiter notre approche proposée sous le titre : « **Une méthode hybride de reconstruction 3D à partir une vidéo** ».

Notre objectif est de combiner l'enveloppe visuelle avec la stéréovision mais d'une façon différente à celles présentées dans le chapitre précédent. Nous allons proposer un nouvel algorithme parallèle qui permet d'estimer la forme de l'enveloppe visuelle dans le moindre temps possible. Avant de passer à ce dernier nous allons présenter une technique qui permet d'isoler un objet de son arrière-plan et de calculer son enveloppe visuelle. Par la suite, le résultat de l'enveloppe visuelle sera utilisé dans le processus de la mise en correspondances pour limiter l'espace de recherche et a cet effet les correspondants incorrects.

2. Système d'acquisition de la vidéo

La méthode de reconstruction 3D proposée dans le cadre de notre travail est appliquée sur une séquence d'images extraites à partir d'une vidéo. L'objet à reconstruire est placé sur un moteur qui tourne avec une vitesse fixe, la vidéo est donc capturée à pas angulaire fixe (approximativement de 3 degré). La rotation complète de l'objet résulte un ensemble de 120 images.

Le système de prise de vues que nous avons utilisé est réalisé à l'aide d'un mini studio (inspiré de celui de la figure V.1), qui représente un environnement spécial pour la capture d'objets. L'idée de base est de placer un objet dans une boîte carrée dont l'arrière-plan est de couleur unique (les couleurs généralement utilisées sont le blanc, le bleu ou bien le vert). Des sources lumineuses sont placées dans les deux côtés du mini studio pour éliminer les ombres éventuelles de l'objet.

Dans notre cas nous avons utilisé un arrière-plan de couleur blanche. L'objectif d'utilisation du mini studio est d'avoir, en premier lieu, des images sans effets d'ombre ayant une influence non favorable sur le résultat final de la silhouette qui doit être extraite d'une façon exacte.

La deuxième raison d'utilisation du mini studio est qu'il facilite l'isolation de l'objet de son arrière-plan. La séparation ou bien l'isolation de l'objet de son fond est une opération qui consiste à distinguer la région qui caractérise l'objet. Le fond unique du mini studio réduit la complexité de cette opération et il assure aussi un bon éclairage de l'objet, ce qui permet aux détails de l'objet d'apparaître parfaitement.



Figure [V.1] : Exemple d'un mini studio.

3. Matériels et langage de programmation

Avant de passer à décrire les étapes du processus de la reconstruction 3D, nous allons présenter les matériels et logiciels utilisés pour l'achèvement de notre travail.

Les vidéos que nous avons capturés, résultent de l'utilisation d'un Smartphone **SAMSUNG Galaxy note 5**. Dont sa caméra est d'une qualité de 16 méga pixel, avec une distance focale de 28 mm, une taille d'ouverture $f1.9$ et une taille de capteur de $1/2.6$.

L'ordinateur que nous utilisons est un **LENOVO W540** qui est doté d'un processeur Intel core i7 vPro 4800MQ 2.70 GHz, une RAM de 16 GO et une carte graphique Nvidia Quadro K2100M. Le système d'exploitation installé sur cette machine est WINDOWS 10 Pro 64 bits

Nous avons choisi l'environnement de programmation MATLAB pour l'implémentation des différents algorithmes, la version utilisée est celle du 2016.

3.1 Programmation avec MATLAB

Le nom MATLAB vient du mot anglais MATrix LABoratory. Ce logiciel prend son nom où la plupart de ces fonctions prédéfinies sont des grandeurs matricielles, et par extension pour les données vectorielles. Les domaines d'applications qui peuvent être traités par MATLAB sont très variés, on trouve par exemple : Le traitement du signal, le traitement d'images, l'automatisme...

Ce langage peut être exécuté sur différentes plateformes : LINUX, WINDOWS, ou bien MAC OS. Les programmes sont écrits sous forme de scripts. Lorsque MATLAB démarre, l'IDE (*Integrated Development Environment*) apparaît à l'écran, c'est l'interface graphique du MATLAB (Voir figure V.2). Cette interface est assez claire pour un utilisateur débutant.

Nous avons choisi ce langage de programmation parce qu'il est très utile pour les calculs numériques sur les vecteurs car nous sommes en train de traiter des images qui sont représentées sur machine sous forme de matrices, ce langage est sacrément évolué et contient beaucoup de fonctionnalités dédiées au domaine de traitement image et vision par ordinateur. Il nous aide beaucoup pour le calcul des temps d'exécution des différents codes en utilisant la fonction tic ()¹ et toc ()². De plus MATLAB aide beaucoup pour tracer les données, pour modifier les couleurs, l'échelle des maillages que nous avons réalisé grâce aux outils graphiques disponibles dans ce langage.

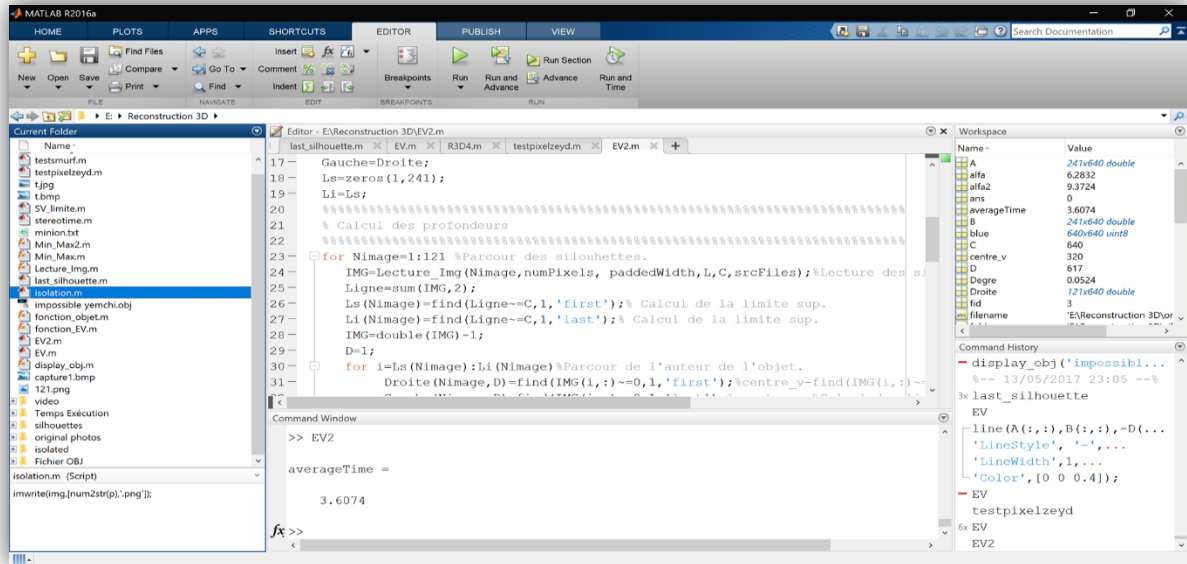


Figure [V.2] : Interface graphique du MATLAB.

4. Démarche proposée du processus de reconstruction 3D

Comme nous avons mentionné ci-dessus, notre approche se base sur la combinaison de la méthode de l'enveloppe visuelle et celle de la stéréovision. L'enveloppe visuelle détecte une approximation de la forme de l'objet. Cette approximation, permet d'accélérer considérablement le processus de la stéréovision en limitant la zone de mise en correspondance.

Le processus de la reconstruction 3D se déroule en cinq phases essentielles :

- 1) L'extraction des images à partir de la vidéo.
- 2) L'isolation de l'objet et extraction de ces silhouettes.
- 3) L'estimation de l'enveloppe visuelle.
- 4) Génération du nuage de points et sauvegarde sous forme d'objet 3D.
- 5) La stéréovision avec limitation de la zone mise en correspondance.

4.1 Extraction des images à partir la vidéo

Le but d'utiliser la vidéo est d'extraire un grand nombre d'images dans seulement quelques secondes. La vidéo filmée est d'une durée de 17 secondes, et comme nous l'avons mentionné précédemment, le degré de rotation du moteur est approximativement 3° , ceci permet d'obtenir un

¹ Tic : Fonction MATLAB qui démarre un chronomètre pour mesurer le temps d'exécution.

² Toc : Fonction MATLAB qui affiche le temps écoulé depuis que la fonction tic a été exécuté.

ensemble de 121 images. Il est connu que chaque vidéo est composée d'un ensemble de cadres (frames en anglais), leur nombre est lié au type de caméra utilisée (Dans notre cas 7 FPS).

La technique utilisée pour l'extraction des images à partir de la vidéo est assez simple, il suffit de lire cette vidéo et de calculer le nombre de frames affichés par seconde, par la suite chaque frame est extrait et transformé en une image d'extension **png** puis sauvegardé dans un répertoire spécifié.

Les images que nous avons extraites sont de taille 640×640.



Figure [V.3] : Quelques images extraites (Objet 1).



Figure [V.4] : Quelques images extraites (Objet 2).

4.2 Isolation d'objet et extraction des silhouettes

Le suivi d'un objet dans une séquence d'images et son isolation de son arrière-plan, est une opération assez complexe et nécessite de spécifier un ensemble de caractéristiques qui permettent de le segmenter dans chaque image.

L'utilisation du mini studio permet de diminuer la complexité de cette tâche à l'aide du fond blanc utilisé. Il facilite, en fait, la séparation de la zone de l'objet en détectant la zone différente du blanc de l'arrière-plan.

Cette phase n'a pas été réalisée simplement, puisque nous avons détecté pendant l'acquisition de la vidéo que l'effet de la lumière et de l'environnement extérieur provoquent un certain bruit, et la couleur de l'arrière-plan n'est pas un blanc pur comme désiré (Le mini studio n'est pas professionnel, et il reste un travail d'amateurs).

Une image couleur est une combinaison de trois matrices essentielles : Rouge, vert et blanc. Afin d'extraire les silhouettes équivalentes, les matrices RGB qui composent l'image générale sont séparées. Pour chaque matrice, nous calculons la valeur minimale et la valeur maximale, et nous utilisons un seuil qui se trouve entre ces deux valeurs. La silhouette est donc obtenue par la comparaison avec le seuil précédemment défini.

Si un pixel de l'image appartient à cet intervalle, cela veut dire qu'il appartient à l'arrière-plan et est donc transformé en blanc. On obtient une image prétraitée comme entrée à notre algorithme qui va convertir chaque pixel en noir s'il est différent au blanc.

La silhouette obtenue est sauvegardée comme image monochrome, dont la matrice représentative contient seulement les valeurs 0 et 1 (0 pour le noir et 1 pour le blanc).

Ce processus est détaillé par l'algorithme suivant :

Algorithme *extraction des silhouettes*

Début

Charger le répertoire d'images

Pour chaque image **faire**

 Lire (image) ;

 Extraire les matrices rouge (R), vert (G) et bleu (B) ;

Pour chaque matrice **faire**

Pour chaque élément des 3 matrices **faire**

Si (élément < seuil_min) && (élément > seuil_max) **alors**

 élément : =blanc ;

Sinon

 élément : =noir ;

Fin si

Fin pour

Fin pour

 image := Regrouper les 3 matrices R,G,B

Fin Pour

Fin.

Les résultats obtenus après application de l'algorithme d'isolation sont les suivants :

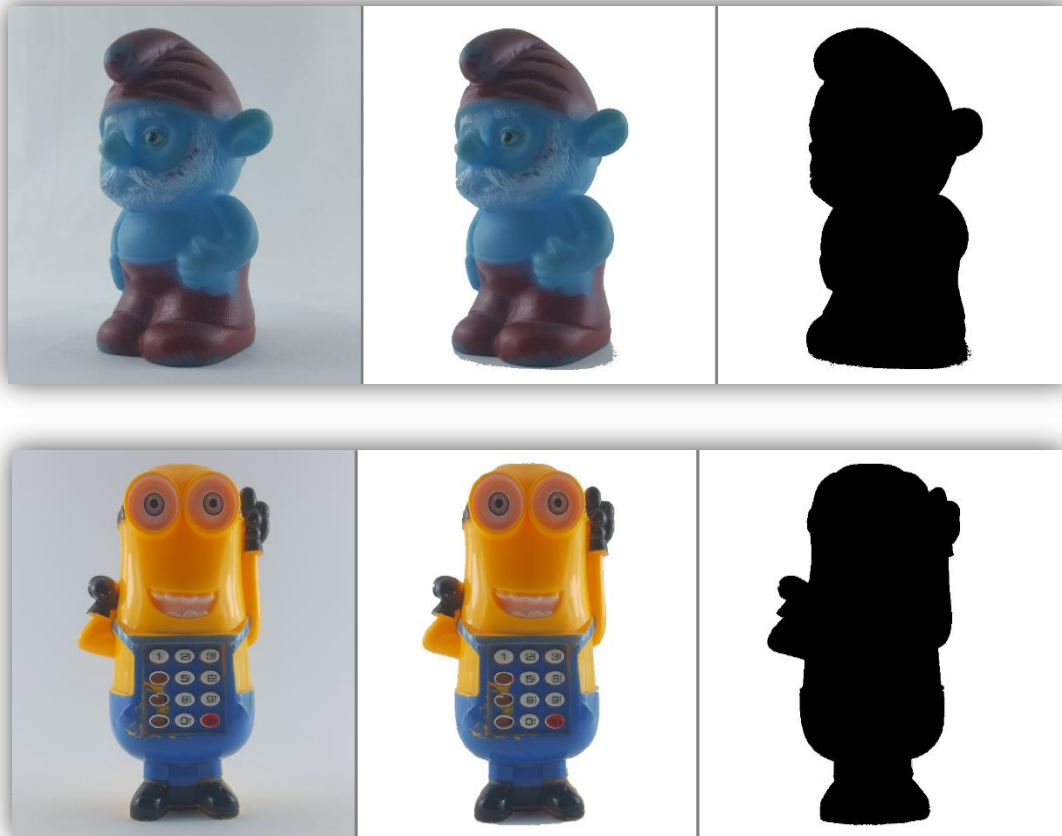


Figure [V.5] : De gauche en droite : L'image originale, image prétraitée et la silhouette obtenue (En haut Objet 1 et en bas Objet 2).

Le déroulement de ce processus est réalisé d'une façon parallèle, les images sont ainsi traitées simultanément et non pas de façon séquentielle. Ceci permet de diminuer considérablement le temps de calcul.

Nous avons détecté que le temps de calcul nécessaire pour traiter toutes les images séquentiellement est environ **44.15** secondes, par contre lorsque on a introduit la notion du traitement parallèle le temps est devenu seulement **16.76** secondes, ceci est pour le premier objet (mignon). Pour le deuxième objet (Smurf) le temps écoulé pour le traitement séquentiel de toutes les images est environ **65** secondes, mais avec l'utilisation de l'algorithme parallèle le temps devient **45** secondes.

L'organigramme ci-dessous résume le fonctionnement de la méthode d'extraction des silhouettes que nous avons abordé, et précise les phases où le parallélisme peut être intégré.

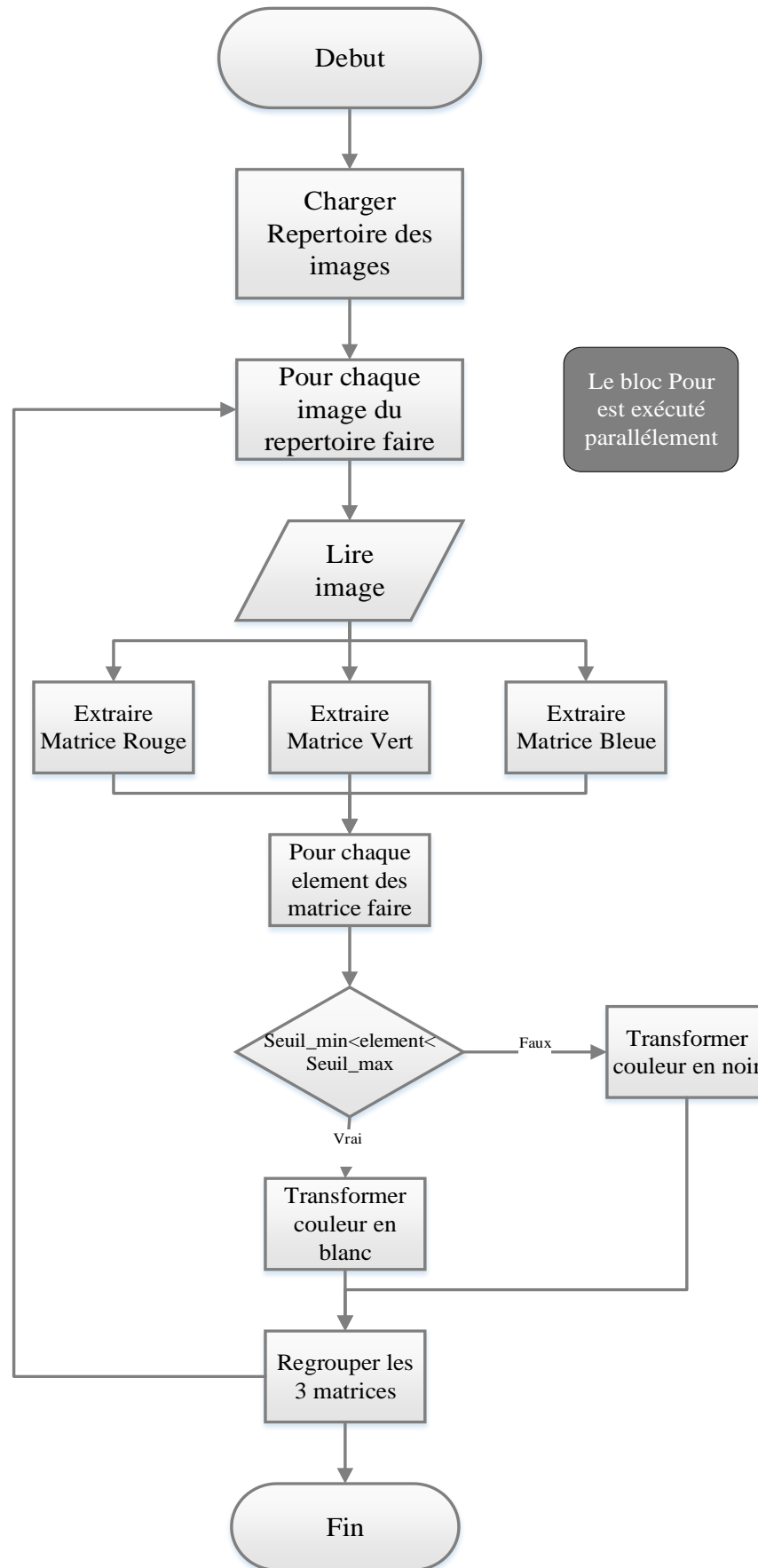


Figure [V.6] : Organigramme de l'algorithme extraction des silhouettes.

4.3 Estimation de l'enveloppe visuelle

L'enveloppe visuelle est calculée à partir de l'ensemble des silhouettes extraites. Plus le nombre de silhouettes augmente, plus le résultat est plus proche de la forme de l'objet réel.

Dans notre cas nous possédons un ensemble de 120 images de silhouettes, et nous n'avons utilisé que 60 d'entre elles. Chacune est utilisée pour le calcul des profondeurs qui représentent les positions 3D de l'objet d'intérêt.

Une silhouette de numéro k est utilisée pour calculer à la fois les profondeurs de l'image k ainsi que celles de l'image $k+60$. Ceci vient du fait que le côté gauche de l'image $k+60$ représente le côté droit de l'image k ; puisque c'est une rotation de 180° ($3*60$). Il est donc inutile de refaire les calculs pour deux images symétriques inversées (Voir figure V.7).

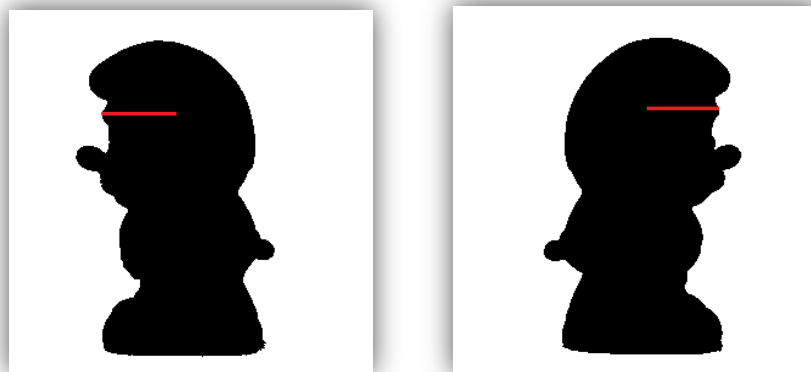


Figure [V.7] : Symétrie inversée de deux silhouettes.

Avant de passer au calcul des profondeurs qui représentent les coordonnées Z perdues lors du processus de formation d'image, nous avons limité la zone de calcul, en cherchant la limite supérieure L_s et la limite inférieure L_i qui représentent les extrémités où l'objet commence (de haut) et termine (en bas) (Voir la figure V.8).

Pour calculer les limites supérieures et inférieures, nous avons réalisé un balayage ligne par ligne de la première silhouette. Le premier pixel noir détecté en descendant, représente le début de l'objet du haut, le numéro de la ligne où se trouve ce pixel représente la limite supérieure. Le même travail est fait pour calculer la limite inférieure, mais cette fois le balayage de la silhouette s'effectue à partir du bas de l'image. Ces deux valeurs sont les mêmes pour toutes les images et le calcul des profondeurs sera limité uniquement pour cette zone.

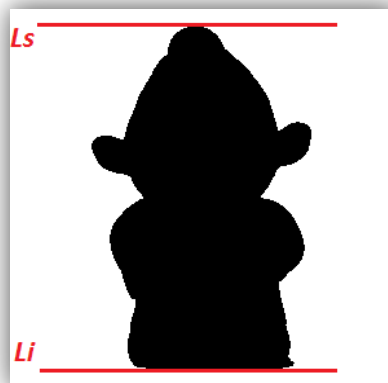


Figure [V.8] : Représentation des limites supérieure et inférieure.

La fonction qui permet de calculer les limite supérieure et inférieure est donnée par :

```

Fonction calcul_limiteSupInf(silhouette) :
Début
    Pour chaque ligne de l'image faire
        tab := Calculer somme des colonnes;
    Fin pour ;
    Pour i=debut_tab:fin_tab faire
        Si somme de colonnes==Largeur d'images alors
            Passer à la ligne suivante ;
        Sinon
            Limite supérieure = indice de ligne ;
        Fin si ;
    Fin pour ;
    Pour i=fin_tab : debut_tab faire
        Si somme de colonnes==Largeur d'images alors
            Passer à la ligne suivante ;
        Sinon
            Limite inférieure = indice de ligne ;
        Fin si ;
    Fin pour ;
Fin.

```

Comme nous avons mentionné ci-dessus, il existe une symétrie inversée entre les images. Alors, pour le calcul des profondeurs il suffit de le faire pour seulement la partie gauche de l'image.

La récupération de la profondeur nécessite une détection des limites gauches et limites droites. Ces dernières représentent les extrémités de l'objet de la partie gauche et droite (En d'autres termes les bordures de l'objet). Pour inférer les limites gauches nous avons fait un parcours horizontal (en commençant par le début d'image) de la partie entre L_s et L_i (Limite supérieure et inférieure). Pour chaque ligne on cherche à trouver le premier pixel noir, ceci représente la limite gauche pour un tel ligne. Pour les limites droites le même principe est réalisé mais dans ce cas le parcours commence par la fin de l'image. L'algorithme de cette phase est le suivant :

```

Fonction limite_gauche_droite(image)
Début
[limite_supérieure, limite_inférieure]=sup_inf(image) ;
  Pour ligne allant de (limite_supérieur) jusqu'a (limite_inférieure) faire
    Tant que colonne <> taille d'image et pixel <> noir faire
      Colonne := colonne+1 ;
      Si pixel est noir alors
        Stocker la position du pixel dans le tableau limite_gauche ;
      Fin si ;
    Fin tant que ;
  Fin pour ;
  Pour ligne allant de (limite_inférieure) jusqu'a (limite_supérieure) faire
    Tant que colonne <> 1 d'image et pixel <> noir faire
      Colonne :=colonne-1 ;
      Si pixel est noir alors
        Stocker la position du pixel dans le tableau limite_droite ;
      Fin si ;
    Fin tant que ;
  Fin pour ;
Fin.

```

A ce moment l'estimation de la profondeur P_i revient à calculer la distance entre le centre vertical de l'image et la limite gauche correspondante à la ligne i comme il est présenté dans la figure [V.9] (a). Il s'agit de la soustraction de la valeur de centre vertical et la valeur de la limite gauche de la ligne i . Ce processus est réalisé pour chaque ligne appartenant à la zone allant de L_s jusqu'à L_i et ce pour toutes images et pour les 60 images. Pour calculer les profondeurs de l'image numéro 61 par exemple, il suffit d'appliquer la même procédure sur la partie droite de la silhouette numéro 1 (Calcule de la distance entre le centre vertical et la limite droite qui correspond à la ligne d'intérêt).



Figure [V.9] : De gauche en droite la profondeur de la ligne (i) calculée à partir de la partie gauche et droite.

A cette étape les profondeurs calculées représentent les coordonnées X des points appartenant à la frontière de l'objet. Le numéro de la ligne représente, quand-à lui, les coordonnées Y. En fin, le numéro de la silhouette analysée représente les coordonnées Z (en appliquant une rotation). Ces valeurs sont stockées dans une matrice *Gauche* pour les calculs appliqués sur la partie gauche et une autre matrice *Droite* pour les calculs appliqués sur les parties droites

La position du voxel 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 exactement le volume de l'objet.

Par conséquent, ce nuage de points peut être sauvegardé et manipulé comme objet 3D par des logiciels de modélisation 3D, dans ce cas nous pouvons appliquer sur le nuage de point obtenu des rotations sur les axes X, Y, Z, nous pouvons aussi appliquer un agrandissement ou atténuation. L'affichage des voxels 3D en espace est réalisé à l'aide de la fonction `plot ()`, `plot3 ()`, `mesh ()` de MATLAB.

Pour l'estimation des positions exactes des pixels dans l'espace, nous avons utilisé une matrice de transformation dans laquelle nous avons mis la matrice de la transformation de la rotation sur l'axe Z, cette matrice est multipliée directement avec la matrice *Gauche* et aussi la matrice *Droite* en prenant en considération le degré de rotation (en radian). L'algorithme suivant représente les différentes étapes suivies dans le processus de la création de l'enveloppe visuelle :

Algorithme Estimation d'enveloppe visuelle

Début

Charger le répertoire des images ;

Pour image1 jusqu'à image 60 **faire**

 Lire (image) ;

 [*Ls*, *Li*]:=calcul_supinf (image) ;

Pour *i* allant de *Ls* jusqu'à *Li* **faire**

 Calculer les profondeurs de la partie gauche et les stocker dans matrice *Gauche* ;

 Calculer les profondeurs de la partie droite et les stocker dans matrice *Droite* ;

Fin pour ;

Fin pour ;

Trans := Calculer matrice de rotation ;

Degré := angle de rotation en radian ;

Multiplier les matrice *Gauche* et *Droite* avec la matrice Trans et Degré ;

Positionner les coordonnées X, Y et Z résultante dans l'espace ;

Fin.

Concernant le temps de calcul dans le cas de l'enveloppe visuelle, nous avons détecté que la durée nécessaire pour obtenir le résultat final après le traitement des 60 silhouettes, est seulement d'environ **1.69** secondes. (Si on utilise toutes les 120 silhouettes, le temps de calcul devient environ **3.05**), ces résultats concernant l'objet (mignon). Pour l'objet (Smurf) le temps écoulé en utilisant 120 images égale à **2.46** secondes, par contre lorsque nous avons appliqué l'algorithme sur 60 images le temps de calcul devient **1.9** secondes.

Malgré la complexité des calculs, la méthode proposée nous a permis de gagner beaucoup de temps, l'utilisateur ne détecte aucune lenteur pour afficher le résultat. De plus le nuage de points obtenu est assez satisfaisant et une grande précision est fournie. L'enveloppe visuelle nous a donné une très bonne approximation du volume de l'objet reconstruit.

Les figures ci-dessous représentent les résultats que nous avons obtenus en appliquant l'algorithme sur deux objets différents.

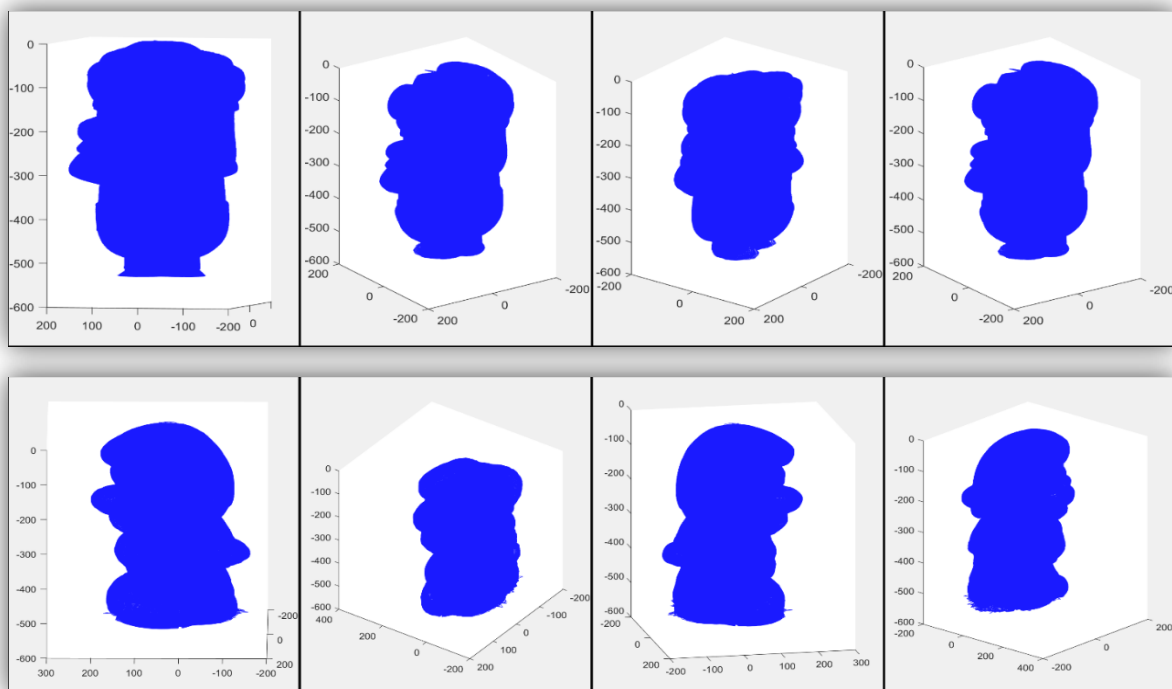


Figure [V.10] : Résultats d'enveloppe visuelle obtenus par MATLAB.

4.4 Génération du nuage de point et sauvegarde en format (obj)

La méthode de l'enveloppe visuelle que nous avons développée nous a permis d'obtenir un nuage de points manipulable par des logiciels de modélisation 3D. Il suffit pour cela de le sauvegarder au format (.obj). Ce format contient une description sur la géométrie 3D. Un fichier .obj est écrit en code ASCII, le tableau suivant décrit les différentes notations utilisées dans ce fichier :

Tableau [V.1] : Différentes notations des fichiers obj

Notations	Signification
o nom	Utilisé pour affecter un nom à l'objet
v (1.0 1.0 1.0)	Utilisé pour créer un sommet.
vt (1.0 1.0)	Utilisé pour définir la coordonnée de texture
l (v1, v2)	Utilisé pour créer une ligne entre 2 sommets
vn (0.0 1.0 0.0)	Utilisé pour définir une normale
f (v1, v2, v3)	Utilisé pour créer une facette entre 3 sommets

Pour réaliser le processus de sauvegarde en format obj, nous avons utilisé les résultats obtenus lors de la phase de calcul des profondeurs. En fait, nous avons les coordonnées X, Y et Z calculées précédemment pour chaque vertex et il suffit de générer un fichier texte respectant la syntaxe des fichiers «.Obj». L'idée est de positionner les différents vertices, et ensuite créer une ligne entre chaque deux vertices adjacentes, l'algorithme de cette étape est comme suit :

Algorithme sauvegarde format_obj :

Début

Créer un fichier texte vide ;

Extraire les coordonnées x, y, z calculés dans la phase d'enveloppe visuelle ;

Pour chaque voxel **faire**

 écrire dans le fichier texte ('v' (x, y, z))

Fin pour ;

Pour i allant de voxel 1 jusqu'à voxel **faire**

 écrire ('f' (voxel[i], voxel [i+1])) ;

Fin pour ;

Fin ;

Après la création de fichier .Obj, ceci peut être manipulable par différents logiciels de modélisation. Nous avons utilisé le logiciel « Blender » pour importer nos deux fichiers créés. On peut donc appliquer n'importe quelle opération sur l'objet reconstruit (création des facettes, plaquage de texture, coloration...). La figure ci-après montre l'un des objets manipulés par Blender.

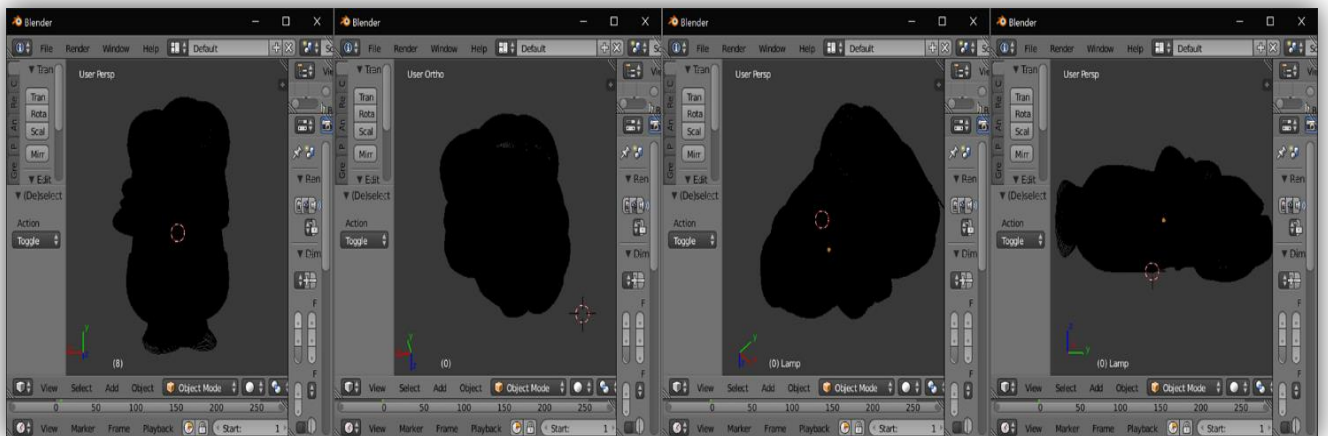


Figure [V.11] : Importation d'objet reconstruit en Blender.

4.5 La stéréovision avec limitation de la zone mise en correspondance

La méthode de l'enveloppe visuelle que nous avons réalisée est suivie par une méthode de stéréovision. Ce processus est réalisé en suivant trois étapes : Calibration, mise en correspondance et triangulation, comme on a vu dans le troisième chapitre (section 3).

Dans notre cas l'objet d'intérêt est positionné perpendiculairement par rapport à la caméra qui est bien alignée, donc les paramètres intrinsèques sont fixés, à ce moment il n'est pas nécessaire de réaliser la phase de la calibration. La calibration a pour rôle de déterminer les paramètres de la scène (matrice fondamentale et essentielle) pour nous permettre de passer à la phase de triangulation et l'obtention des points 3D, dans notre cas en a pas besoin de la calibration puisque nous sommes en train de calculer la carte de disparité seulement. Notre intérêt est donc de passer directement à la phase de mise en correspondance.

▪ Méthode de mise en correspondance choisie :

Pour le calcul des couples correspondants nous avons utilisé une approche locale d'appariement- détaillée dans le troisième chapitre (section 3.2.5.3) -. Le principe de base est de choisir une fenêtre de taille bien définie dans la première image et de chercher la fenêtre correspondante dans la deuxième image.

Notre objet tourne avec un angle de 3° , et cette rotation implique que les lignes épipolaires ne soient pas strictement parallèles. Et d'autre part, l'algorithme appliqué pour la mise en correspondance nous oblige à spécifier la distance sur laquelle nous cherchons. A cet effet, nous avons cherché les correspondants seulement dans le sens horizontal (puisque l'angle est très petit).

Afin de traiter le problème du petit désalignement angulaire, la recherche de la fenêtre correspondante est faite de droite en gauche - et vice-versa - sur la même ligne épipolaire. Cette idée est appliquée après plusieurs tentatives et expérimentations et en se basant sur les résultats des cartes de disparité obtenues. (Nous avons constaté que les résultats sont erronés lorsque nous balayons l'image dans un seul sens, et par test nous avons compris l'origine de la faute que nous avons corrigée par un nouveau balayage dans le sens inverse).

▪ Taille de fenêtre de corrélation :

Les objets que nous avons utilisés contiennent beaucoup de zones homogènes ainsi que des parties mal texturées, ce qui rend la recherche des correspondants assez délicate. C'est pour cette raison que nous avons choisi une fenêtre de corrélation de taille très petite : Taille de 3×3 pixels. Nous avons essayé d'augmenter la taille de cette fenêtre, dans ce cas le résultat de la carte de disparité obtenu contient moins de bruit mais beaucoup de détails sont perdus, par contre le coût de traitement se diminue avec des modèles plus grands.

▪ Valeur de disparité :

Pour la valeur de la disparité qui exprime le décalage des pixels de l'objet, nous avons choisi de façon expérimentale une valeur entière égale à **5** qui représente la distance en pixel entre l'emplacement d'une région dans une image et son emplacement dans l'autre image. Des valeurs plus grandes augmentent énormément le temps de recherche et donnent des résultats erronés, ceci est aussi dû à l'homogénéité des objets utilisés.

▪ Calcul de disparité :

La méthode utilisée pour réaliser la phase de la mise en correspondance se base sur la méthode **SAD** (Sum of Absolute Differences), la similarité entre deux blocs différentes est

obtenue en calculant la différence absolue entre chaque pixel de bloc dans la première image et le pixel correspondant de bloc de la deuxième image, ce processus est répété pour calculer la différence absolue entre le bloc de la première image et tous les blocs qui se trouvent dans la même ligne épipolaire dans la seconde image, à ce moment parmi les différentes valeurs obtenus, la valeur de différence minimale signifie que ces deux blocs sont correspondants.

▪ **Carte de disparité :**

L'algorithme de la mise en correspondance permet d'obtenir une carte de disparité qui est une image numérique qui contient les informations sur les correspondances par pixel entre deux images. Le contenu de cette carte signifie que les pixels avec de grandes disparités sont plus proches de la caméra (donc éloignés de l'arrière-plan de l'objet), et que les pixels avec des disparités plus petites sont éloignés de la caméra (plus proches de l'arrière-plan de l'objet). Et c'est ici qu'intervient la calibration (matrices essentielles et fondamentale) pour permettre une triangulation et de suite le positionnement exact des voxels dans l'espace).

Algorithme stéréovision originale

Début

Charger répertoire d'images ;

Pour i allant de 1 jusqu'à dernière image **faire**

Lire image(i) ;

Lire image(i+1) ;

Convertir les 2 images en niveau de gris ;

Affecter la valeur de disparité ;

Spécifier la taille de fenêtre de corrélation ;

Pour ligne allant de 1 jusqu'à hauteur **faire**

Pour colonne allant de 1 jusqu'à largeur **faire**

Positionner la fenêtre de corrélation dans la première image ;

Chercher la fenêtre correspondante dans l'autre image dans la même ligne ;

Sélectionner meilleure résultat

Fin pour ;

Fin pour ;

Fin pour ;

Fin.

La méthode de la stéréovision que nous avons travaillée avec utilise la globalité de l'image pour la recherche des couples correspondants en appliquant un balayage horizontal et vertical de tous les pixels de l'image, la zone de recherche est donc très grande ce qui implique des complexités en temps de calcul.

La partie qui nous intéresse pour extraire la carte de disparité est la partie qui contient notre objet seulement, il n'est pas nécessaire de prendre en considération la partie qui représente l'arrière-plan. Pour détecter la zone qui englobe l'objet seulement nous avons utilisé l'idée qui consiste à une hybridation de l'enveloppe visuelle avec la stéréovision.

L'élimination de la zone de recherche avec l'enveloppe visuelle permet de réduire les anomalies des correspondants incorrectes, ainsi que le bruit qui peut apparaître en dehors de la frontière de résultat. De plus cette combinaison a un très grand avantage qui est la diminution de temps de traitement.

Dans la phase de l'enveloppe visuelle nous avons calculé les limites gauches et droites ainsi que les limites supérieure et inférieure, ces valeurs représentent la forme exacte de l'objet utilisé. A ce moment au lieu de parcourir toute la largeur et la hauteur de l'image, on utilise seulement la partie

allant de la limite supérieure jusqu'à la limite supérieure en vertical, et en horizontal le parcours est dû en commençant par la limite gauche d'une ligne jusqu'à la limite droite de la même ligne, ceci est réalisé pour chaque ligne dans la partie verticale parcourue.

Comme nous avons travaillé avec une fenêtre de corrélation de taille 3×3 pixels, pour chaque trois lignes nous avons calculé la valeur minimale des limites gauches de ces trois lignes, et aussi en calculant la valeur maximale des limites gauches, ces deux valeurs représentent les extrémités où en va chercher les correspondants.

L'algorithme de calcul des valeurs minimales et maximales ainsi que la stéréovision limitée sont les suivants :

```
fonction [min,max] = Min_Max (tab_gauche,tab_droite)
```

```
Début
```

```
Pour i allant de limite_sup jusqu'à limite_inf faire
```

```
    min(i)=calculer le minimum des 3 lignes de tab_gauche
```

```
    max(i)=calculer le maximum des 3 lignes de tab_droite
```

```
    passer au 3 lignes suivantes
```

```
fin pour
```

```
fin
```

Algorithme stéréovision

Début

Charger répertoire d'images

Pour i allant de 1 jusqu'à dernière image **faire**

Lire image(i) ;

Lire image(i+1) ;

Convertir les 2 images en niveau de gris ;

Affecter la valeur de disparité ;

Spécifier la taille de fenêtre de corrélation ;

Pour chaque trois ligne **faire**

Calculer les valeur min et max ;

Pour ligne allant de Ls jusqu'à Li **faire**

Pour colonne allant de min jusqu'à max **faire**

Positionner la fenêtre de corrélation dans la première image ;

Chercher la fenêtre correspondante dans l'autre image dans la même ligne ;

Choisir la fenêtre avec la disparité minimale comme fenêtre correspondante ;

Calculer une estimation de disparité par une interpolation ;

Fin pour ;

Fin pour ;

Fin pour ;

Fin pour ;

Fin

La carte de disparité obtenue comme résultat de cet algorithme est représentée dans la figure suivante :

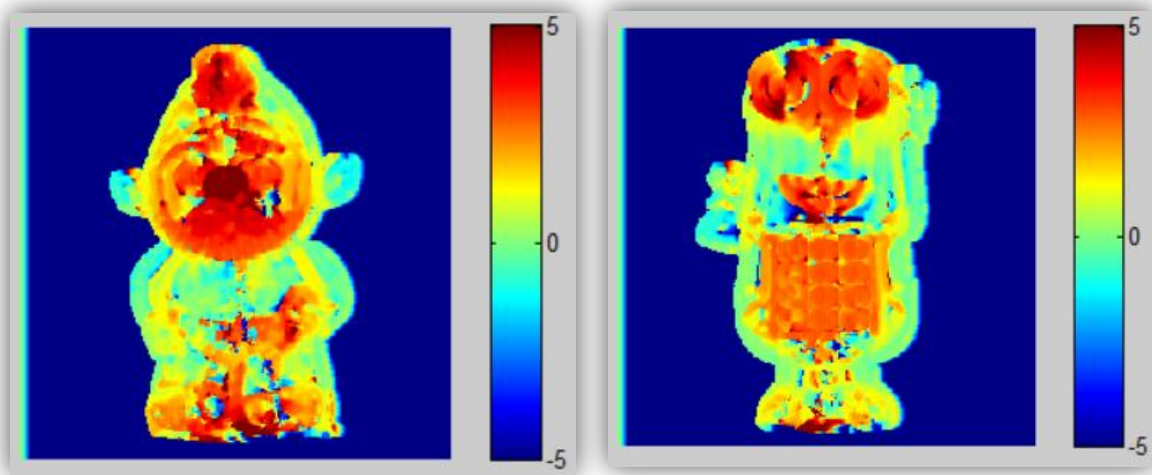


Figure [V.12] : Cartes de disparité résultantes des objets utilisés.

Le degré de couleur représenté dans la figure [V.12] signifie que l'objet est plus proche de la caméra pour les régions rouges, le jaune jusqu'à la couleur vert veut dire que la zone est plus encore éloignée. En arrivant à la couleur bleue qui désigne que cette partie est la plus écarté par rapport à la caméra.

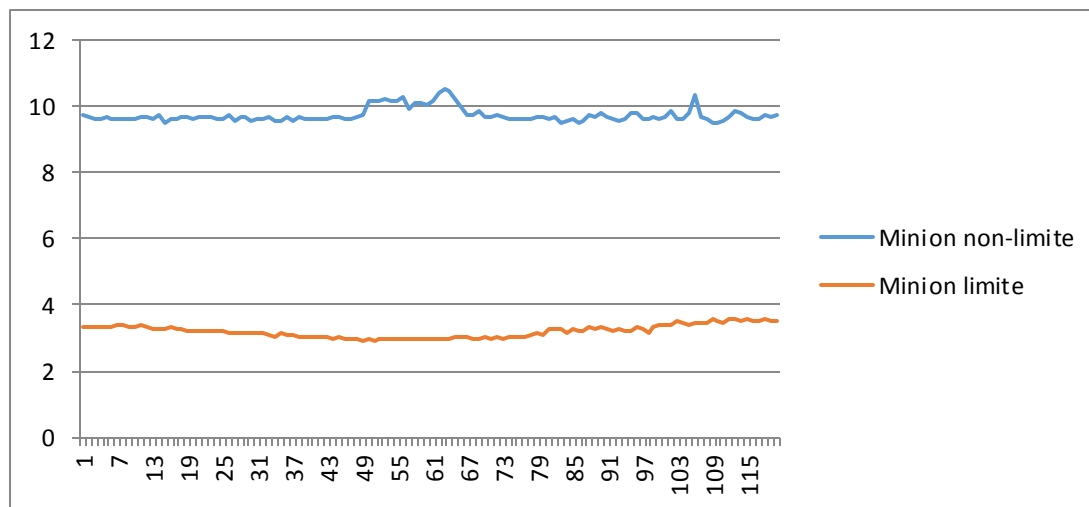
Après l'implémentation des deux algorithmes de stéréovision, l'original et celui limité par l'enveloppe visuelle que nous avons calculée. Nous avons détecté que la combinaison de ces deux méthodes a un grand avantage en ce qui concerne la réduction du coût de recherche des points correspondants et aussi dans l'élimination des fausses correspondances.

Le tableau suivant résume les différentes valeurs comparatives des temps de calculs obtenus pour ces deux approches :

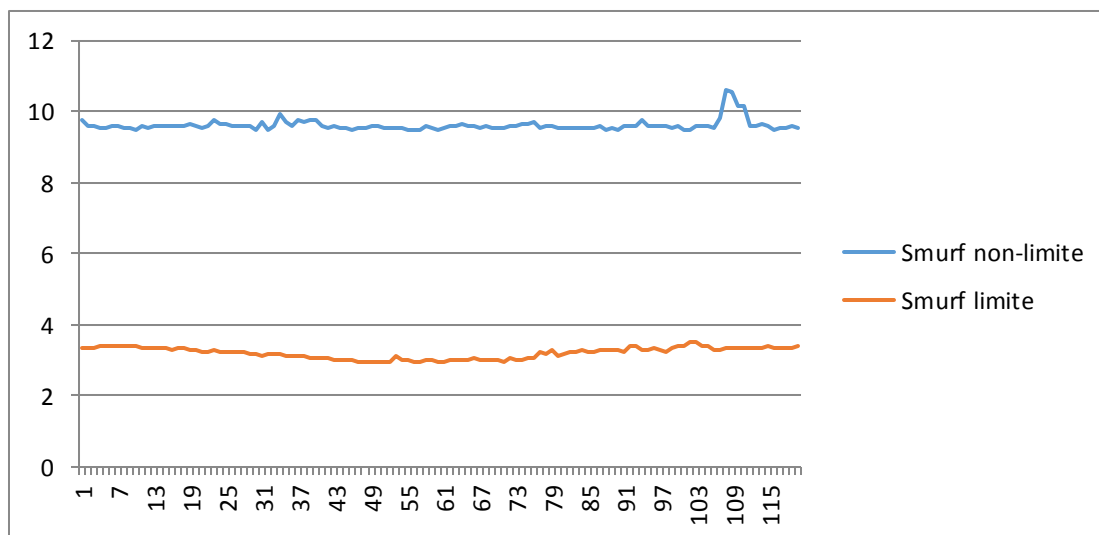
Tableau [V.2] : Tableau comparatif de temps de calcul de stéréovision avec stéréovision limitée

Tâche	Temps de calculs en secondes	
	Minion	Smurf
Min non limité	9,49	9,45886
Min limité	2,93	2,951388
Max non limité	10,51	10,59226
Max limité	3,60	3,525198
Moyenne non limité	9,73	9,605716
Moyenne limité	3,20	3,208618
Total non limité	1168,49	1152,686
Total limité	386,72	387,2942
Taux de réduction	66,90%	66,4%

Pour plus de détail, les graphiques suivant illustrent de façon plus claire les différentes améliorations apportées sur les temps de calcul :



(a) Minion



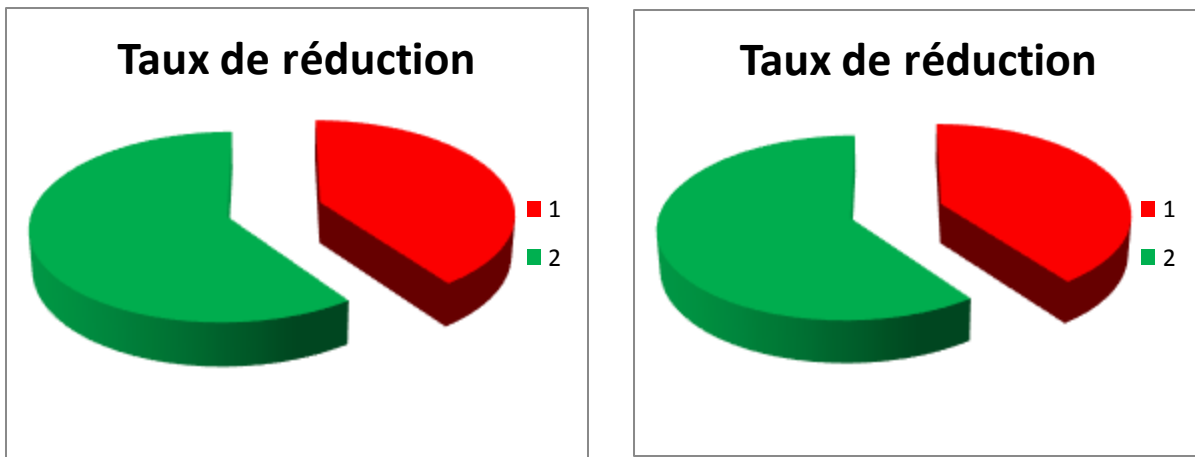
(b) Smurf

Figure [V.13] Courbes comparatives des temps de calculs : (en rouge, le temps d'exécution avec limitation), (en bleu, le temps d'exécution sans limitation).

Les courbes ci-dessus, montre clairement la différence des temps de calculs des correspondants entre l'algorithme de stéréoscopie sans limitation et celui limité par la méthode que nous proposons. Les lignes en rouges (stéréoscopie limitée), n'ont pas dépassé le seuil de **3.7** secondes pour les deux objets, tandis que les lignes en bleu (stéréoscopie non limitée), atteignent jusqu'à **10.4** secondes. Ce qui valide notre hypothèse de réduction de temps de la reconstruction 3D.

De même, les diagrammes suivants représentent des valeurs statistiques intéressantes concernant les temps d'exécution avec et sans limitation de la zone de recherche.

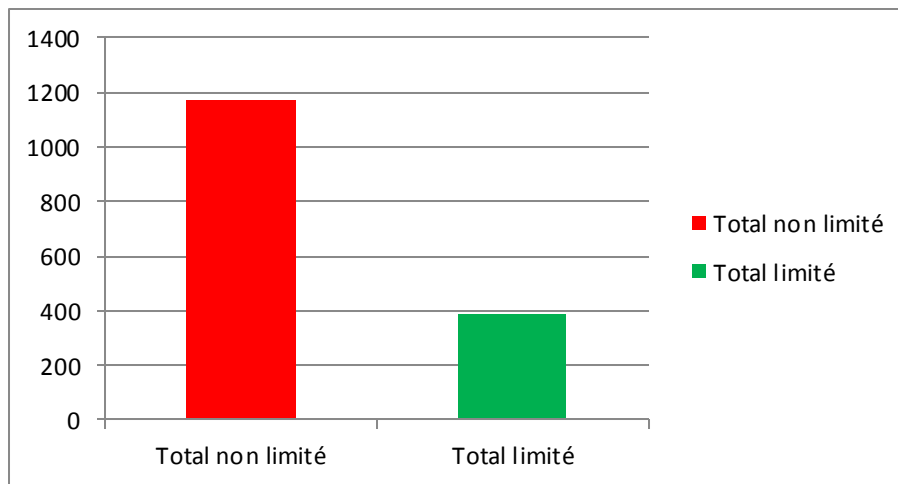
Les deux premiers, représentent les taux globaux de réduction des temps de calcul pour les deux méthodes appliquées sur les deux objets, les deux autres représentent successivement, les temps globaux et moyens de traitement :



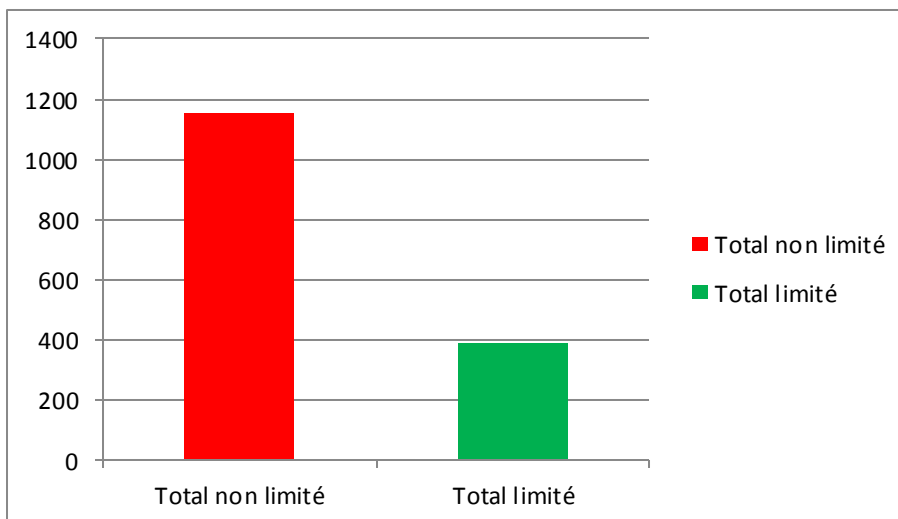
(a) Minion

(b) Smurf

Figure [V.14] : Diagramme circulaire représente le pourcentage de minimisation de temps de calcul d'algorithmme proposée.



(a) Minion



(b) Smurf

Figure [V.15] : Diagramme de secteur représente le temps de calcul écoulé par la stéréovision et la stéréovision limitée.

Conclusion

Dans ce chapitre nous avons proposé une méthode qui combine l'enveloppe visuelle et la stéréovision pour la reconstruction 3D en utilisant une vidéo.

Nous avons bien détaillé la démarche de notre travail en commençant par l'extraction des images à partir d'une vidéo des objets réels à l'aide d'un mini studio, nous avons présenté aussi la technique utilisée pour isoler l'objet de son arrière-plan, ce qui nous a permis d'extraire les silhouettes de l'objet. Dans cette étape on a utilisé un algorithme parallèle dans le but de minimiser le temps de calcul au maximum.

Par la suite nous avons exposé notre algorithme d'enveloppe visuelle où on a utilisé 121 silhouettes qui est relativement un grand nombre, ce qui nous a permis d'estimer une forme 3D très proche de la réalité sans affecté le temps d'exécution.

Après l'estimation de l'enveloppe visuelle, nous avons enregistré les résultats dans des fichiers « OBJ » qui peut être ouverts et manipulés par des logiciels de modélisation 3D.

Le résultat de l'enveloppe visuelle est utilisé par la suite pour limiter la zone de recherche dans la phase de la stéréovision, cela a permis de minimiser le temps d'exécution plus de la moitié.

Enfin nous avons présenté les résultats obtenus et quelques statistiques pour la preuve de notre proposition.

Conclusion Générale et Perspectives

Nous nous sommes intéressés à la problématique de la reconstruction 3D à partir d'images qui appartient au domaine d'infographie et de vision par ordinateur, et qui représente un sujet de recherche très actif dans les dernières années.

La modélisation 3D classique consiste à générer des modèles 3D en suivant deux phases essentielles, une phase de modélisation et une phase de rendu. Cette approche est assez complexe, surtout dans le cas des grandes scènes, de plus les logiciels de modélisation sont coûteux et nécessitent une phase d'apprentissage et des utilisateurs experts.

Dans ce mémoire nous avons présenté quelques notions de base liées au domaine tel que : le système de coordonnées, le processus de la formation de l'image et le modèle de caméra. Nous avons aussi présenté le principe de la reconstruction 3D à partir d'images qui est connue par le terme *IBMR (Image Based Modeling and Rendering)*. Ce type d'approches consiste à calculer la troisième dimension à partir des informations disponibles dans les photographies avec le moindre de coût et avec une meilleure qualité.

En s'introduisant à ce domaine, nous avons présenté une classification multicritère des différentes méthodes de reconstruction 3D existantes, cette classification de bien choisir la méthode de reconstruction 3D qui convient au cas étudié.

Parmi les méthodes *IBMR* existantes, on trouve la méthode de l'enveloppe visuelle et celle de la stéréovision. Après l'étude que nous avons réalisée sur ces deux méthodes, nous avons constaté que ces deux approches ont quelques limitations qui peuvent être traitées si on les combine ensemble pour obtenir des objets 3D plus réalistes et dans des temps minimaux.

Avant de passer à la méthode proposée, nous avons présenté quelques travaux qui utilisent l'idée de la combinaison de l'enveloppe visuelle avec la stéréovision, en détaillant le déroulement de chaque travail et les résultats obtenus.

Dans notre cas, nous avons proposé une méthode hybride de reconstruction 3D qui consiste aussi à combiner l'enveloppe visuelle avec la stéréovision. L'enveloppe visuelle est une première forme de l'objet qui limite au maximum l'espace englobant cet objet. Cette forme est utilisée par la suite dans la phase de la stéréovision pour réduire les calculs inutiles en dehors de la zone de l'objet.

Pour la validation de notre approche, nous avons capturé deux vidéos de deux objets réels à l'aide d'un mini studio. L'utilisation de vidéo facilite la récupération instantanée d'un grand nombre d'images, et nous avons extrait un ensemble de 121 photographies pour chaque objet.

Pour l'extraction des silhouettes, nous avons adopté un algorithme parallèle (c'est-à-dire traiter les images d'une façon parallèle) qui permet de séparer l'objet de son arrière-plan, ces silhouettes sont utilisées par la suite pour l'estimation de l'enveloppe visuelle.

L'algorithme de l'enveloppe visuelle implémenté, donne comme résultats des formes très proches de celles des objets à reconstruire et l'utilisateur ne détecte aucune lenteur. Le problème rencontré dans cette approche est qu'elle ne peut pas traiter les zones concaves sur sa surface, cette anomalie est raffinée avec la stéréovision qui est de sa part optimisée par l'enveloppe visuelle qui limite les correspondances incorrectes en appliquant la recherche seulement au pixels qui se trouvent dans la bordure de l'enveloppe visuelle. Nous avons obtenu un excellent taux de réduction du temps de calculs qui a atteint 69 % du temps global requis sans limitation de la zone de recherche.

Notre but principal était de réduire le coût du traitement le plus que possible. Après l'implémentation d'un algorithme de stéréovision qui utilise la globalité de l'image et notre

algorithme qui combine l'enveloppe visuelle avec la stéréovision, nous avons montré que l'hybridation de ces deux méthodes permet de minimiser le temps de calcul avec un taux de 67%.

Ceci n'a pas seulement réduit le temps de calcul, la carte de disparité que nous avons obtenue ne contient aucun bruit ou de fausses correspondances, malgré l'utilisation d'objets avec beaucoup de zone homogènes qui sont difficiles à différencier par le processus de la mise en correspondance qui est assez délicat dans le cas des zones homogènes et des objets mal texturés. D'autre part, l'introduction de la notion de parallélisme a permis d'accélérer quelques tâches comme par exemple l'extraction des silhouettes. De plus l'utilisation d'un grand nombre de silhouette nous permet de d'aboutir à des résultats d'enveloppe visuelle satisfaisants.

Parmi les perspectives immédiates de notre travail, on peut citer les suivantes :

- Nous envisageons aussi d'implémenter un algorithme d'isolation d'objet de son arrière-plan sans avoir besoins d'un mini studio.
- Nous espérons de compléter l'algorithme de sauvegarde au format obj, en ajoutant des facettes entre les différents vertices. De cette façon, l'objet 3D résultant peut-être même imprimé facilement avec des imprimantes 3D.
- Et enfin nous comptons d'implémenter une application Android de reconstruction 3D en temps réel.

Références

- [1] E. BOYER, 'Cours vision par ordinateur', M2IICAO-UFRIMA.
- [2] B.G BAUMGART, 'Geometric Modeling for Computer Vision'. PhD thesis, University of Stanford, 1974.
- [3] D. MARR, 'A computational investigation into the human representation and processing of visual information', Freeman, 1982.
- [4] R. HARTLEY et P. STRUM, 'Triangulation', CVIU, 1997.
- [5] S. DUBUISSON, 'Plaquage de texture, Algorithmique graphique et modélisation', 2008.
- [6] Y. MATSUMOTO, T. FUJIMURA, T. KITAMURA, 'shape from silhouette/, Stereo and its application to 3D Digitizer', JAPAN, 1999.
- [7] H. ISHIKAWA, D. GEIGER, 'Occlusions, discontinuities and epipolar lines in stereo', ECCV, 1998.
- [8] JN. BEURY, 'Système de coordonnées cartésiennes, cylindriques, sphériques', Système de coordonnées pages 35-500.
- [9] J. CROWLY, 'Cours de formation et analyse d'image', ENSIMAG 3 - Option IRV, 2010.
- [10] V. BRANDOU, 'Stéréovision Locale et Reconstruction 3D/4D', Université de Nice-Sophia, 2008.
- [11] B. BOUFAMA, 'Reconstruction tridimensionnelle en vision par ordinateur' : cas des caméras non étalonnées, L'institut national polytechnique de Grenoble, 2004.
- [12] P. Sturm, 'Quelques note pour le cours de la vision par ordinateur', Equipe-projet PERCEPTION.
- [13] C. ZHANG, T. CHEN. A Survey on Image-Based Rendering. Technical Report AMP 03-03 Advanced Multimedia Processing Lab Carnegie Mellon University Pittsburgh, 2003.
- [14] 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.
- [15] P. DEBEVEC, C. J. TAYLOR, J. MALIK. Modeling and rendering architecture from photographs: A hybrid geometry and image-based approach. SIGGRAPH, 1996.
- [16] Manuel M. OLIVEIRA, 'Image-Based Modeling and Rendering Techniques: A survey', UFRGS, Porto Alegre, RS, Brazil.
- [17] R. ZHANG, Ping-sing. TSAI, J. Edwin et M. SHAH, 'Shape from shading: A survey', University of central Florida.
- [18] B.K.P. Horn. 'Obtaining shape from shading information'. In P.H. Winston, éditeur, The Psychology of Computer Vision. McGraw-Hill, New York, 1975.
- [19] A. MEYER, H. BRICENO PULIDO et S. BOUAKAZ, 'User-guided Shape from Shading to Reconstruct Fine Details from a Single Photograph'. Asian Conference on Computer Vision, 2007.
- [20] G. HARO, 'Shape from Silhouette Consensus', Université Pompeu Fabra, Espagne.

- [21] A.M. LOH et R. HARTLEY, 'Shape from non-homogeneous, non-stationary, anisotropic, perspective texture'. Proc. Vision Conference, pages 69–78, 2005.
- [22] A. BOTTINO, A. LAURENTINI, 'A silhouette based technique for the reconstruction of human movement', 2001.
- [23] N. ASADA, H. FUJIWARA, T. MATSUMAYA, 'Edge and depth from focus', 1998
- [24] P. DAMIEN, 'les imposteurs', 2005.
- [25] D. ROUABHIA, M.C. BABAHENINI, 'Multi-criteria classification of 3D reconstruction approaches', Mohamed Khider University, Biskra, Algeria, 2009.
- [26] A. BOULCH Reconstruction automatique des maquettes numériques 3D Thèse d'université Paris. 2014.
- [27] N. CHAHATA, F. JUNG, M. Pierrot-Deseilligny, J. STAMON, 'Détection et reconstruction de facettes 3D par approche hiérarchique par régions à partir de couples d'images satellite en milieu urbain', Conférence internationale (Science of electronic), Tunisie, 2004.
- [28] M. LEVOY, 'The Digital Michelangelo Project: 3D Scanning of Large Statues', Proc. Siggraph, ACM, 2000.
- [29] E. SAMAIN, 'Téléométrie laser', France
- [30] D. SCHARSTEIN, R. SZELISKI, 'High-accuracy stereo depth maps using structured light', Middlebury Coll, VT, USA, 2003
- [31] D. NISTÉR, 'Automatic passive recovery of 3D from images and video' 3DPVT, 2004.
- [32] A. LAURENTINI, 'The Visual Hull Concept for Silhouette Based Image Understanding', IEEE Trans, Pattern Anal, 1994.
- [33] T. MATSUMAYA, S. NOBUHARA, T. TAKAI, T. TUNG, '3D video and its application', 2012
- [34] C. R. WERN, A. AZARBAYEJANI, T. DARRELL, et A. P. PENTLAND, 'Real-time tracking of the human body', IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997.
- [35] L. BENOIT, 'Enveloppe visuelle et reconstruction 3D', 2009.
- [36] R. CHARLES, 'Volumetric scene reconstruction from multiple views. Foundations of image understanding', Boston, 2001.
- [37] R. HORAUD, O. MONGA Vision par ordinateur outils fondamentaux. Hermès, 1995.
- [38] Z. hengyou ZHANG, 'A flexible new technique for camera calibration', IEEE Trans. Pattern Anal. Mach. Intell, 2000.
- [39] G. GALES, 'Mise en correspondance de pixels pour la stéréovision binoculaire par propagation d'appariements de points d'intérêt et sondage de régions', Université Toulouse 3, 2011.

- [40] N. FAKHFAKH, 'Détection et localisation tridimensionnelle par stéréovision d'objets en mouvement dans des environnements complexes : application aux passages à niveau', Ecole centrale de Lille, 2012.
- [41] R.A. Hummel et S.W. ZUCKER, 'On the foundations of relaxation labeling processes', IEEE Transactions on Pattern Analysis and Machine Intelligence, 1983.
- [42] S. Roy et I.J. Cox, 'A maximum formulation of the n-camera stereo correspondence problem', In Proceedings of 6th International Conference on Computer Vision, 1998.
- [43] E. NOWARK, 'Amélioration d'une reconstruction 3D par voxels à l'aide de la stéréovision', 2003.
- [44] Ming Li, H. SCHIRMACHER, M. MAGNOR et H. P. SEIDEL, 'Combining Stereo and Visual Hull Information for On-line Reconstruction and Rendering of Dynamic Scenes', Germany, 2002.
- [45] G. BERTRAND, M. CUPRIE et L. P. EROTTON, 'Discrete Geometry for Computer Imagery', 8th International Conference, Marne-la-Vallée, France, 1999.