

LA FORME 3D À PARTIR DE CHAMPS DE GRADIENT
ET DES POINTS DE CONTRÔLE.

Présentée par
Fernando Graça

POUR OBTENIR LE GRADE DE
MASTER OF ARTS
AT
UNIVERSITÉ PARIS 8
PARIS, FRANCE
JUN 2011

© Fernando Graça, 2011

UNIVERSITÉ PARIS 8
DÉPARTEMENT DE
UFR ARTS ET TECHNOLOGIES DE L'IMAGE

Dated: Juin 2011

Directrice:

Marie-Hélène Tramus

Examineur:

Anne-Laure George-Molland

Jean-François Jégo

Table des matières

Table des matières	v
List of Figures	vii
Résumé	ix
Introduction	1
1 Éléments de base	7
1.1 La caméra	8
1.2 Caméra sténopé	9
1.3 Projective Geometry	10
1.4 Étalonnage de caméra	12
1.5 Structure à partir du mouvement	13
1.6 Résumé	14
2 L’approche photométrique	15
2.1 État de l’art	15
2.2 Définitions radiométriques	16
2.2.1 La forme à partir de la photométrie	18
2.2.2 La normal à partir de l’image	18
2.2.3 Profondeur à partir du gradient	20
3 Fusion de donnés	23
3.1 État de l’art	24
3.2 Le système	26
3.3 Aperçu du Processus	27
3.3.1 Déformation de la maillage	29
3.4 Résultats expérimentaux	32

3.4.1	Contexte de l'évaluation	32
3.4.2	Génération de données de synthèse	36
3.5	Résumé	36
4	Discussion et conclusion	38
	Bibliographie	40

List of Figures

1.1	Modèle sténopé de caméra. Le plan image I c'est un plane 2D déterminé par x_I, y_I . Le point p c'est la intersection de $(0, 0, f)$ du axe optique avec le plan image.	8
2.1	Dans cette image je représente les trois angles sur la surface S . V, L et N sont la direction de la caméra, lumière (le rayon) et la normal de la surface. Le angle incident i c'est le angle entre le rayon et la surface; le angle émergent c'est la somme des angles i et g	16
2.2	La disposition des lumières dans la scène.	17
2.3	Le résultat de l'utilisation de la distribution des lumières de la Fig.2.2.	18
2.4	Les gradients x et y calculées.	19
2.5	Une surface triangulaire élémentaire définit dans le espace \mathfrak{R}^3 . On peut écrire la relation suivante, liant les coordonnées cartésiennes A, B et C , $\vec{N} = \vec{AB} \otimes \vec{AC}$	20
2.6	La représentation du déplacement dans la matrice.	21
3.1	(a) démontre les nuages des points vues à partir de la caméra dans une plan 2D. L'image (b) c'est le résultat final de la photométrie stéréo vu à partir de la caméra en (a). On projette, ensuite, la nuage des points sur la surface d'objet (c). En utilisant la représentation <i>Laplace</i> d'objet chaque point va applique des déformations sur la surface, (d) démontre l'objet obtenu.	26

3.2	Pour les résultats expérimentaux j'ai considéré la scène (a). L'image (b) c'est le résultat du algorithme structure à partir du mouvement.	27
3.3	Le système.	28
3.4	À gauche c'est l'objet original, au mieux la surface avec des déformations issues des points de contrôle, et à droite la surface sans déformations.	29
3.5	Pour les résultats j'ai sélectionné dans la séquence d'images le'intervalle de l'image 34 à 54, cette séquence utilise une écart-type 2. Les images générées ont une taille de 1024×768 pixels et ont été pris avec le système de capture muni de trois lumières (voir l'image Fig. 3.2 (a)).	33
3.6	Le résultat de la surface calculé après de la déformation.	34
3.7	Ici je présent quelques images de reconstructions 3D produites à l'aide de mon système. L'image (d) met en évidence les limitations de la déformation à partir des points de contrôle.	35

Résumé

Ce projet propose une méthode pour reconstruire la surface 3D à partir d'une caméra en mouvement avec trois lumières attachées de couleurs différents. Ce algorithme intègre un champ de gradients et une carte de points de contrôle pour évaluer la surface 3D.

Le problème est formulé d'une manière qui concilie photométrie stéréo et *structure-from-motion* (SfM). Le premier fait une estimation initial de la surface 3D de chaque image de la séquence en utilisant la radiométrie de la scène. Le deuxième rendre une carte de points de disparité — points de contrôle — de la scène capturé. Chaque point encode l'information du volume de la scène et effectue des déformations en utilisant la représentation laplacien de la surface. Ce modèle permet de préserver les détails de la géométrie autant que possible. Ils améliorent ainsi la précision du volume de la géométrie finale. En utilisant les avantages des ces deux techniques, on peut explorer le maximum d'information de la scène. Le premier donne une représentation dense de la scène avec de gros erreurs de volume, est particulièrement approprié aux objets de surface lisse. D'autant que le deuxième permet de améliorer les premières estimations et convient aux objets de surface rugueuse.

De ce fait, on obtient une collection de maillages qui sont, après, fusionnées afin d'obtenir le modèle final. Les expérimentations menées sur des séquences d'images de scènes virtuelles démontre l'efficacité et les inconvénients de cette approche. Les expériences on montrées que algorithme est de calcul intensif, en revanche la simplicité du dispositif peut être facilement mettre en place, est peut coûteux et a la capacité de capturer des scènes extensifs.

Introduction

Lors le processus de formation d'une image, l'environnement 3D est projeté sur un support 2D, le plan de l'image. L'information 3D est donc perdue. Un des problèmes majeures de la vision par l'ordinateur est de restituer la structure tridimensionnelle des scènes.

Généralement, les méthodes qui se basent sur capteurs optique sont classés en deux catégories: passive et active. La vision passive se base sur l'acquisition des images de la scène à des points de vue différents. Alors que la vision active consiste à combiner des caméras avec des sources de lumière structurée.

L'adjectif passive vient du fait qu'on ne considère que la perception des radiations visibles de la scène. Les méthodes proposées dans cette catégorie tendent mettre à profit les indices visuels qu'on peut trouver dans une image et à analyse géométriques. La structure 3D de la scène réelle et les paramètres intrinsèques et extrinsèques sont estimés en utilisant techniques algèbres pour la reconstruction de la scène. On peut donner l'exemple de structure à partir du mouvement (ou *structure-from-motion*, SfM, dans la littérature anglo-saxonne), liée au déplacement de la caméra; triangulation de laser; stereo-vision; analyse de l'ombre.

Les techniques qui utilisent des principes photométriques repose sur l'analyse radiométrique de l'image, évalue la distribution de l'intensité dans une image pour générer la surface 3D. On peut considérer les méthodes *shape from shading* (SfS) et stéréo photométrie. En supposant que la source de lumière directionnelle est connue la loi *Lambert* donne des solutions pour calculer la normal sur la surface en chaque point en connaissant la réflectance de la surface. En plus ce dernières méthodes font

des reconstruisons lisses.

Dans les méthodes passives les scanners n'ont pas besoin de contrôler l'éclairage. La vision stéréoscopique c'est un exemple on utilise deux images pris à partir de angles de vue différent. Pour estimer la position d'un point dans la scène, on trouve des correspondances entre images, après on utilise la géométrie de triangulation [Turk and Levoy, 1994, Curless and Levoy, 1996]. On peut appliquer cette méthode a plusieurs angles de vue. Les problèmes de ces méthodes sont scènes avec des surfaces lisses sans texture qui pose de problèmes pour le calcul de correspondances. Les méthodes classiques de stéréovision atteignent leurs limites notamment lorsque la scène contient des objets transparents. En effet, pour tels objets, la mise en correspondance devient prohibitive car la mesure tiendra fortement compte la scène placée derrière les surfaces transparents. L'algorithme livrera des résultats erronés.

Plusieurs alternatives ont été proposés, en utilisant des conditions de lumière bien contrôlés: l'arrière-plan et silhouette bien définis. On peut donner l'exemple *shape-from-silhouette* [Laurentini, 1994], *depth-of-field* ou on modifie l'ouverture de la lentille, en utilisant le *point-spread-function* (PSF). En photographie le focus contrôle la distance qui semble plus net dans la scène et n'affecte pas la luminosité de l'image. Si nous voulons que le objet soit net il doit se trouver dans le *Depth Of Field* (DOF). Plus on s'éloigne de la DOF, plus de détails sont perdus. Il y a également un effet secondaire qui provoque des distorsions géométriques. L'ouverture de la lentille (*aperture*) contrôle la quantité de lumière pénètre dans l'appareil photo. Il produit deux effets, première, grande ouverture recueille plus de lumière. Deuxièmement, grande ouverture augmente la quantité de flou conduisant à la réduction de la profondeur de champ. Un objet placé au long de la surface où l'image est mise au point est précisément imagé, tandis que d'autres loin de cet objet sont floues. Le niveau de flou est en relation directe avec la distance entre la surface, où l'image est mise au point, et la position spatiale de l'objet dans la scène. De cette façon, en comparant plusieurs photos avec différents niveaux de focus (par modification de l'ouverture de la lentille)

on peut estimer la profondeur de chaque pixel en analysant le flou local. Les PSF d'un système défocalisé optique agit comme un filtre passe-bas. Pour cette raison, l'image doit afficher de hautes fréquences spatiales, auquel il est normalement obtenu en utilisant des surfaces rugueuses. On peut identifier certain algorithmes, certain formulent le problème comme la minimization d'une fonction [Ens and Lawrence, 1993, Subbarao and Surya, 1994, Ziou, 1997, Xiong and Shafer, 1993, Marshall et al., 1996, Schechner, 1999]. On peut faire remonter l'idée d'utiliser focus/defocus à Shree K. Nayar [Nayar, 1989, Nayar and Nakagawa, 1994], qui ont développé un opérateur pour mesurer la mise au point afin d'estimer la profondeur. Paolo Favaro et Stefano Soatto ([Favaro and Soatto, 2002]) ont proposé un méthode pour reconstituer des formes 3D à partir d'une collection d'images défocalisé. Les deux méthodes de profondeur focus/defocus utilise la géométrie optique pour reconstituer la scène réelle et sont basées sur le *thin lens law*. Samuel W. Hasinoff et Kiriakos N. Kutulakos [Hasinoff, 2008], présente une nouvelle approche basée sur les méthodes précédentes, confocal stereo, propre pour reconstruire des scènes avec un haut degré de complexité géométrique et des fins détails comme le cheveux. Ils introduisent *confocal constancy property*, pour chaque variation de l'ouverture du diaphragme, l'intensité du pixel d'un point dans la scène varie de façon indépendante. Ceci est prévu dans deux manières: d'une part ils construisent un modèle de lentille détaillée d'un *Single Lens Reflex* (SLR) afin de calculer les distorsions de la lentille. Ce qui permet de mesurer la relation entre l'ouverture de la lentille et le focus pour chaque pixel. De cette façon, ils peuvent évaluer les mesures de focus pour chaque pixel. *Confocal stereo* calcule la profondeur en utilisant une seule caméra, et exploite les capacités de haute résolution de les SLR caméras. Ana Levin et al. [Levin et al., 2007, Zhou and Nayar, , Zhou et al., 2009] introduisent une nouvelle approche pour calculer la profondeur à partir d'une seule image. En insérant dans l'ouverture du diaphragme un motif codé pour occlure les zones de lumière. De cette manière, les rayons propagation de la lentille, dans le capteur de la caméra, ne sont pas uniformément réparties, ce qui peut aider à la mesure

de la profondeur.

Le principe de la vision active est de remplacer une des caméras d'un système stéréoscopique par une source de lumière afin de projeter des motifs visuelles (normalement des colonnes verticales noir et blancs) dans la scène - projection de lumière structurée. Par conséquent, les problèmes soulevés en vision passive peuvent être résolus en se basant sur l'extraction d'indices générés par les motifs projetés et d'établir facilement des correspondances. Ce système consiste en une caméra et une source de lumière qui projette un faisceau laser sur la surface. Et calculer sa profondeur par triangulation, e.g. *Structured Light Scanners* sont des dispositifs qui utilisent la projection d'une image sur la scène (normalement des colonnes verticales noir et blancs). La position d'objet illuminé est déterminé en utilisant la triangulation [Curless and Levoy, 1995].

Motivation

Mon travail se place dans le contexte général de la reconstruction 3D à partir d'images. L'information 3D constitue, en effet, une matière indispensable pour la capture de performance, l'analyse de la forme, par exemple l'archéologie. Je me intéresse ici à des données acquises par des capteurs optiques.

Le choix d'une technique pour la reconstruction de formes par vision dépend essentiellement de la nature des surfaces (réflexion, surfaces lisses ou rugueuses) et de la précision désirée. La vision passive et active se différencient par la résolution de la reconstruction obtenue. La vision passive est connue pour la haute résolution puisque le calcul est effectué pour chaque pixel de l'image. L'inconvénient de la stéréovision est qu'elle requiert la présence d'indices visuelles dans la scène afin d'établir des correspondances. En combinant les deux différentes classes nous permet de combiner les avantages de chacune.

Les informations provenant de cette source n'est pas suffisante pour faire une reconstruction 3D d'objet avec le volume correcte. Cette imprécision est corrigée par

les points de contrôle. La méthode proposée dans ce manuscrit s'applique seulement à des objets fixes. Dans mon étude je considère que l'objet est observé par un capteur optique muni de trois lumières que se déplace autour d'objet.

Problématiques

Les problèmes abordés sont multiples:

1. *la intégration des données;*
2. *établir un système que soit facile mettre en place et utilisable dans une situation réelle;*
3. *le choix des algorithmes d'acquisition;*
4. *la déformation de la surface;*

Ce sont essentiellement ces contraintes qui vont guider la conception du projet.

Contribution et objectifs

L'objectif de ce projet est multiple, tout d'abord, il consiste en la mise en œuvre d'un système de reconstruction 3D à partir d'une séquence d'images de plusieurs angles de vue. Dans ce projet j'ai considéré plusieurs contextes de application: l'analyse de la forme, par exemple en archéologie; capture de performance; l'inspection de qualité de la surface.

Il est important de préciser que le but de ce projet est de proposer un système de reconstruction qui produit des modèles 3D dont l'apparence est la plus fidèle possible au modèle original et de restituer une reconstruction dense de la scène.

Trame du document

Le manuscrit comprend cinq chapitres. Le chapitre 1 repose sur une introduction à description de modèles de caméras. On y décrit la projection réalisée par une caméra lors de l'acquisition d'images. On expose ensuite, les différentes méthodes de

calibration de la caméra. Enfin, on décrit le méthode d'estimation de la structure de la scène à partir du mouvement de la caméra.

Dans le chapitre 2, j'aborde le problème photométrique stéréo. Il s'agit de mesurer l'information radiométrique de l'image pour obtenir la surface 3D.

Le chapitre 3 décrit plus précisément le contexte de l'étude et notamment la configuration de prise de vue étudié qui fait en partie l'originalité du travail. Je présente également le jeux de donnés de synthèse pour la mise en point et l'évaluation des méthodes. On détaille le choix de la représentation de l'information tridimensionnelle de la maillage. Je propose de modéliser les déformations dans la représentation *Laplace* du objet. Je présent la méthode et leurs modèles mathématiques.

Enfin, je consacre le dernière chapitre pour la discussion des résultats, de voies d'amélioration, plusieurs extensions de cette méthode sont proposées pour aboutir à la mise en œuvre d'un système.

Chapter 1

Éléments de base

Dans ce chapitre je introduise les éléments de base nécessaires à la compréhension du mémoire. Je introduit Le modèle et propriétés de la caméra. Cette comporte un aspect géométrique e un aspect radiométrique, je m'intéresse ici au seul premier point. Dans un premier temps, je fais la description de différent types de caméra. Puis, je présente la géométrie epipolar. Enfin, je aborde le problème gouvernant à l'étalonnage et les techniques associés.

Les données

Variables	Description
\Re	the subspace of real numbers
\equiv	identical symbol
\leftarrow	replacer le valeur de la droite à gauche
$X_R = [x_R, y_R, z_R]^T$	coordonnées d'un point exprimées dans le repère R
$x^{(k)}$	le valeur x au instant k
C	repère de la caméra, $C \in \Re^3$
M	repère monde, $M \in \Re^3$
I	repère de l'image, $I \in \Re^2$
p	un point dans la surface, $p \in S$

\mathbb{R}^3 au \mathbb{R}^2 . Le modèle le plus courant est le modèle projectif linéaire ou sténopé. Dans ce modèle, la caméra réalise une projection dans l'espace euclidien sur une surface. La géométrie euclidienne est caractérisée par les transformations euclidiennes: rotations et les translations.

La transformation T est muni d'une matrice de rotation $R \in \mathbb{R}^{3 \times 3}$ déterminée pour trois valeurs, Euler rotation angles, et une vecteur de déplacement t . Par exemple pour convertir un point du repère de la scène pour le de la caméra, on fait:

1.2 Caméra sténopé

Une caméra est un capteur fournissant de nombreuses informations sur le monde extérieur, à l'instant t , une caméra capture une image et enregistre ainsi une grande quantité d'informations photométriques.

Le modèle sténopé ne s'agit que d'une approximation d'un système réel (ceci est composé de lentilles qui sont sources de distorsions).

Dans le modèle sténopé, la caméra est représentée par un plan 2D R appelée le plan image et un point 3D, le centre optique (ou de projection) que l'on notera t , c'est une droite qui passe à travers le centre optique et est perpendiculaire au plan de l'image. La distance entre ce centre optique et le plan image est la focale f de la caméra. La projection d'un point 3D sur le plan image est opérée en dessinant un rayon optique partant du point 3D et allant vers le centre optique de la caméra. L'interception formée par ce rayon optique et le plan image correspond à l'image du point 3D. Pour exprimer mathématiquement ces projections nous introduisons le repère local de la caméra C . Son origine, le centre optique, coïncide avec l'axe z , aussi appelé axe optique. Ce axe rencontre le plan image au point principal c .

On considère un point $X_C = (x, y, z)^T$ défini par rapport au repère de la caméra C . Pour savoir les coordonnées d'un point X_C dans le plan de l'image $X_I = (\hat{u}, \hat{v})$, on utilise le théorème de Thalès [Faugeras and Luong, 2001]:

$$u = -b\frac{x}{z}, \quad v = -b\frac{y}{z} \quad (1.2.1)$$

La transformation entre le repère monde et le repère de la caméra:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (1.2.2)$$

Tous les points provenant d'un rayon optique ont une projection unique sur le plan image I . Il est impossible de déterminer les profondeurs z des points projetés sur une seule image. Cette remarque permet d'introduire les coordonnées homogènes. Si (x, y) sont les coordonnées euclidiennes d'un point sur un plan, ses coordonnées homogènes sont définis par $(\lambda x, \lambda y, \lambda)$.

1.3 Projective Geometry

Avant de définir la caméra projective, je présente un cadre théorique, la géométrie projective. On représente un point $x = (x, y, z)^T$ dans le espace projective comme en utilisant ces coordonnées homogènes. La caméra projective est la représentation plus générique du modèle sténopé. Un problème central c'est faire l'inverse allant d'espace 2D pour 3D. Remarquons qu'un point m dans le plan de l'image contient un rayon de lumière. Par définition le rayon de lumière contient le centre optique. Pourtant, pour trouver les respectives coordonnées $(x, y, z)^T$, on utilise la direction de ce rayon. En utilisant les coordonnées projectives l'équation (1.2.1) devient:

$$m = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ T \end{bmatrix} = P_0 M \quad (1.3.1)$$

Ainsi $u = x/z, v = y/z, x = X/T, y = Y/T, z = Z/T$. Les points avec $T = 0$ ont pas de correspondance affines, sont points défnit au infinie. En utilisant la géométrie projective nous amène une mieux description du problème. On appelle matrice de projection P associée à une caméra, la matrice de $\mathfrak{R}^{3 \times 4}$ qui représente la projection des points de \mathfrak{R}^3 sur le plan image. En connaissant la matrice de projection, on peut à partir des coordonnées d'un pixel localiser le rayon optique.

La matrice P_0 c'est un cas simple, en général l'espace de coordonnées de la caméra et de la scène ne sont pas alignés. La forme général pour la matrice de projection a la forme:

$$P = \underbrace{\begin{bmatrix} \alpha_u & \gamma & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_A P_0 \underbrace{\begin{bmatrix} R & t \\ 0_3^T & 1 \end{bmatrix}}_D = A [Rt] \quad (1.3.2)$$

1. A décrit les paramètres intrinsèques de la caméra. Ils regroupent les 5 entrées de la matrice: α_u et α_v représentent la distance focal, c'est à dire la distance entre le plan image et le centre optique. Cette distance est donnée en unités de pixel en chaque direction. (u_0, v_0) dénote le point principal. Le paramètre γ représente de distorsions que décrivent les défauts optiques, si le pixel est carrée $\gamma = 0$.

2. D décrit l'orientation de position de la caméra par rapport au repère du monde et sont appelés les paramètres extrinsèques.

1.4 Étalonage de caméra

L'étalonnage de caméra consiste à déterminer de manière précise les paramètres internes de la caméra: f_u , f_v et du point principal c . Nous supposons que les pixels sont carrés donc le facteur *skew* est nul. À ces quatre paramètres s'ajoutent les 5 paramètres de distorsion radiale. La pose de la caméra comprenant l'orientation et la position sont estimés simultanément. Dans le cas des optiques pour caméras disponibles sur le marché ne suivent pas le modèle sténopé. Par conséquent comportent des distorsions non linéaires qui peuvent conduire à accumulation d'erreurs dans la plupart des algorithmes de vision. Une manière classique de corriger ces distorsions passe par l'utilisation d'une mire de calibration [Brown, 1971]. La méthode consiste à détecter les segments de droite distordus présents dans l'image par une détection de contours. La majorité des approches se basent sur la minimisation de la courbure des lignes de l'image qui devraient être droites [Mohr and Triggs, 1996].

Pour réaliser cette opération il existe différentes méthodes, dans [Zhang, 2000] on trouve un aperçu complet sur les procès.

Dans la calibration classique on prend plusieurs images d'une mire de calibrage et de chercher les paramètres de la caméra qui minimise l'erreur de reprojection [Clarke and Fryer,]. Ces méthodes nécessitent, donc d'avoir une caméra et une mire. Le principe base est de photographier une mire dont on connaît les formes et la position dans l'espace, et de détecter ces formes sur les images afin de calculer les paramètres intrinsèques. On peut considérer des méthodes non-linéaires, en général plus robustes. Les données utilisées en général sont des points, lignes ou des ellipses.

La calibration automatique est surtout appliquée dans la robotique. Les paramètres intrinsèques peuvent être faites une fois, mais les paramètres extrinsèques doivent être calculés à chaque mouvement, afin de connaître la position du appareil [Faugeras et al., 1992]. Ici il n'y a pas besoin de mires. En effet, ils utilisent des correspondances entre points entre deux ou plusieurs images au cours d'un placement ou changement de orientation

pour déterminer les paramètres l'aide de géométrie épipolaire.

Dans les travaux présentes dans ce travail on considère que les caméras sont calibrées en distorsion radiale - les images sont rectifiées.

1.5 Structure à partir du mouvement

Dans cette section je présente le thème structure à partir du mouvement (*Structure-from-Motion*, SfM) que vise à l'estimation de la structure 3D d'une scène à partir d'une séquence d'images. La problématique de l'estimation de la structure tridimensionnelle d'une scène à partir d'images est donc:

1. *Identifier points sur la séquence d'images [Lowe, 2004]*
2. *estimer la matrice fondamentale entre les images*
3. *récupérer les poses des caméras*

La première étape est de trouver des correspondances des points entre images. Un algorithme de détection automatique de coins, permet d'établir plus facilement ces correspondances. De là, les matrices de projection perspective des images peuvent être calculées à une transformation projective grâce aux propriétés de la géométrie épipolaire. SfM fait l'analyse du mouvement à partir d'une séquence d'images, ensuite est créée une carte de mouvement.

La connaissance du mouvement d'une caméra a de nombreuses applications. Dans le projet *Photo Tourism* [Snavely et al., 2006], [Snavely et al., 2008] la reconstruction s'effectue à partir d'images non calibrées issues de prises de vues différents et non continus. Il s'agit ici d'extraire le mouvement, dans le cadre d'une calibration extrinsèque partielle et imprécise.

1.6 Résumé

Ce chapitre fait la présentation du contexte de la géométrie de la caméra et à mis en évidence les données nécessaires à la compréhension des chapitres prochains.

Les paramètres extrinsèques sont composés par une transformation de rotation et de déplacement. Je désigne la pose de caméra l'orientation et position de la caméra.

Le processus de formation de l'image est défini pour les paramètres intrinsèques de la caméra soit spécifique de la lentille ou du sensor de la caméra. Pour les systèmes avec une lentille nous avons besoin de compenser les distorsions de la lentille. Ici les coordonnées de l'image sont définies dans le repère de l'image. Pour traduire la projection d'un point dans ce repère image on utilise les paramètres suivants: la focale, le point principal et le factor *skew*. L'ensemble de ces factors compose la matrice de calibrage de la caméra. Les paramètres varient suivant les modèles de la caméra et de la lentille et peuvent évoluer au cours du temps.

Le procès étalonnage de la caméra consiste de manière précise déterminer les paramètres intrinsèques de la caméra et la distorsion radiale. Le procédé plus simple est de prendre plusieurs images images d'une mire de calibrage — on minimise la distance entre les observations et les positions estimées. Il est possible d'auto-étalonner une caméra sans mire.

Les algorithmes *structure-from-motion* sont capables d'estimer les déplacements d'une caméra. Le procédé est le suivant: on détecte et suit des points dans la séquence d'images; ensuite, on estime la collection de poses et la trajectoire de la caméra; la trajectoire de la caméra est raffiné par une méthode optimale. On minimise les distances entre les éléments de la scène réelle et les éléments obtenus. Cette dernière étape apporte la précision à la reconstruction de la scène - bundle adjustment.

Chapter 2

L'approche photométrique

Ce chapitre présente l'approche photométrique, celui-ci explore l'information radiométrique du pixel dans l'image. L'information photométrique donne seulement l'information sur les gradients de la surface d'objet. Nous avons besoin d'avoir l'information sur la direction de vue de la caméra, de la lumière et les propriétés de réflexion d'objet. Cette méthode donne la profondeur de la surface à condition que l'information de la éclairage e des propriétés de réflexion soient disponibles. Premier on calcule la normal (ou le gradient) de la surface. Après les normales sont utilisées pour intégrer la surface 3D. Dans le cas d'une surface *Lambert*, le gradient peut être calculé en utilisant trois images.

2.1 État de l'art

On trouve beaucoup de travaux sur des systèmes de reconstruction à partir d'images, il serait bien de les présenter tous. Je vais donc passer en revue les méthodes les plus représentatives. Des ombres à partir de plusieurs images sous différents conditions de lumière contiennent information sur la forme d'objet [[Wöhler and Hafezi, 2005](#), [Schlüns, 1997](#)]. Hafezi and Wöhler présentent une méthode pour acquérir la forme à partir de l'ombre [[Wöhler and Hafezi, 2005](#)], Yu et Chang calculent l'ombre à partir de plusieurs images [[Schlüns, 1997](#)]. Les approches photométriques sont basées sur

l'intensité du pixel, cette méthode permet d'estimer la surface des gradients. La technique *shape-from-shading* utilise seulement une image [Horn, 1990, Klette and Schlüns, 1996], ce que peut amener a plusieurs solutions. Pour surmonter ce problème on utilise plusieurs images – photométric stéréo [Lohse and Heipke, 2004]. Pour la reflectance non uniforme on peut nommer les travaux de [Wöhler and Hafezi, 2005, R. Lena and Fattinanzi,]. On peut encore ajouter à ces méthodes les propriétés de polarisation de la surface, *shape from polarisation* [S. K. Nayar and Boulton, 1993, Georgiades, 2003].

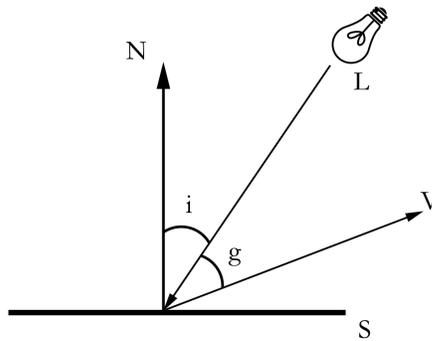


Figure 2.1: Dans cette image je représente les trois angles sur la surface S . V , L et N sont la direction de la caméra, lumière (le rayon) et la normal de la surface. Le angle incident i c'est le angle entre le rayon et la surface; le angle émergent c'est la somme des angles i et g .

2.2 Définitions radiométriques

La partie de lumière reflété pour la surface d'objet dans une direction dépend: de les propriétés optiques du matériel de la surface, de la structure de la surface, la distribution spectrale et la polarisation de la lumière incident. On dénote les propriétés de reflet par la fonction $\phi(i, e, g)$ avec trois angles, i , e , et g appelé incident, émergent et

phase. Les angles i et e sont définis par rapport à la normal de la surface. $\phi(i, e, g)$ détermine le ration de la radiance de la surface et irradiance mesuré par unité d'aire.

On considère la surface suivante:

$$\phi_a(i, e, g) = \rho \cos(i) \quad (2.2.1)$$

La fonction de reflectance correspond à une surface parfaitement diffuse - la quantité de lumière et fixe quelque soit la direction de vue. ρ c'est le factor de reflectance et cosine du angle incident. La reflectance correspondent est donné par:

$$R_a(p, q) = \frac{\rho(1 + pp_s + qq_s)}{\sqrt{1 + p^2 + q^2} \sqrt{1 + p_s^2 + q_s^2}} \quad (2.2.2)$$

En utilisant la reflectance, l'équation élémentaire de la formation de l'image est donné par:

$$I(x, y) = R(p, q) \quad (2.2.3)$$

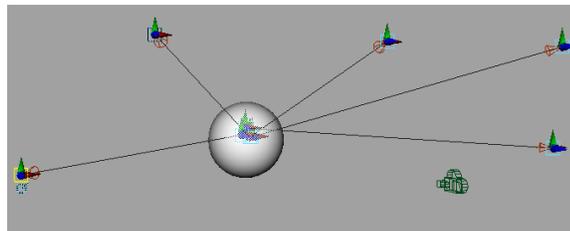


Figure 2.2: La disposition des lumières dans la scène.

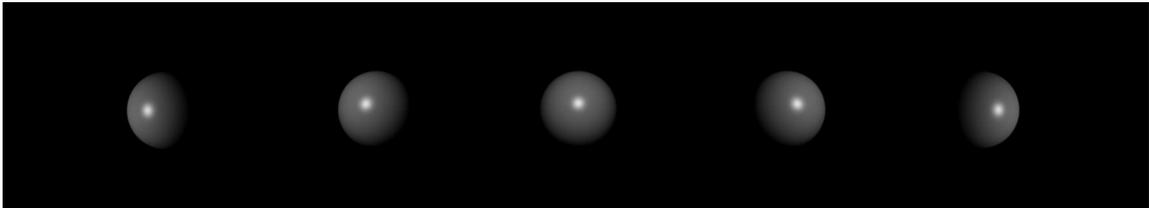


Figure 2.3: Le résultat de l'utilisation de la distribution des lumières de la Fig.2.2.

2.2.1 La forme à partir de la photométrie

L'idée de photométrie stéréo (*photometric-stereo*, PhS) c'est de varier la direction de la lumière \vec{s} , pendant la direction de la caméra (la position de vue) \vec{v} est fixe [Woodham, 1980]. Ce qui permet d'obtenir l'information nécessaire pour calculer l'orientation de la surface pour chaque point (u, v) de l'image I . On utilise les valeurs radiométriques enregistrés pour chaque direction de la lumière pour estimer la forme d'objet.

Dans notre scène on considère la surface $z(x, y)$ à construire est éclairée pour une lumière e vu pour une caméra, placées dans la direction \vec{s} et \vec{v} . Pour chaque pixel (u, v) de l'image on veut estimer la profondeur $z(x, y)$. En supposant que la source de lumière directionnelle L est connue, la loi de *Lambert* donne des solutions pour calculer la normale N sur la surface en chaque point en connaissant la réflectance de la surface.

2.2.2 La normal à partir de l'image

Des contraintes supplémentaires sont nécessaires pour rendre le problème bien posé. Les difficultés peuvent être surmontées en acquérant plusieurs images avec différents éclairages. Dans ce cas, chaque image fournit une contrainte sur le calcul de la normale. Par conséquent, trois images sont suffisantes pour obtenir une solution unique pour l'estimation de la normale à la surface en chaque point.



Figure 2.4: Les gradients x et y calculés.

La photométrie c'est une technique qui utilise au moins trois images pour calculer l'équation (2.2.3). On considère trois images $I_1(x, y)$, $I_2(x, y)$ et $I_3(x, y)$, obtenues en changeant la direction de la lumière, une fois que la géométrie est la même chaque position (x, y) correspond au même objet, pour cette raison au même gradient (p, q) . La variation de la direction de la lumière va varier la reflectance map $R(p, q)$.

Ces images permet d'identifier uniquement la orientation de la surface en chaque position (x, y) . Si on considère $I = [I_1, I_2, I_3]^T$, en plus:

$$\begin{aligned} n_1 &= [n_{11}, n_{12}, n_{13}] \\ n_2 &= [n_{21}, n_{22}, n_{23}] \\ n_3 &= [n_{31}, n_{32}, n_{33}] \end{aligned} \tag{2.2.4}$$

qui définis trois directions de lumière. Ainsi on peu construire la matrice:

$$N = \begin{pmatrix} n_{11} & n_{12} & n_{13} \\ n_{21} & n_{22} & n_{23} \\ n_{31} & n_{32} & n_{33} \end{pmatrix} \tag{2.2.5}$$

On considère maintenant la normale de la surface au point (x, y) définit par $\eta = [n_1, n_2, n_3]^T$. Donc $I = \rho N \eta$, si l'inverse de N existe on peut écrire $\rho \eta = N^{-1} I$. Avec

trois ou plus lumières incidentes, on peut calculer $g = \rho\eta$ en utilisant un méthode de moindres carrés [Hayakawa, 1994], de cette façon on peu écrire la fonction:

$$Q = \sum_i (I_i - L_i g^T) \quad (2.2.6)$$

La fonction objective Q est minimisé selon g . Ce que permet de calculer les normales η de chaque pixel. Ensuite on peut estimer le factor reflectance ρ avec la méthode de moindres carrés. Enfin, à travers d'une méthode de optimization on calcule la surface qui correspond mieux aux normales.

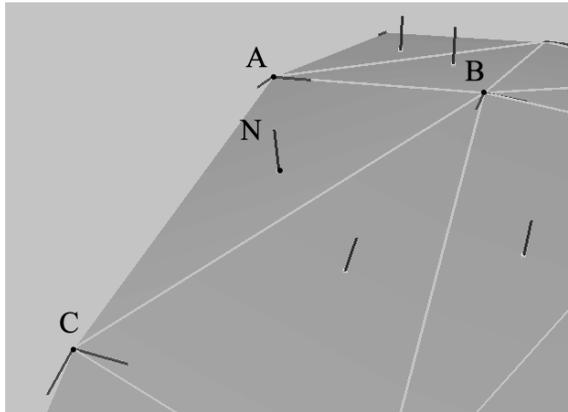


Figure 2.5: Une surface triangulaire élémentaire définit dans le espace \mathfrak{R}^3 . On peut écrire la relation suivante, liant les coordonnées cartésiennes A , B et C , $\vec{N} = \vec{AB} \otimes \vec{AC}$

2.2.3 Profondeur à partir du gradient

Le problème traité dans cette partie est le suivant: comment estimer la surface à partir du gradient de la surface. Cela revient donc définir la fonction:

$$z(x, y) = f(u, v), \quad (2.2.7)$$

où x , y et z correspondent aux coordonnées cartésiennes d'un point de la surface. Ce problème a été traité notamment dans le cadre du *shape-from-shading*, où le problème est résolu en utilisant les cartes des dérivées partielles de la surface $p(x, y) = \frac{\partial z(x, y)}{\partial x}$ et $q(x, y) = \frac{\partial z(x, y)}{\partial y}$

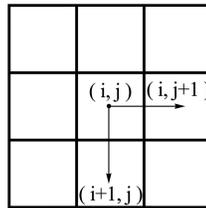


Figure 2.6: La représentation du déplacement dans la matrice.

Système d'équations linéaires en utilisant la relation de la normal \vec{N} décrit en 2.5.

$$\begin{aligned}\vec{a} &= (i + 1, j, f(i + 1, j)) - (i, j, f(i, j)) = (1, 0, f(i + 1, j) - f(i, j)) \\ \vec{b} &= (i, j + 1, f(i, j + 1)) - (i, j, f(i, j)) = (0, 1, f(i, j + 1) - f(i, j))\end{aligned}\quad (2.2.8)$$

On peut generalizer les équations 2.2.8 pour la forme de la matrice en utilisant les indices définis en 2.2.9.

$$\begin{cases} A[2 \times index, index + 1] = z \\ A[2 \times index, index] = -z \\ A[2 \times index + 1, index + width] = z \\ A[2 \times index + 1, index] = -z \end{cases}\quad (2.2.9)$$

$$\begin{cases} b[2 \times index] = -(n_x) \\ b[2 \times index + 1] = -(n_y) \end{cases}\quad (2.2.10)$$

L'a matrice A et le vecteur b démontre un exemple de remplissage en utilisant la méthode précédent. On voit que chaque paire de lignes de la matrice A correspond:

$$\underbrace{\begin{bmatrix} -z & z & 0 & 0 & 0 & 0 \\ -z & 0 & 0 & z & 0 & 0 \\ 0 & -z & z & 0 & 0 & 0 \\ 0 & -z & 0 & 0 & z & 0 \\ 0 & 0 & -z & z & 0 & 0 \\ 0 & 0 & -z & 0 & 0 & z \end{bmatrix}}_A, \underbrace{\begin{bmatrix} x \\ y \\ x \\ y \\ x \\ y \end{bmatrix}}_b$$

Ils existent de nombreuses méthodes permettant d'obtenir la surface à partir des cartes de gradients réparties en trois catégories [Frankot et al., 1988, Kovesei, 2005]: locales, itératives et globales. Le choix de la méthode devra se faire suivant le type de surface à construire, le temps de calcul disponible et le bruit présent dans les images de gradients.

Chapter 3

Fusion de données

On a vu que le problème de la reconstruction 3D peut être abordé pour méthodes photométriques et géométriques. Dans les premiers, les imprécisions de cette estimation 3D est due principalement aux valeurs radiométriques: le reflet de lumière sur la surface. Ces problèmes inhérents aux conditions de prise de vues expliquent le choix d’une deuxième méthode pour faire de corrections. Ils justifient également l’exploitation de la totalité de la séquence d’images plutôt que une image. Les méthodes géométriques utilisent la corrélation de points entre des images, l’inconvénient c’est que ces méthodes dépend de la texture de la scène.

Photométrie stéréo est capable de faire reconstructions 3D de haute qualité, mais est sujette à des erreurs dues à l’éclairage ou d’étalonnage non idéales. De cette façon, je propose l’usage des points de contrôle pour faire des corrections sur les estimations photométriques. Ces points — obtenues à partir du méthode SfM — améliorent la qualité de la surface obtenue de la méthode photométrique. Ce chapitre introduit une méthode permettant d’intégrer des données du champ de gradients issues de photométrie stéréo avec les points de contrôles. Je présente alors la représentation *Laplace* d’objet. Cette représentation permet d’appliquer des déformations non-linéaires ce qui permet de préserver le détail de la surface. Ensuite, je détaille l’emploi de cette technique dans les expériences.

3.1 État de l'art

Ju Yong Chang et al [Chang et al., 2007] proposent une méthode pour combiner plusieurs champs de normal de différents angles de vue en utilisant une fonction d'énergie qui est minimisé. Ils utilisent la méthode *level-set* pour obtenir l'objet final.

Zhenglong Zhou et Ping Tan [Zhou and Tan, 2010] utilisent systèmes de capture munis de lumières attachées à une caméra. Li Zhang et al. [Zhang et al., 2003] proposent une méthode pour calculer le mouvement, la forme et couleur à partir d'une séquence d'images. Son système utilise SfM allié avec stéréo-vision et photométrie pour estimer la surface d'objets rigides. Ils considèrent: des matériaux du type *Lambert*; que l'objet est non modifiable (rigide); projection orthographique; il n'y a pas de occlusions, d'ombres, surfaces specular, et d'inter-reflections. En premier ils font l'estimation du mouvement de la caméra, pour cela ils font le *tracking* des points identifiables en utilisant l'équation *Lukas-Kanade*, et après ils appliquent la factorisation *Tomasi-Kanade* pour récupérer la position de la caméra. Ensuite, pour chaque pixel c'est calculé la normal qui est après intégrée pour obtenir la composante z de la surface. À cause d'imprécisions dans le calcul de la direction de la lumière, on utilise les valeurs obtenues de de stéréo-vision, pour corriger la surface 3D.

Pablo d'Angelo et Christian Whöler [d'Angelo and W,] combinent les propriétés de la reflection, polarisation, et les donnés de profondeur. Pour estimer le champ des gradients ils propose un système pour minimiser d'un coté la relation entre reflection e polarisation, et d'autre coté le champ de la profondeur obtenu de la surface de gradients et les donnés de profondeur. Dans son système l'information de la profondeur est acquis à partir d'algorithme de vision-stéréo. Itzik Horovitz et Nahum Kiryati [Horovitz and Kiryati, 2004] utilisent des points de contrôle comme contraintes pour estimer la profondeur. Les points de contrôle éparses sont obtenues avec la projection d'une matrice de points sur la surface des objets. Ça permet de créer un groupe des points aisément identifiable dans la scène. Après, à chaque point est associé un poids

pour déformer la surface selon les poids associés. Ils utilisent le full-multigrid avec successive over-relaxation.

Carlos Hernández et al. [Hernández et al., 2007] propose une méthode pour la capture d’objets non rigides en utilisant de méthodes photométriques pour récupérer la surface d’objet dans une séquence d’images. Son système est particulièrement adapté pour la capture de tissus. La principale différence des méthodes précédents, c’est que son système permet de capturer la déformation d’objet. L’idée est d’utiliser la première surface calculée (dans la séquence d’images), et appliquer des déformations successives sur la surface en appliquant un vecteur de déformation \vec{t} à chaque vertex en utilisant une fonction d’énergie pour régulariser les déformations (éviter de grosses déformations de la maillage). Les vecteurs de déplacement sont basés sur le *calcul optical-flow* entre les images. Dans ce cadre, on peut encore nommer le travail [Zhang et al., 2004] qui permettent de capturer des objets non rigides.

C. Wöhler propose un système pour intégrer méthodes qui utilisent techniques photométriques et le flou de l’image [Fassold et al., 2004, Horovitz and Kiryati, 2004, Wöhler,] pour obtenir des résultats précis et nettes. À partir du flou de l’image on calcule la profondeur et les nuages de points.

Yasutaka FuruKawa et Jean Ponce [Furukawa and Ponce, 2010] proposent une algorithme basée sur *patches* pour stéréo vision de plusieurs angles de vue. L’algorithme commence par détecter la corrélation de points sur une séquence d’images qui permettra d’obtenir une première collection de *patches*. Ensuite, ils appliquent une procédure pour augmenter la quantité initiale et un filtre pour effacer le bruit. À la fin, l’algorithme rend une collection dense de *patches* qui couvrent la surface d’objet. Cette collection est ensuite convertie dans une *mesh*. Son méthode utilise aussi les gradients de la surface obtenus à partir des méthodes photométriques.

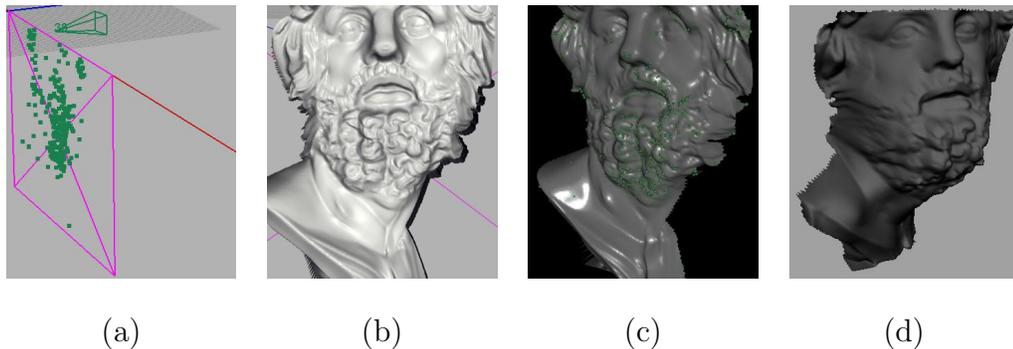


Figure 3.1: (a) démontre les nuages des points vues à partir de la caméra dans une plan 2D. L'image (b) c'est le résultat final de la photométrie stéréo vu à partir de la caméra en (a). On projette, ensuite, la nuage des points sur la surface d'objet (c). En utilisant la représentation *Laplace* d'objet chaque point va applique des déformations sur la surface, (d) démontre l'objet obtenu.

3.2 Le système

Nous définissons le système de capture comme une caméra muni de trois lumières de couleurs différents [Zhou and Tan, 2010, Paterson and Fitzgibbon, 2002, Hernández et al., 2007, Paterson et al.,]. Ou la méthode SfM mesure les déplacements de la caméra dans l'environnement. Notons que le vecteur $x^{(i)}$ constitué des informations de localisation, position $t \in \mathbb{R}^3$ et d'orientation R du système. Puisque le système est mobile, sa localisation évolue au cours du temps, indicé par la variable t .

L'approche consiste à localiser un système en reconstruisant simultanément l'environnement obtenue à partir du SfM. On considère que le repère du système est le centre optique de la caméra, et l'on cherchera localiser celui-ci. Les paramètres de notre modèle se composent d'un historique de la localisation du système $x_{t \in [1, \dots, N]}$. Nous notons X l'espace des vecteurs de paramètres $x \in X$.

Je donne un aperçu du système dans la figure (en bas). Le système mobile possède son propre repère noté W et évolue dans l'environnement, symbolisé par le repère

monde W_s . Le système possède trois lumières, de couleur rouge, vert et bleu, que lui sont rigidement liés que délivrent des mesures dans leur propre repère W (montrer une figure avec caméra e lumière).

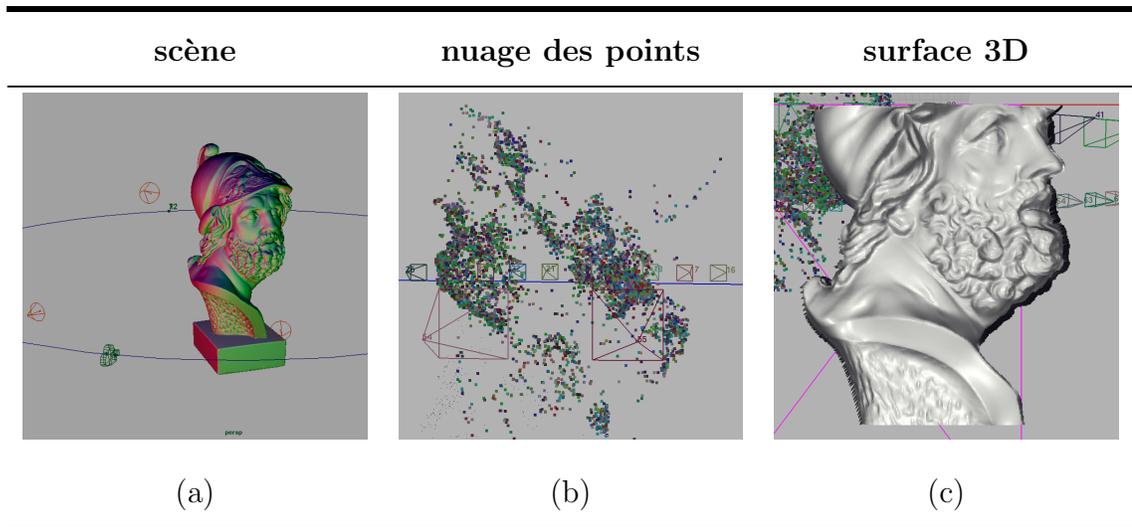


Figure 3.2: Pour les résultats expérimentaux j'ai considéré la scène (a). L'image (b) c'est le résultat du algorithme structure à partir du mouvement.

3.3 Aperçu du Processus

On prend la suite des images $S = \{s_i\}, i = 1, \dots, N$ autour d'objet.

Je résume les étapes de la méthode dans l'algorithme. Nous proposons un exemple d'utilisation de cette méthode dans la section suivante.

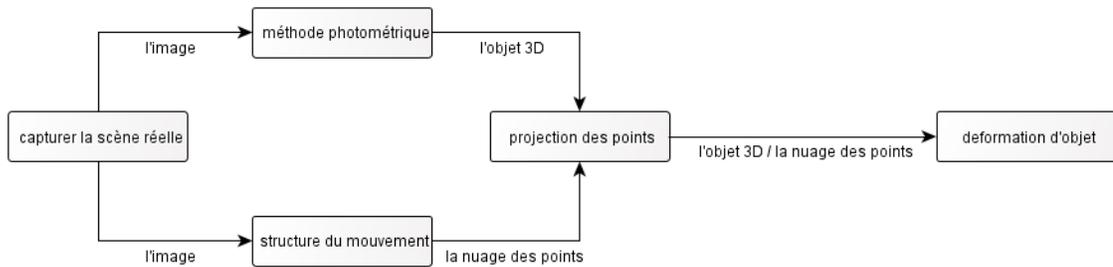


Figure 3.3: Le système.

Input: la suite des images $S = \{s_i\}, i = 1, \dots, N$

Output: la surface 3D

$C \leftarrow \text{StructurefromMotion}(S);$

foreach s_i **do**

$\zeta_i \leftarrow \text{computeHeightGrad}(s_i);$
 $\tau_i \leftarrow \text{placeControlPoints}(c_i, \zeta_i);$
 $\lambda_i \leftarrow \text{laplaceRepresentation}(\zeta_i);$
 $\psi_i \leftarrow \text{laplaceDeform}(\lambda_i, \tau_i);$

end

$F \leftarrow \text{mergeSurfaces}(\Psi);$

Algorithm 1: L'algorithme pour estimer la surface 3D. Cette algorithme est cumulatif, calcul de la déformation d'une séquence de surfaces pris en instants différents. C contient l'information de la pose (orientation et position dans le repère monde) de chaque caméra. En plus est possible de savoir quel sont les points visibles par caméra.

La première étape de la reconstruction à partir de plusieurs images est de construire un modèle pour chaque image. Les points visibles de chaque caméra sont transformés et ajoutés au modèle comme le montre la figure. Une fois que nous avons une reconstruction pour chaque image, nous pouvons les fusionner afin d'obtenir le modèle global.

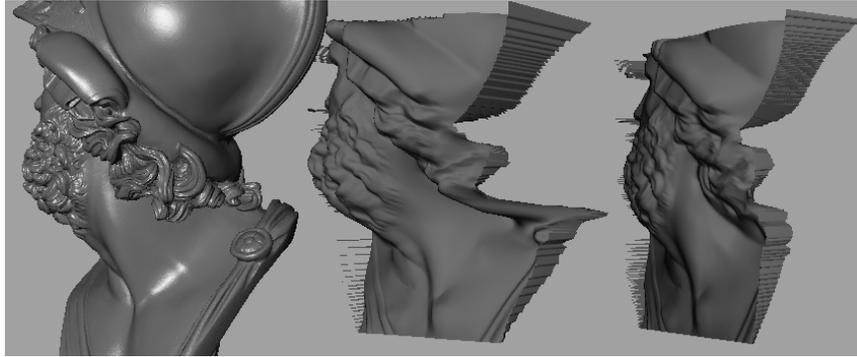


Figure 3.4: À gauche c'est l'objet original, au milieu la surface avec des déformations issues des points de contrôle, et à droite la surface sans déformations.

3.3.1 Déformation de la maillage

Dans le contexte de ce travail je m'intéresse aux déformations non linéaires qui permettent de conserver le détail de la surface de l'objet et d'avoir des déformations lisses. Dans la littérature on peut trouver de nombreuses travaux [Sorkine and Alexa, 2007, Sorkine, 2006, Lipman et al., 2004, Sorkine et al., , Sorkine et al., 2004, Zhou et al., 2005a, Zhou et al., 2005b, Nealen et al., 2005a, Nealen et al., 2005b] qui utilisent la représentation *Laplace* pour modifier l'objet 3D. Dans ces projets, à chaque vertex est associé un poids ce qui permet de réguler les déformations de la surface. En plus, utilisent une fonction fonctionnel [Terzopoulos et al., 1987] qui simule courbature et extension en transformant la surface $S \rightarrow S'$, ce qui permet d'avoir des déformations lisses.

Représentation des triangles:

$$\begin{aligned}
 V &= \{v_1, \dots, v_n\} \\
 E &= \{e_1, \dots, e_h\}, e_i \in V \times V \\
 F &= \{f_1, \dots, f_m\}, f_i \in V \times V \times V \\
 P &= \{p_1, \dots, p_n\}, p_i \in \mathbb{R}^3
 \end{aligned}
 \tag{3.3.1}$$

Ce type de déformations travaillent avec la représentation des coordonnées différentielles, ce type de coordonnées utilisent *differential calculus* pour définir des propriétés de la surface et courbes [Botsch et al., 2007, Hormann et al., 2007, Yang and Ivriissimtzi, 2010, Desbrun et al.,]. Dans une surface continue $S \subset \mathfrak{R}^3$, est donnée pour les équations paramétriques suivantes [Botsch et al.,]

$$X(u, v) = \begin{pmatrix} x(u, v) \\ y(u, v) \\ z(u, v) \end{pmatrix}, (u, v) \in \mathfrak{R}^2 \quad (3.3.2)$$

ou les équations x , y et z sont fonctions différentielles. Ainsi:

$$\begin{aligned} X_u(u_0, v_0) &= \frac{\partial X}{\partial u}(u_0, v_0) \\ X_v(u_0, v_0) &= \frac{\partial X}{\partial v}(u_0, v_0) \end{aligned} \quad (3.3.3)$$

La normal de la surface est orthogonal par rapport aux deux vecteurs, de cette façon on peut calculer la normal de la forme suivante:

$$n = \frac{X_u \times X_v}{\|X_u \times X_v\|} \quad (3.3.4)$$

Definition 3.3.1. L'opérateur Laplace Δ est défini comme la divergence du gradient, $\Delta = \nabla^2 = \nabla \cdot \nabla$. Dans le cas d'une fonction de deux variables $f(u, v)$, on peu écrire l'opérateur différentiel comme la somme de la deuxième dérivatif partiel: $\Delta f(u, v) =$

$$\operatorname{div} \nabla f = \operatorname{div} \begin{pmatrix} f_u \\ f_v \end{pmatrix} = f_{uu} + f_{vv}.$$

L'opérateur *Laplace-Beltrami* prolonge cette définition pour les fonctions définies sur les surfaces. L'objectif est de calculer l'opérateur *Laplace-Beltrami* pour chaque vertex dans une intervalle $N_n(v)$, *n-ring*.

Definition 3.3.2. Pour calculer le *Laplace-Beltrami*:

$$\Delta f(v_i) = \frac{1}{2A_i} \sum_{v_j \in N_1(v_i)} (\cot \alpha_{i,j} + \cot \beta_{i,j}) (f_j - f_i) \quad (3.3.5)$$

Remark 3.3.1. Un point p_i dans la surface S est un point de contrôle si $p_i \in R = S \setminus (H \cup F)$, c'est à dire c'est un point où nous pouvons appliquer des transformations déplacement per vertex du type $p_i \rightarrow p_i + d(p_i)$. Pour les parties fixes de la surface S (où il n'y a pas des points de contrôle) $d(p_i) = 0$, c'est à dire il n'existe aucun déplacement. Pour simplifier la notation on peut écrire:

$$\begin{aligned} d(p_i) &= \bar{d}_i, \forall p_i \in H \\ d(p_i) &= 0, \forall p_i \in F \end{aligned}$$

À ce stade de l'étude, j'ai défini une méthode pour déformer la forme 3D. La technique retenue repose sur la voisinage spatial. Maintenant, l'objectif est d'intégrrer les points issues de SfM dans la surface du modèle. Ces points sont ensuite utilisés en association de la représentation *Laplace* pour restituer la profondeur original à la surface dans lequel il opère. De cette façon, il faut d'abord convertir la représentation du modèle pour *Laplace*. Ensuite on applique les transformations issues de SfM à chaque point de la surface. Enfin, on retourne pour la représentation original du modèle.

Definition 3.3.3. Un point de contrôle est définit comme $p_i \in H \subset S$. Chaque point enregistre la position vectoriel \mathfrak{R}^3 délivré par SfM. Ce vecteur permet de faire transformations sur chaque vertex de la surface: $p \rightarrow p + d(p)$.

$$-k_s \Delta d + k_b \Delta^2 d = 0, \quad (3.3.6)$$

où k_s et k_b sont utilisées pour contrôler la résistance à extension et courbature de la surface.

Definition 3.3.4. Pour déformer la surface on impose les contraintes suivantes pour chaque point p_i :

$$\begin{aligned} \Delta s(p_i) &= 0, p_i \in R \\ s(p_i) &= 1, p_i \in H \\ s(p_i) &= 0, p_i \in F \end{aligned}$$

Dans la définition 3.3.4, $s(p_i)$ dénote la représentation *Laplace* du vertex. En utilisant le remarque 3.3.1 on peut calculer la fonction 3.3.6 dans le domaine discrète par vertex:

$$\begin{aligned} -k_s \Delta d_i + k_b \Delta^2 d_i &= 0, p_i \in R \\ d_i &= \bar{d}_i, p_i \in H \\ d_i &= 0, p_i \in F \end{aligned} \quad (3.3.7)$$

3.4 Résultats expérimentaux

Dans cette section je présente les résultats expérimentaux dans une scène virtuelle, en utilisant donnés synthétiques, voir Fig. 3.3.1.

3.4.1 Contexte de l'évaluation

Dans cette section je présente les résultats expérimentaux dans une scène virtuelle. Dans les travaux présentées, pour capturer l'objet j'ai utilisé une caméra muni de trois

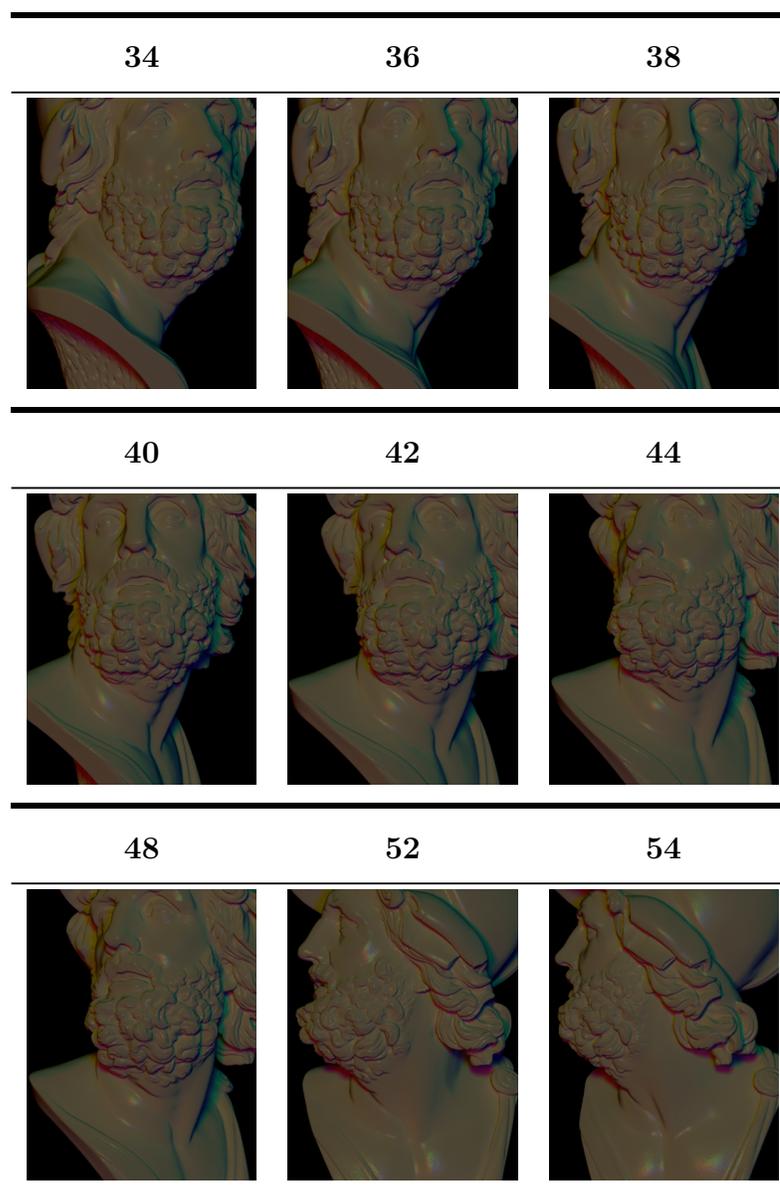


Figure 3.5: Pour les résultats j'ai sélectionné dans la séquence d'images l'intervalle de l'image 34 à 54, cette séquence utilise une écart-type 2. Les images générées ont une taille de 1024×768 pixels et ont été pris avec le système de capture muni de trois lumières (voir l'image Fig. 3.2 (a)).

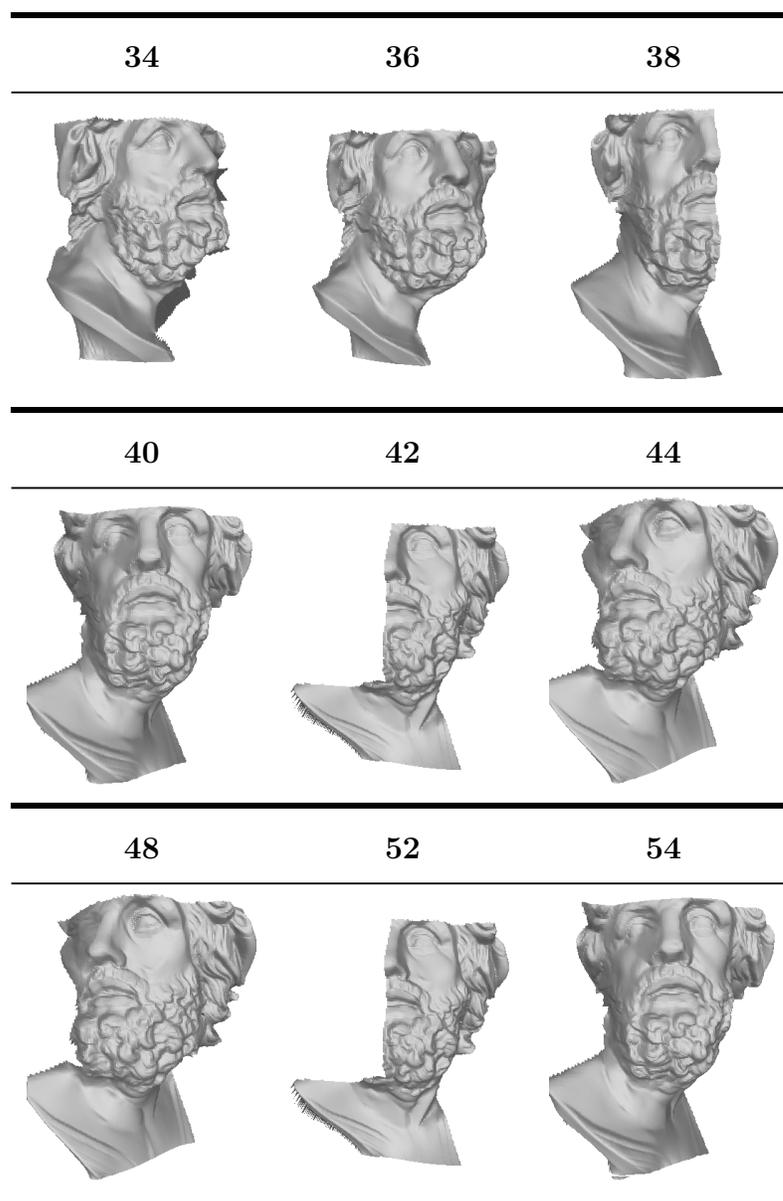


Figure 3.6: Le résultat de la surface calculé après de la déformation.

lumières de couleurs différents: rouge, vert et bleu. La position de chaque lumière par rapport à caméra est connue et la distorsion radial est négligeable

Les algorithmes photométriques ont été implémenté dans la langage C++, en

utilisant la librairie *Trimesh* ¹ pour enregistrer e visualizer le model 3D et la librairie Matrix Template Library ² les équations linéaires. J'ai utilisé le système *Bundler* de Noah Snavely ³ pour calculer la nuage des points à partir de la séquence d'images.

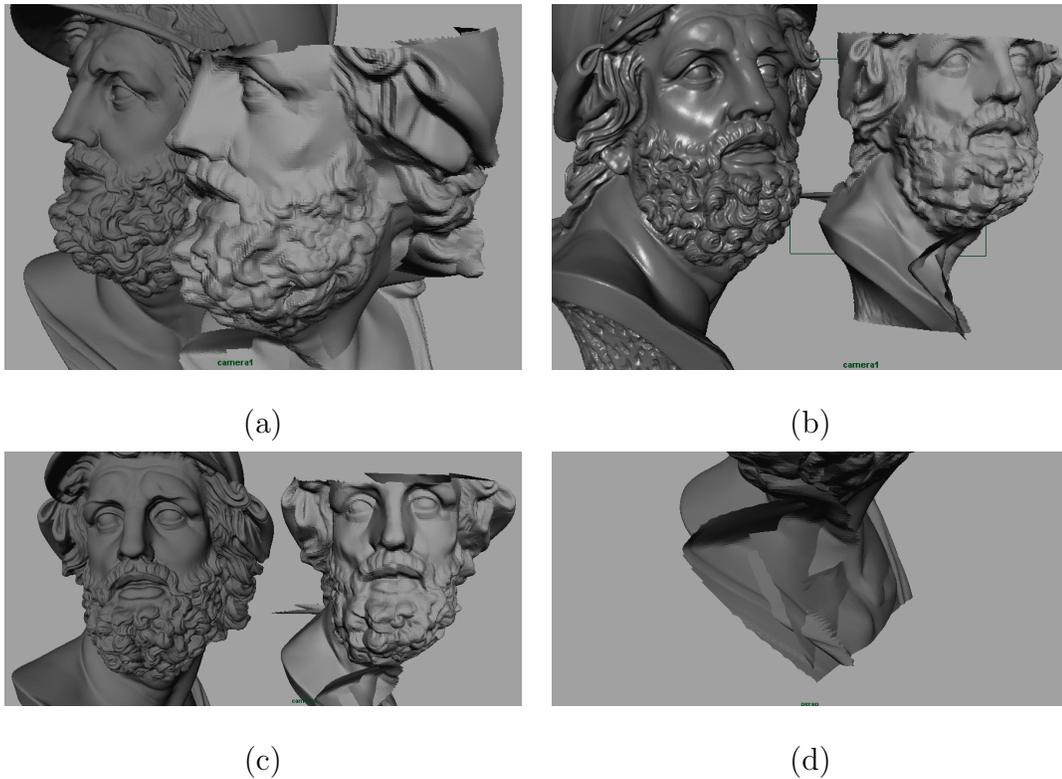


Figure 3.7: Ici je présent quelques images de reconstructions 3D produites à l'aide de mon système. L'image (d) met en évidence les limitations de la déformation à partir des points de contrôle.

¹<http://graphics.stanford.edu/software/trimesh/>

²<http://www.simunova.com/de/node/24>

³<http://phototour.cs.washington.edu/bundler/>

3.4.2 Génération de données de synthèse

L'intérêt de données de synthèse est de maîtriser parfaitement les conditions de prise de vue et la structure 3D afin de faciliter l'évaluation des méthodes de reconstruction 3D. On peut en effet comparer la surface estimée au modèle utilisé.

On dispose pour la synthèse d'un modèle numérique d'évaluation. Toutes ces informations sont utilisées premièrement pour un outil pour obtenir la structure 3D de la scène. Deuxièmement, par un outil OpenGL qui génère une suite des surfaces 3D — photométrique stéréo —, dans lesquels sont projetés les points de contrôles sur chaque surface. Enfin, en utilisant la représentation *Laplace* de la surface, on fait des déformations sur l'objet.

La préparation de la scène a été réalisée à l'aide du logiciel *Maya*. La figure (of the virtual scene) illustre la configuration de la scène simulée et la figure (optitrack) montre la localisation des caméras pour capturer le mouvement.

Ceci explique en partie le manque de points de contrôle pour déformer le maillage. Ce qui résulte dans une accumulation d'erreurs. La forte redondance des surfaces peut être utilisée pour estimer une structure tridimensionnelle plus riche et dense.

3.5 Résumé

Dans ce chapitre je présente la méthode permettant de reconstruire la forme 3D à partir d'une séquence d'images. Après d'avoir présenté les différentes méthodes de reconstruction: photométrique stéréo et structure-from-motion. Ici je me intéresse à la reconstruction 3D 'partir d'une séquence d'images acquises d'un capteur optique muni de trois lumières de couleurs (Rouge, Vert et Bleu) et direction différents. L'approche consiste à intégrer les informations supplémentaires des méthodes géométriques et photométriques. Pour déformer l'objet on considère l'utilisation de la représentation *Laplace*

L'image 3.4.1 (e) nous fait soulever un problème importante dans cette approche, que l'on retrouve fréquemment. C'est le fait que la collection de points de contrôle est insuffisante pour déformer l'objet. Ces obstacles produisent des effets indésirables. La solution possible est de utiliser un méthode pour régulariser la déformation de la surface de forme à réduire de grosses déplacements sur les vertex.

Chapter 4

Discussion et conclusion

Dans ce manuscrit j'ai présenté un système de reconstruction à partir d'une séquence d'images. Dans une première phase, j'ai fourni un rappel de notions fondamentales associés à la géométrie de caméras. Ensuite dans le chapitre 2 j'aborde le problème de la estimation de la surface à partir des gradients en utilisant la méthode photométric stéréo.

Cette approche privilégie l'apparence du modèle obtenu qui doit être plus fidèle possible à l'original. Pour cela j'ai utilisé des approches géométriques et radiométriques, comme la structure à partir du mouvement et photométrie multi-spectral. Cette approche d'un côté est bien adaptée pour reconstruire des objets rigides, d'autre côté est simple de mettre en place.

On peut résumer la tâche en cinq étapes:

1. *calcul de la nuage des points à partir SfM*
2. *estimer la surface 3D à partir des gradients*
3. *projeter les points de contrôles vues d'une caméra sur la respective surface*
4. *déformer chaque surface en utilisant sa représentation Laplace*
5. *fusionner les modèles obtenus*

Dans le cadre des travaux menés au cours de ce travail, je suis intéressé particulièrement à la conception d'un système capable de fournir des reconstructions 3D précises. C'es pourquoi la choix de la méthode photométrie stéréo. Cette est connue

pour la haute résolution puisque le calcul est effectué pour chaque pixel de l'image. Les résultats obtenus sont satisfaisantes mais ils mettent en valeur un certain nombre de limites. Fig.3.2 met en évidence l'utilisation des points de contrôle pour déformer le modèle. Néanmoins, on peut contourner cette difficulté en utilisant. On a alors constaté que la déformation provoque des effets indésirables quand la collection des points de contrôle est insuffisant Fig.3.4.1 (c) et (d). Puisque cette algorithm est cumulatif, calcul de la déformation d'une séquence de surfaces pris en instants différents peut atténuer ce type d'erreurs. Notons qu'il est nécessaire que la estimation du mouvement de la caméra ne soit trop erronée car en dépend les points de contrôle, sur lesquelles sont basés les déformations.

Les images Fig.3.4.1 (a) et (b) montrent que cette technique d'ajustement de la surface permet d'approcher la surface initial au objet original. Les points de contrôle améliorent ainsi la précision du volume de la forme global. La méthode ne prend pas en compte le phénomène de occultation. Pour améliorer ces résultats on peut d'abord tester l'algorithme dans une scène réelle avec des vidéo réelles, appliquer différentes méthodes de régularisation et heuristiques pour améliorer la déformation.

Bibliography

- [Botsch et al.,] Botsch, M., Kobbelt, L., Pauly, M., Alliez, P., and Levy, B. Polygon mesh processing.
- [Botsch et al., 2007] Botsch, M., Pauly, M., Kobbelt, L., Alliez, P., Lévy, B., Bischoff, S., and Rössl, C. (2007). Geometric modeling based on polygonal meshes video files associated with this course are available from the citation page. In *ACM SIGGRAPH 2007 courses*, SIGGRAPH '07, New York, NY, USA. ACM.
- [Brown, 1971] Brown, D. C. (1971). Close-range camera calibration. *PHOTOGRAMMETRIC ENGINEERING*, 37(8):855–866.
- [Chang et al., 2007] Chang, J. Y., Lee, K. M., and Lee, S. U. (2007). Multiview normal field integration using level set methods. In *CVPR*.
- [Clarke and Fryer,] Clarke, T. and Fryer, J. The development of camera calibration methods and models.
- [Curless and Levoy, 1995] Curless, B. and Levoy, M. (1995). Better optical triangulation through spacetime analysis. Technical report, Stanford, CA, USA.
- [Curless and Levoy, 1996] Curless, B. and Levoy, M. (1996). A volumetric method for building complex models from range images. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, SIGGRAPH '96, pages 303–312, New York, NY, USA. ACM.

- [d'Angelo and W,] d'Angelo, P. and W, C. Image-based 3d surface reconstruction by combination of sparse depth data with shape from shading and polarisation.
- [Desbrun et al.,] Desbrun, M., Schroder, P., and Wardetsky, M. Discrete differential geometry an applied introduction.
- [Ens and Lawrence, 1993] Ens, J. and Lawrence, P. (1993). An investigation of methods for determining depth from defocus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(2):XX–XX.
- [Fassold et al., 2004] Fassold, H., Danzl, R., Schindler, K., and Bischof, H. (2004). Reconstruction of archaeological finds using shape from stereo and shape from shading. In *9 th Computer Vision Winter Workshop (CVWW), February, Piran*, pages 21–30.
- [Faugeras and Luong, 2001] Faugeras, O. and Luong, Q.-T. (2001). *The Geometry of Multiple Images: the laws that govern the formation of multiple images of a scene and some their applications*. The MIT Press, XXXX.
- [Faugeras et al., 1992] Faugeras, O. D., Luong, Q.-T., and Maybank, S. J. (1992). Camera self-calibration: Theory and experiments. In *Proceedings of the Second European Conference on Computer Vision, ECCV '92*, pages 321–334, London, UK. Springer-Verlag.
- [Favaro and Soatto, 2002] Favaro, P. and Soatto, S. (2002). Learning depth from defocus. In *IEEE European Conference on Computer Vision*.
- [Frankot et al., 1988] Frankot, R. T., Chellappa, R., and Member, S. (1988). A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10:439–451.

- [Furukawa and Ponce, 2010] Furukawa, Y. and Ponce, J. (2010). Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:1362–1376.
- [Georghiadis, 2003] Georghiadis, A. S. (2003). Recovering 3-d shape and reflectance from a small number of photographs. In *Proceedings of the 14th Eurographics workshop on Rendering, EGRW '03*, pages 230–240, Aire-la-Ville, Switzerland, Switzerland. Eurographics Association.
- [Hasinoff, 2008] Hasinoff, W. S. (2008). *Variable Aperture Photography*. Phd, University of Toronto.
- [Hayakawa, 1994] Hayakawa, H. (1994). Photometric stereo under a light source with arbitrary motion. *Optical Society of America*, 11(11):XX–XX.
- [Hernández et al., 2007] Hernández, C., Vogiatzis, G., Brostow, G. J., Stenger, B., and Cipolla, R. (2007). Non-rigid photometric stereo with colored lights. In *In Proc. of ICCV*, pages 1–8.
- [Hormann et al., 2007] Hormann, K., Lévy, B., and Sheffer, A. (2007). Mesh parameterization: theory and practice video files associated with this course are available from the citation page. In *ACM SIGGRAPH 2007 courses*, SIGGRAPH '07, New York, NY, USA. ACM.
- [Horn, 1990] Horn, B. K. P. (1990). Height and gradient from shading. *International Journal of Computer Vision*, 5(1):37–75.
- [Horowitz and Kiryati, 2004] Horowitz, I. and Kiryati, N. (2004). Depth from gradient fields and control points: Bias correction in photometric stereo. *Image and Vision Computing*, 22:681–694.

- [Klette and Schlüns, 1996] Klette, R. and Schlüns, K. (1996). Height data from gradient fields. In *Proceedings of SPIE (the international Society for Optical Engineering) on Machine Vision Applications, Architectures, and Systems Integration*, pages 204–215.
- [Kovesi, 2005] Kovesi, P. (2005). Shapelets correlated with surface normals produce surfaces. In *In ICCV05*, pages 994–1001.
- [Laurentini, 1994] Laurentini, A. (1994). The visual hull concept for silhouette-based image understanding. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16:150–162.
- [Levin et al., 2007] Levin, A., Fergus, R., Durand, F., and Freeman, W. T. (2007). Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph.*, 26.
- [Lipman et al., 2004] Lipman, Y., Sorkine, O., Cohen-Or, D., Levin, D., Rössl, C., and Seidel, H.-P. (2004). Differential coordinates for interactive mesh editing. In *Proceedings of Shape Modeling International*, pages 181–190. IEEE Computer Society Press.
- [Lohse and Heipke, 2004] Lohse, V. and Heipke, C. (2004). Multi-image shape from shading: derivation of planetary digital terrain models using clementine images.
- [Lowe, 2004] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60:91–110.
- [Marshall et al., 1996] Marshall, J. A., Marshall, J. A., Ariely, D., Burbeck, C. A., Aricly, T. D., Rolland, J. P., and Martin, K. E. (1996). Occlusion edge blur: A cue to relative visual depth. *Intl. J. Opt. Soc. Am. A*, 13:681–688.
- [Mohr and Triggs, 1996] Mohr, R. and Triggs, B. (1996). Projective geometry for image analysis.

- [Nayar, 1989] Nayar, S. (1989). Shape from focus.
- [Nayar and Nakagawa, 1994] Nayar, S. and Nakagawa, Y. (1994). Shape from focus. In *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 824–831.
- [Nealen et al., 2005a] Nealen, A., Sorkine, O., Alexa, M., and Cohen-Or, D. (2005a). A sketch-based interface for detail-preserving mesh editing. *ACM Trans. Graph.*, 24:1142–1147.
- [Nealen et al., 2005b] Nealen, A., Sorkine, O., Alexa, M., and Cohen-Or, D. (2005b). A sketch-based interface for detail-preserving mesh editing. In *ACM SIGGRAPH 2005 Papers*, SIGGRAPH '05, pages 1142–1147, New York, NY, USA. ACM.
- [Paterson et al.,] Paterson, J. A., Claus, D., and Fitzgibbon, A. W. Brdf and geometry capture from extended inhomogeneous samples using flash photography. *Comput. Graph. Forum*.
- [Paterson and Fitzgibbon, 2002] Paterson, J. A. and Fitzgibbon, A. W. (2002). Flexible bump map capture from video. In *Annual Conference of the European Association for Computer Graphics (Eurographics)*.
- [R. Lena and Fattinanzi,] R. Lena, C. Wöhler, M. T. B. and Fattinanzi, C. A combined morphometric and spectrophometric study of the complex lunar volcanic region in the south of petavius.
- [S. K. Nayar and Boult, 1993] S. K. Nayar, X.-S. F. and Boult, T. (1993). Removal of specularities using color and polarization. In *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR '93., 1993 IEEE Computer Society Conference*, pages 583–590.
- [Schechner, 1999] Schechner, Y. Y. (1999). The optimal axial interval in estimating depth from defocus. In *In Proc. of the Intl. Conf. of Comp. Vision*, pages 843–848.

- [Schlüns, 1997] Schlüns, K. (1997). Shading based 3d shape recovery in the presence of shadows. In *In Proc. Intern. Conf. on Digital Image and Vision Computing*, pages 10–12.
- [Snavely et al., 2006] Snavely, N., Seitz, S. M., and Szeliski, R. (2006). Photo tourism: exploring photo collections in 3d. In *ACM SIGGRAPH 2006 Papers*, SIGGRAPH '06, pages 835–846, New York, NY, USA. ACM.
- [Snavely et al., 2008] Snavely, N., Seitz, S. M., and Szeliski, R. (2008). Modeling the world from internet photo collections. *Int. J. Comput. Vision*, 80:189–210.
- [Sorkine, 2006] Sorkine, O. (2006). Differential representations for mesh processing. *Computer Graphics Forum*, 25(4):789–807.
- [Sorkine and Alexa, 2007] Sorkine, O. and Alexa, M. (2007). As-rigid-as-possible surface modeling. In *Proceedings of the fifth Eurographics symposium on Geometry processing*, pages 109–116, Aire-la-Ville, Switzerland, Switzerland. Eurographics Association.
- [Sorkine et al.,] Sorkine, O., Cohen-Or, D., Lipman, Y., and Alexa, M. Laplacian surface editing.
- [Sorkine et al., 2004] Sorkine, O., Cohen-Or, D., Lipman, Y., Alexa, M., Rössl, C., and Seidel, H.-P. (2004). Laplacian surface editing. In *Proceedings of the Eurographics/ACM SIGGRAPH Symposium on Geometry Processing*, pages 179–188. ACM Press.
- [Subbarao and Surya, 1994] Subbarao, M. and Surya, G. (1994). Depth from defocus: A spatial domain approach. *International Journal of Computer Vision*, 13:271–294.
- [Terzopoulos et al., 1987] Terzopoulos, D., PLatt, J., Barr, A., and Fleischer, K. (1987). Elastic deformable models.

- [Turk and Levoy, 1994] Turk, G. and Levoy, M. (1994). Zippered polygon meshes from range images. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, SIGGRAPH '94, pages 311–318, New York, NY, USA. ACM.
- [Wöhler,] Wöhler, C. *3D Computer Vision: Efficient Methods and Applications*.
- [Wöhler and Hafezi, 2005] Wöhler, C. and Hafezi, K. (2005). A general framework for three-dimensional surface reconstruction by self-consistent fusion of shading and shadow features. *Pattern Recogn.*, 38:965–983.
- [Woodham, 1980] Woodham, R. J. (1980). Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144.
- [Xiong and Shafer, 1993] Xiong, Y. and Shafer, S. (1993). Depth from focusing and defocusing. In *In Proc. of the DARPA Image Understanding Workshop*, pages 68–73.
- [Yang and Ivriissimtzis, 2010] Yang, Y. and Ivriissimtzis, I. (2010). Polygonal mesh watermarking using laplacian coordinates. *Computer Graphics Forum (proceedings of Eurographics/ACM SIGGRAPH Symposium on Geometry Processing)*, 29(5).
- [Zhang et al., 2003] Zhang, L., Curless, B., Hertzmann, A., and Seitz, S. M. (2003). Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. *Computer Vision, IEEE International Conference on*, 1:618.
- [Zhang et al., 2004] Zhang, L., Snavely, N., Curless, B., and Seitz, S. M. (2004). Spacetime faces: high resolution capture for modeling and animation. *ACM Trans. Graph.*, 23:548–558.
- [Zhang, 2000] Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22:1330–1334.

- [Zhou et al., 2009] Zhou, C., Lin, S., and Nayar, S. (2009). Coded Aperture Pairs for Depth from Defocus. Kyoto, Japan.
- [Zhou and Nayar,] Zhou, C. and Nayar, S. What are good apertures for defocus deblurring? *Computational Photography (ICCP), 2009 IEEE International Conference.*
- [Zhou et al., 2005a] Zhou, K., Huang, J., Snyder, J., Liu, X., Bao, H., Guo, B., and Shum, H.-Y. (2005a). Large mesh deformation using the volumetric graph laplacian. *ACM Trans. Graph.*, 24:496–503.
- [Zhou et al., 2005b] Zhou, K., Huang, J., Snyder, J., Liu, X., Bao, H., Guo, B., and Shum, H.-Y. (2005b). Large mesh deformation using the volumetric graph laplacian. In *ACM SIGGRAPH 2005 Papers*, SIGGRAPH '05, pages 496–503, New York, NY, USA. ACM.
- [Zhou and Tan, 2010] Zhou, Z. and Tan, P. (2010). Ring-light photometric stereo. In *Proceedings of the 11th European conference on Computer vision: Part II, ECCV'10*, pages 265–279, Berlin, Heidelberg. Springer-Verlag.
- [Ziou, 1997] Ziou, D. (1997). Passive depth from defocus using a spatial domain approach. In *In Proc. of the Intl. Conf. of Computer Vision*, pages 799–804.