



HAL
open science

Reconstruction 3D par stéréophotométrie pour la vision omnidirectionnelle

Jordan Caracotte

► **To cite this version:**

Jordan Caracotte. Reconstruction 3D par stéréophotométrie pour la vision omnidirectionnelle. Vision par ordinateur et reconnaissance de formes [cs.CV]. Université de Picardie Jules Verne, 2021. Français. NNT : 2021AMIE0031 . tel-03828820

HAL Id: tel-03828820

<https://theses.hal.science/tel-03828820>

Submitted on 25 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Thèse de Doctorat

Mention Sciences pour l'Ingénieur

Spécialité Vision par Ordinateur

présentée à l'École Doctorale en Sciences, Technologie, Santé (ED 585)

à l'Université de Picardie Jules Verne

par

Jordan CARACOTTE

pour obtenir le grade de Docteur de l'Université de Picardie Jules Verne

**Reconstruction 3D par stéréophotométrie
pour la vision omnidirectionnelle**

Soutenue le 06 juillet 2021, après avis des rapporteurs, devant le jury d'examen :

M. J.-D. DUROU, MdC HDR, Université Paul Sabatier, Toulouse

Rapporteur

M. P. STURM, DR, Inria Grenoble Rhône-Alpes

Rapporteur

M^{me} S. BOUCHAFA-BRUNEAU, PU, Université d'Évry

Examineur

M. Y. QUÉAU, CR CNRS, GREYC, Caen

Examineur

M. E.M. MOUADDIB, PU, Université de Picardie Jules Verne

Directeur de thèse

M. F. MORBIDI, MdC, Université de Picardie Jules Verne

Co-encadrant

Remerciements

La préparation de cette thèse de doctorat constitue une aventure scientifique et avant tout humaine, pendant laquelle j'ai pu faire de nombreuses rencontres, et je tiens à remercier les personnes qui m'ont accompagné pendant tout ou partie du chemin jusqu'à l'aboutissement de cette thèse.

Tout d'abord, je pense aux membres du jury qui ont accepté d'évaluer mes travaux de thèse et d'y avoir apporté des critiques dans le but d'enrichir ces derniers : aux rapporteurs Jean-Denis DUROU et Peter STURM pour la lecture attentive et les riches conseils qui ont suivi, à Samia BOUCHAFA-BRUNEAU pour avoir accepté de présider le jury à distance, ainsi qu'à Yvain QUÉAU pour les échanges nourris qui ont eu lieu pendant la soutenance et qui m'ont apporté de nouvelles pistes.

Je pense naturellement à mon directeur de thèse El Mustapha MOUADDIB ainsi qu'à mon encadrant Fabio MORBIDI, qui ont été mes enseignants tout au long de ma formation, de l'IUT d'Amiens au département EEA, avant de me guider pendant la préparation de cette thèse. Je les remercie aussi de m'avoir transmis leurs approches pour l'enseignement dans les parcours de formation du département EEA. À ce sujet, je remercie aussi Alex POTELLE, Frederic COLLET, Djemâa KACHI, Abdelhamid RABHI, Jérôme BOSCHE, Pierre LEBRUN, Hervé MIDAVAINÉ, Gaëlle ALLART ainsi que les autres membres du département EEA pour les échanges formels et informels, autour d'un café ou non, que nous avons pu avoir. Cela a fortement contribué à mon investissement dans les enseignements que j'ai pu réaliser.

Je pense à Guillaume CARON, enseignant pendant ma formation puis chef d'équipe au laboratoire pendant les premières années de préparation de la thèse. Je me souviendrai longtemps de ce séjour de recherche au Japon, dont les premiers jours ont été guidés par Fabio, Guillaume et Cédric DEMONCEAUX. Je remercie encore à cette occasion Takuya FUNATOMI ainsi que toute l'équipe de *Optical Media Interface Lab* pour leur accueil.

Un grand merci aux membres du laboratoire MIS pour les moments scientifiques et/ou de convivialité que nous avons partagés. À cet instant, je pense à Gilles DEQUEN, Claude PÉ-GARD, Corinne LUCET-VASSEUR, Céline JOIRON, Juliette DUBOIS, David DURAND et Yu LI. Une pensée à la marche du vendredi dont il ne reste plus que la récompense.

Je remercie l'équipe de l'école doctorale STS dont son directeur Ahmed EL HAJJAJI, pour le soutien qui m'a été apporté pendant les épreuves personnelles rencontrées pendant la préparation.

Une pensée particulière aux jeunes chercheurs du laboratoire avec lesquels j'ai pu échanger, trouver des réponses, prendre du recul voire avec qui j'ai construit des activités de vulgarisation : Clément LECAT, Romuald CARETTE, Charlotte PIAU, Monika TRIMOSKA, Olivier GERARD, Nathan CROMBEZ, Zaynab HABIBI, Alejandro Eder RODRIGUEZ MARTINEZ, Nathan SANCHIZ, Farouk ACHAKIR, Yassine AHMINE, Seif-Eddine GUERBAS, Thibault POTIN, Tawsif GOKHOOK, Sarah DELMAS, Audrey AMBLES, ainsi que Julien DUCROCQ.

Pour finir, je remercie mes proches, famille et amis, qui m'ont soutenu pendant la préparation de cette thèse, dont Loïc BOULLEMANT et Cyrielle ANDRÉ qui a pu découvrir ce manuscrit complet la première.

Table des matières

Introduction générale	1
1 Introduction à la reconstruction 3D par stéréophotométrie	5
1.1 Formation des images	6
1.1.1 Modèle géométrique	6
1.1.2 Modèle photométrique	8
1.2 Estimation des normales	10
1.2.1 Cas idéal	10
1.2.2 Rapports photométriques	11
1.2.3 Sources d'erreur	13
1.3 Estimation de la profondeur	15
1.3.1 Gradients de profondeur	16
1.3.2 Intégration des gradients de profondeur	18
1.4 Jeux de données	20
2 Stéréophotométrie pour les capteurs à projection centrale	23
2.1 Modèle unifié pour les capteurs à projection centrale	24
2.1.1 Présentation du modèle	24
2.1.2 Représentation d'une image sphérique	25
2.2 Estimation des gradients orientés	26
2.2.1 Équation de l'irradiance pour les images sphériques	26
2.2.2 Rapports photométriques	29
2.3 Estimation des distances radiales	30
2.3.1 Résolution adaptée aux images panoramiques	30
2.4 Expérimentations	32
2.4.1 Validation en simulation	33
2.4.2 Évaluation en conditions réelles et contrôlées	38
3 Extension aux caméras <i>twin-fisheye</i>	48
3.1 Caméra <i>twin-fisheye</i> : applications	50
3.2 Adaptation de la chaîne de traitements	56
3.2.1 Modélisation du capteur	56
3.2.2 Fusion des données	59
3.2.3 Estimation des distances radiales	60

3.3	Expérimentations	60
3.3.1	Validation en simulation	61
3.3.2	Évaluation en conditions réelles et contrôlées	63
3.3.3	Discussion	67
3.4	Analyse des reconstructions à partir de données réelles	68
3.4.1	Éclairage non directionnel	68
3.4.2	Atténuation de l'éclairage	69
	Conclusion générale	76
	Bibliographie	81

Introduction générale

Depuis plus de quatre décennies, les techniques de reconstruction 3D par stéréophotométrie usuelle offrent la possibilité de retrouver la géométrie d'une surface. Dans sa configuration la plus simple, l'objet à reconstruire est immobile et illuminé par au moins trois éclairages différents et successifs. Pour chaque éclairage, une caméra statique observe la scène et capture une image. À partir de ces observations et pour chaque pixel de l'image qui représente l'objet, Woodham a proposé dans [Woodham 1980] une solution afin de retrouver l'orientation de la surface, au prix de plusieurs hypothèses. Parmi ces hypothèses, la surface observée doit être lambertienne, elle doit restituer la même quantité de lumière, peu importe la position de l'observateur. Les ombres et les rebonds de lumière (inter-réflexions) sont ignorés et d'autres contraintes sont imposées au niveau des éclairages : la lumière ne subit pas d'atténuation, les rayons lumineux sont tous parallèles et la source est homogène. De fait, les images capturées au fur et à mesure des illuminations successives sont obtenues dans des environnements clos et contrôlés. Ces méthodes, qui se distinguent de celles exploitant un ensemble de points de vue différents, suscitent toujours un intérêt particulier dans la recherche en vision par ordinateur (voir [Ackermann & Goesele 2015, Furukawa & Hernández 2015]).

De nombreuses extensions de la méthode originale ont été proposées dans la littérature. Par exemple, dans [Abrams *et al.* 2012, Ackermann *et al.* 2012, Abrams *et al.* 2013, Jung *et al.* 2015], les contraintes fortes imposées sur l'éclairage ont été relâchées et des images provenant de scènes en extérieur, illuminées naturellement, ont été exploitées pour réaliser des reconstructions 3D de bâtiments ou de rues entières. Lorsque nous ne disposons pas d'informations sur l'éclairage (le nombre de sources, la direction des rayons) ou sur la surface observée, le problème de stéréophotométrie "non calibrée" apparaît et, bien que mal posé mathématiquement, de nombreuses approches tentent de le résoudre (voir [Hayakawa 1994, Basri *et al.* 2007, Quéau *et al.* 2015, Quéau *et al.* 2017, Papadimitri & Favaro 2013]). Cette version dégradée du problème de stéréophotométrie peut rappeler les techniques de *shape-from-shading* qui cherchent à retrouver l'orientation d'une surface à partir d'une seule observation et pour laquelle la direction de l'éclairage est connue (par exemple, voir [Prados & Faugeras 2006]).

D'autres améliorations se sont concentrées sur une modélisation davantage complète du comportement de la surface vis-à-vis de la lumière qu'elle reçoit puis restitue. Très simplement, les fonctions de distribution de la réflectivité bidirectionnelle, ou *Bidirectional Reflectance Distribution Functions* (BRDF) en anglais, permettent d'établir un lien entre la quantité d'énergie des rayons lumineux reçus par une portion d'une surface et celle restituée par réflexion. La prise en compte d'éclairages proches, c'est-à-dire, non directionnels (par exemple, voir [Mecca *et al.* 2014, Logothetis *et al.* 2017]) ou la réduction des contraintes

sur le placement de la caméra pour permettre de combiner les techniques de reconstruction multi-vues et stéréophotométrie (par exemple, voir [Hernández *et al.* 2008, Vlastic *et al.* 2009, Logothetis *et al.* 2019]) sont aussi des axes d'amélioration de la méthode originale.

Les nombreux efforts pour perfectionner l'approche de Woodham ont permis de l'étendre au modèle de projection perspective (par exemple, voir [Tankus & Kiryati 2005]). Néanmoins, à notre connaissance et malgré une première tentative dans [Kondo *et al.* 2008], nous n'avons pas trouvé de résultats de travaux réalisant des reconstructions 3D par stéréophotométrie et impliquant les capteurs à projection centrale, de façon générale. Ces capteurs de vision garantissent l'unicité du point de vue, peu importe la taille du champ de vue dont ils disposent. Par exemple, en couplant une caméra perspective avec un miroir hyperbolique, il est possible d'obtenir un système de vision panoramique permettant de capturer des images à 360°. Avec ce type d'image, il est possible de capturer l'intégralité de la scène se trouvant autour de la caméra en un faible nombre de points de vue. Cet avantage devient évident dans une scène en intérieur : par exemple, si nous souhaitons réaliser la reconstruction 3D de la portion d'un escalier hélicoïdal où se trouve une caméra, nous pouvons capturer toute la scène en utilisant quelques points



FIGURE 1 – Image d'illustration d'un escalier hélicoïdal obtenue à partir d'une caméra perspective.

de vue avec un capteur de vision omnidirectionnelle. Avec une caméra perspective, il serait nécessaire de combiner de nombreuses images comme celle de la figure 1, obtenues à partir de nombreux points de vue qu'il faut finalement aligner pour obtenir une observation complète de la scène. Malgré tout, ces capteurs panoramiques à projection centrale ne sont pas parfaits et souffrent d'autres limitations : le très large champ de vue offert est à comparer avec la résolution spatiale hétérogène et des distorsions peuvent apparaître en fonction de la conception et/ou de la réalisation du capteur.

Contributions

Dans ce manuscrit, nous nous intéresserons aux capteurs de vision pour lesquels le modèle unifié de projection centrale peut être appliqué et nous chercherons à résoudre le problème de stéréophotométrie en exploitant des images simulées et réelles de plusieurs capteurs avec un champ de vue croissant.

Les contributions originales introduites et présentées dans ce document sont :

1. Une modélisation unifiée du problème de stéréophotométrie, compatible avec tous les capteurs de vision à projection centrale ;
2. Une chaîne de traitements adaptée et modulable pour la reconstruction 3D par stéréophotométrie à partir de capteurs étalonnés, exploitant la modélisation unifiée et une approche pour la phase d'intégration qui tire profit des images sphériques ;
3. Une adaptation de la méthode présentée dans [Schnieders & Wong 2013] pour l'étalonnage des directions des éclairages pour la rendre compatible avec le modèle de projection ;
4. Une extension de la chaîne de traitements proposée pour permettre la reconstruction 3D par stéréophotométrie à partir des images acquises depuis des caméras *twin-fisheye*.

Ces contributions s'inscrivent dans la thématique « Cartographie du patrimoine monumental » de l'équipe Perception Robotique du laboratoire Modélisation, Information et Systèmes. Cette thématique regroupe les initiatives proposées, les efforts fournis et les travaux réalisés dans le but d'obtenir des sauvegardes numériques, les plus précises possibles, de grandes structures architecturales telles que la cathédrale Notre-Dame d'Amiens (voir [Mouaddib *et al.* 2025]).

Résumé du contenu

Le contenu du manuscrit est structuré en trois chapitres. Dans le **premier**, nous aborderons la reconstruction 3D par stéréophotométrie usuelle ainsi que les prérequis nécessaires à

la compréhension des chapitres suivants. Pour commencer, nous présenterons les modèles mathématiques considérés pour la formation des images, puis nous poursuivrons plus en détails sur la résolution du problème de stéréophotométrie en deux étapes. Enfin, nous présenterons quelques jeux de données utiles pour le développement et/ou la comparaison des techniques de reconstruction par stéréophotométrie. Dans le **deuxième chapitre**, nous introduirons le modèle unifié pour les capteurs à projection centrale. Après avoir détaillé la représentation des images sphériques que nous utiliserons, nous détaillerons la chaîne de traitements qui permet de réaliser des reconstructions 3D par stéréophotométrie à partir de n'importe quel capteur de vision à projection centrale. Nous illustrerons la méthode avec une caméra perspective et un capteur catadioptrique. Dans le **troisième chapitre**, nous reprendrons les notions introduites dans le chapitre 2 et présenterons une adaptation de la chaîne de traitements dédiée aux caméras *twin-fisheye*. Nous terminerons ce chapitre avec une étude des limitations apparentes de la méthode proposée. Dans la conclusion générale, nous rappellerons le contenu de l'intégralité du document puis nous listerons et détaillerons quelques perspectives d'amélioration des travaux présentés tout au long du manuscrit.

Publications

L'ensemble des contributions abordées dans ce manuscrit peuvent être retrouvées dans les deux articles suivants qui ont été soumis et acceptés pendant la préparation de la thèse :

- [Caracotte *et al.* 2020] J. Caracotte, F. Morbidi et E.M. Mouaddib. *Photometric stereo with central panoramic cameras*. *Comput. Vis. Image Und.*, vol. 201, article 103080, 2020.
- [Caracotte *et al.* 2021] J. Caracotte, F. Morbidi et E.M. Mouaddib. *Photometric Stereo with Twin-Fisheye Cameras*. In *Proc. IEEE Int. Conf. Pattern Recogn.*, pages 5270-5277, 2021.

En complément des travaux de recherche sur la stéréophotométrie adaptée aux capteurs de vision omnidirectionnelle, la publication suivante s'appuie sur le modèle unifié, exploité dans ce manuscrit :

- [Hajjami *et al.* 2020] J. Hajjami, J. Caracotte, G. Caron et T. Napoleon. *ArUcOmni : detection of highly reliable fiducial markers in panoramic images*. In *Proc. IEEE/CVF Conf. Comput. Vis. and Pattern Recogn. Workshops*, pages 634-635, 2020.

Introduction à la reconstruction 3D par stéréophotométrie

Contenu du chapitre

1.1	Formation des images	6
1.1.1	Modèle géométrique	6
1.1.2	Modèle photométrique	8
1.2	Estimation des normales	10
1.2.1	Cas idéal	10
1.2.2	Rapports photométriques	11
1.2.3	Sources d'erreur	13
1.3	Estimation de la profondeur	15
1.3.1	Gradients de profondeur	16
1.3.2	Intégration des gradients de profondeur	18
1.4	Jeux de données	20

Résumé

Ce chapitre est dédié à la stéréophotométrie ainsi qu'à l'exposé de quelques approches proposées depuis la formulation originale de Woodham pour reconstruire la géométrie d'une surface observée. Dans une première section, nous détaillerons les modèles utilisés pour la formation des images numériques, d'un point de vue géométrique puis photométrique. Dans une deuxième section, nous nous intéresserons en particulier à la résolution du problème de stéréophotométrie dans un cadre idéal. Nous poursuivrons en introduisant quelques améliorations présentes dans la littérature. Dans une troisième section, nous aborderons ensuite les gradients de profondeur ainsi que les approches existantes permettant de reconstruire la géométrie de la surface observée, après avoir résolu le problème de stéréophotométrie. Enfin, nous présenterons quelques jeux de données disponibles publiquement qui peuvent être utilisés afin d'évaluer une méthode en cours de développement pour l'estimation des normales et/ou la reconstruction de la surface observée.

1.1 Formation des images

Dans cette première partie, nous nous concentrons sur le modèle de formation des images numériques. En commençant par les modèles géométriques des caméras orthographiques et perspectives, nous terminerons avec les modèles photométriques qui permettent d'obtenir l'intensité mesurée pour chaque pixel.

1.1.1 Modèle géométrique

Une surface observée par une caméra est composée d'un ensemble de points $\mathbf{X} = [X, Y, Z]^T \in \mathbb{R}^3$. Nous supposons que les coordonnées de chaque point sont définies dans le repère $\{O; \vec{X}, \vec{Y}, \vec{Z}\}$ de la caméra. L'origine O de ce repère correspond au centre optique du capteur, le plan XY est parallèle au plan image métrique xy et l'axe optique est confondu avec l'axe \vec{Z} . Lorsque les coordonnées des points 3D sont définies dans un repère différent du repère de la caméra, nous pouvons facilement ramener les points 3D dans le repère attendu en utilisant une transformation rigide. Dans la suite du manuscrit, pour simplifier les expressions, nous considérerons que les points 3D qui composent les surfaces observées sont toujours définies dans le repère de la caméra. L'objectif de la modélisation géométrique est d'établir le lien entre les coordonnées des points 3D et leurs coordonnées $\mathbf{u}(u, v)$ dans le plan image. Après projection, chaque point visible de la surface est défini dans les plans images métrique et numérique, respectivement aux coordonnées (x, y) et (u, v) .

Des distorsions géométriques liés aux imperfections des capteurs peuvent apparaître et ne sont pas intégrées à la modélisation. Ces perturbations sont différentes des déformations observables avec l'utilisation d'une caméra grand angle ou d'un capteur catadioptrique.

Projection orthographique

Dans le cas d'une projection orthographique sans changement d'échelle (projection ortho-normale), nous plaçons le plan image métrique xy pour qu'il soit confondu avec le plan XY (voir la figure 1.1). En coordonnées homogènes, la transformation appliquée aux coordonnées des points \mathbf{X} qui permet de passer d'un point 3D sur la surface de l'objet à un point 2D dans le plan image métrique, est :

$$\mathbf{x}(x, y) = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}. \quad (1.1)$$

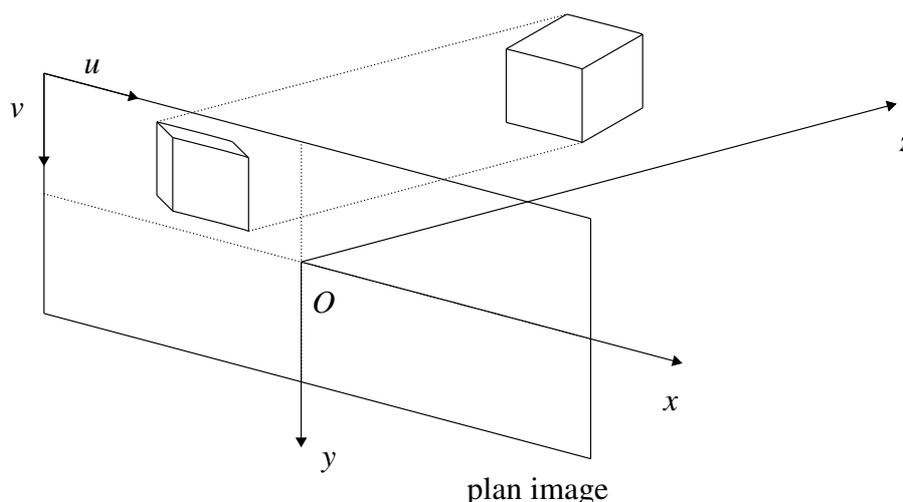


FIGURE 1.1 – Formation d’une image par projection orthographique.

Pour obtenir $\mathbf{u}(u, v)$, il faut appliquer une deuxième transformation afin de prendre en compte les paramètres internes de la caméra :

$$\mathbf{u}(u, v) = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} k_u x + u_0 \\ k_v y + v_0 \\ 1 \end{bmatrix}, \quad (1.2)$$

avec le point principal (u_0, v_0) et les coefficients de densité de pixels de la matrice photosensible k_u et k_v (exprimés en pixels par mm, par exemple), le long des directions \vec{u} et \vec{v} , respectivement. Ces paramètres peuvent être estimés par étalonnage du capteur et permettent de passer du plan image métrique au numérique (et vice versa).

Projection perspective

Dans le cas d’une projection perspective, le plan image métrique xy n’est plus confondu avec le plan XY (voir la figure 1.2). Il est perpendiculaire à l’axe optique de la caméra et situé à une distance égale à la longueur focale f de l’objectif. Plus précisément, cette distance est mesurée entre le centre optique et le plan image. La transformation devient alors :

$$\mathbf{x}(x, y) = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} f \frac{X}{Z} \\ f \frac{Y}{Z} \\ 1 \end{bmatrix}. \quad (1.3)$$

De la même manière qu’avec la projection orthographique, la transformation (1.2) permet de passer au plan image numérique.

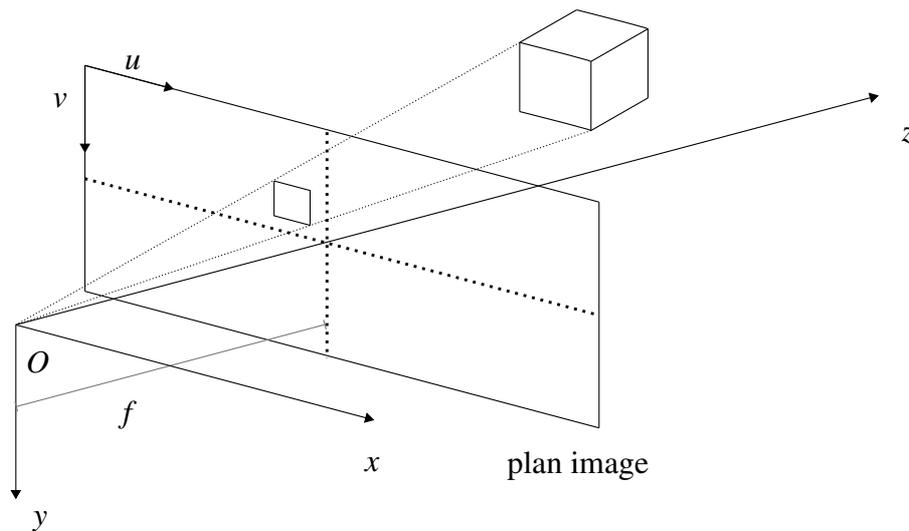


FIGURE 1.2 – Formation d’une image par projection perspective.

1.1.2 Modèle photométrique

Alors que dans la partie 1.1.1, nous avons présenté la manière avec laquelle les coordonnées 3D se projettent dans l’image, cette section va introduire la modélisation des intensités lumineuses qui seront associées aux points \mathbf{X} . Pour cela, nous supposons l’utilisation d’une caméra monochrome (ou d’une conversion des images couleurs en nuances de gris) ainsi que d’une source de lumière blanche qui illumine la surface décrite par les points \mathbf{X} (voir la figure 1.3).

Nous introduisons le vecteur $\mathbf{N}(\mathbf{X}(X, Y, Z)) = [N_x, N_y, N_z]^T$ qui décrit l’orientation de la surface à chaque point \mathbf{X} . Après projection (orthographique ou perspective) du point \mathbf{X} , nous pouvons écrire :

$$\mathbf{N}(\mathbf{x}(X, Y, Z)) = \mathbf{N}(\mathbf{u}(u, v)),$$

ou, par souci de simplicité, $\mathbf{N}(u, v)$. Nous notons l’éclairage incident au point \mathbf{x} avec le vecteur de direction $\mathbf{L}(\mathbf{X}) = [L_x, L_y, -1]^T$. Lorsque nous supposons que l’éclairage est directionnel, nous utilisons simplement \mathbf{L} , puisque la direction et l’intensité sont identiques pour tout point \mathbf{X} .

Une surface lambertienne émet la même quantité de lumière peu importe l’angle entre la ligne de vue et le vecteur normal à la surface au point étudié (voir [Ma et al. 2004, Chapitre 3]). En respectant les conditions d’éclairage homogène et directionnel, la quantité de lumière réfléchiée par la surface est proportionnelle à la luminance émise au point \mathbf{X} . Elle reste identique quelque soit la position de l’observateur par rapport au point \mathbf{X} car elle dépend du fait que la surface est lambertienne. Lorsque la caméra capture une image de la scène, le niveau de gris pour chaque point $\mathbf{u}(u, v)$ dans l’image I est mesuré. Selon les hypothèses d’un éclairage directionnel, homogène et constant ainsi que d’une surface lambertienne, l’irradiance mesurée par la

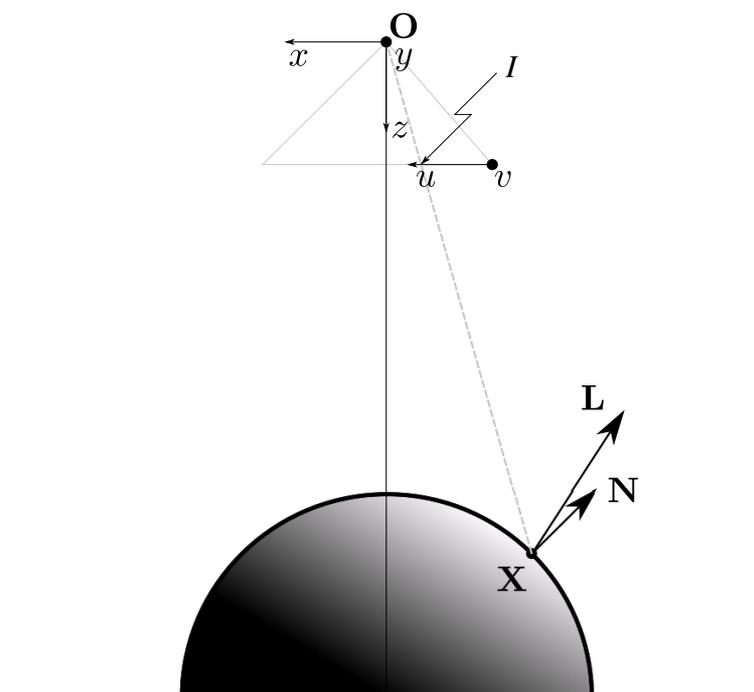


FIGURE 1.3 – Schéma en coupe d’une demi-sphère éclairée par une source directionnelle dont la direction est représentée par le vecteur \mathbf{L} . L’ombre générée n’est pas photoréaliste et sert d’illustration.

caméra peut être modélisée par le modèle suivant :

$$I(u, v) = \alpha(u, v) \max \{0, \langle \mathbf{N}(u, v), \mathbf{L} \rangle\}, \quad (1.4)$$

où $\alpha(u, v) > 0$ (simplification de $\alpha(\mathbf{u}(u, v))$) est l’albedo¹ de la surface au point \mathbf{X} . $\langle \mathbf{N}(u, v), \mathbf{L} \rangle$ correspond au produit scalaire entre les vecteurs $\mathbf{N}(u, v)$ et \mathbf{L} et l’opérateur max permet d’éliminer toutes les valeurs négatives qui correspondent, par construction, à des points de la surface orientée dans une direction opposée à la caméra. Ce modèle est une simplification de l’équation de la radiance totale² fréquemment employée lors de l’exposé du problème de stéréophotométrie (voir [Ackermann & Goesele 2015, Quéau 2015]).

L’utilisation d’un éclairage directionnel et l’observation d’une surface lambertienne sont deux hypothèses fortes. Nous reviendrons sur ces hypothèses dans les chapitres 2 et 3.

1. Il est usuel de noter l’albedo avec le symbole ρ . Cependant, nous utiliserons dans les chapitre 2 et 3 ce symbole pour noter la distance radiale dans une image sphérique.

2. Pour davantage de détails, voir [Nicodemus *et al.* 1992] pour la radiance totale et [Forsyth & Ponce 2012, Chapitre 4] pour une étude des phénomènes impliqués d’un point de vue radiométrique.

et $\mathbf{m}(u, v) = [m_x, m_y, m_z]^T = \alpha(u, v)\mathbf{N}(u, v)$. Le système (1.6) est maintenant linéaire par rapport à la nouvelle inconnue $\mathbf{m}(u, v)$. Il est à noter que nous pouvons définir $\alpha(u, v) = \|\mathbf{m}(u, v)\|$ et $\mathbf{N}(u, v) = \frac{\mathbf{m}(u, v)}{\|\mathbf{m}(u, v)\|}$, puisque $\mathbf{N}(u, v)$ est un vecteur unitaire.

La résolution du système (1.6) dépend du nombre d'images n et du rang de la matrice contenant les directions d'éclairages. Trois situations sont possibles :

- Le rang de la matrice Λ est égal à 3 et $m = 3$. Dans ce cas, le système admet une solution unique :

$$\mathbf{m}(u, v) = \Lambda^{-1} \begin{bmatrix} I_0(u, v) \\ \vdots \\ I_{n-1}(u, v) \end{bmatrix}$$

- Le rang de la matrice Λ est égal à 3 et $n > 3$; le système est surcontraint et une solution est approchée par les moindres carrés :

$$\mathbf{m}(u, v) = \Lambda^\dagger \begin{bmatrix} I_0(u, v) \\ \vdots \\ I_{n-1}(u, v) \end{bmatrix}$$

- Le rang de la matrice Λ est inférieur à 3 ; le système admet plusieurs de solutions et le problème est mathématiquement mal posé.

Cette approche fonctionne lorsque les données sont exactes et fidèles au modèle présenté. Naturellement, des biais peuvent apparaître lorsque les données s'éloignent des différentes hypothèses (surface lambertienne ainsi qu'un éclairage uniforme, constant et directionnel). Nous verrons dans la partie 1.2.3 que des biais peuvent aussi apparaître dans certaines configurations, alors que les hypothèses sont respectées.

1.2.2 Rapports photométriques

Nous nous concentrons à présent sur une amélioration de l'approche présentée dans la partie 1.2.1. Le but est de parvenir à estimer l'orientation de la surface sans devoir se préoccuper de l'albedo $\alpha(u, v)$.

Nous considérons à nouveau une caméra perspective étalonnée et fixe qui observe un objet statique sous différentes conditions d'illumination $\mathbf{L}_i, i \in \{0, n - 1\}$. D'après les travaux de [Tankus & Kiryati 2005], nous pouvons noter $\mathbf{X} = [X, Y, Z(X, Y)]^T$ avec $Z(X, Y)$ la fonction qui définit la profondeur. Après projection du point \mathbf{X} dans le plan image métrique, nous notons $Z(X, Y) = z(x, y)$ (de la même manière, nous pouvons noter $Z(X, Y) = z(u, v)$). L'équation (1.4)

est simplifiée afin d'obtenir $I(x, y) \in [0, 1]$:

$$I(x, y) = \alpha(x, y) \frac{\langle \mathbf{N}(x, y), \mathbf{L} \rangle}{\|\mathbf{L}\|} = \alpha(x, y) \langle \mathbf{N}(x, y), \widehat{\mathbf{L}} \rangle, \quad (1.7)$$

avec $\widehat{\mathbf{L}}(x, y) = \frac{\mathbf{L}(x, y)}{\|\mathbf{L}(x, y)\|}$. L'albedo $\alpha(x, y)$ est à nouveau compris dans l'intervalle $[0, 1]$.

Théorème 1 (Équation de l'irradiance pour la projection perspective [Tankus & Kiryati 2005]). *D'après les hypothèses précédentes et la notation définie, l'équation de l'irradiance pour la projection perspective est,*

$$I(x, y) = \alpha(x, y) \langle \widehat{\mathbf{N}}(x, y), \widehat{\mathbf{L}} \rangle = \frac{-\alpha(x, y) \mathbf{L}^T}{\|\mathbf{L}\| \sqrt{(xp + yq + 1)^2 + f^2(p^2 + q^2)}} \begin{bmatrix} fp \\ fq \\ xp + yq + 1 \end{bmatrix}, \quad (1.8)$$

où

$$p \triangleq \frac{1}{z} \frac{\partial z}{\partial x} = \frac{\partial \ln z}{\partial x}, \quad q \triangleq \frac{1}{z} \frac{\partial z}{\partial y} = \frac{\partial \ln z}{\partial y}. \quad \blacksquare$$

Avec cette formulation du problème, l'équation (1.8) repose sur les dérivées partielles de $\ln z(x, y)$. Pour estimer $z(x, y)$, nous allons d'abord estimer $\ln z(x, y)$. Puisque le logarithme naturel est une fonction bijective et par convention $z(x, y) > 0$, nous obtiendrons ensuite $z = \exp(\ln z(x, y))$. Ce point sera à nouveau abordé dans la partie §1.3.1.

L'équation de l'irradiance pour la projection perspective pour la i -ième image est :

$$I_i(x, y) = \alpha(x, y) \langle \widehat{\mathbf{N}}(x, y), \widehat{\mathbf{L}}_i \rangle = \frac{-\alpha(x, y) \mathbf{L}_i^T}{\|\mathbf{L}_i\| \sqrt{(xp + yq + 1)^2 + f^2(p^2 + q^2)}} \begin{bmatrix} fp \\ fq \\ xp + yq + 1 \end{bmatrix}. \quad (1.9)$$

En divisant la i -ième image par la k -ième³, nous pouvons construire le rapport photométrique suivant :

$$I_i(x, y) = \frac{\|\mathbf{L}_k\|((x - fL_{x_i})p + (y - fL_{y_i})q + 1)}{\|\mathbf{L}_i\|((x - fL_{x_k})p + (y - fL_{y_k})q + 1)} \quad (1.10)$$

Nous pouvons remarquer que l'albedo est absent de l'équation (1.9) et que les inconnues sont à présent p et q . Nous pouvons écrire l'équation (1.9) sous la forme compacte suivante :

$$A_{ik}p + B_{ik}q + C_{ik} = 0 \quad (1.11)$$

3. Nous supposons que toutes les intensités nulles ont été filtrées au préalable pour éviter toute division par zéro. Nous reviendrons sur ces intensités nulles dans la partie 1.2.3.

où,

$$\begin{aligned} A_{ik} &= I_i(x, y) \|\mathbf{L}_i\| (x - fL_{x_k}) - I_k(x, y) \|\mathbf{L}_k\| (x - fL_{x_i}), \\ B_{ik} &= I_i(x, y) \|\mathbf{L}_i\| (y - fL_{y_k}) - I_k(x, y) \|\mathbf{L}_k\| (y - fL_{y_i}), \\ C_{ik} &= I_i(x, y) \|\mathbf{L}_i\| - I_k(x, y) \|\mathbf{L}_k\|. \end{aligned}$$

En utilisant trois images, il est donc possible de trouver une solution analytique⁴ pour p et q . Bien que nous ayons deux inconnues, trois images sont nécessaires car deux images ne permettent de construire qu'une équation. Même s'il s'agit de *shape-from-shading* et non de stéréophotométrie, les développements mathématiques présents dans [Tankus *et al.* 2005] peuvent aider à la reproduction du détail des calculs.

À partir de p et q , il est enfin possible de retrouver le vecteur normalisé $\hat{\mathbf{N}}(x, y)$ au moyen de l'expression suivante :

$$\hat{\mathbf{N}}(x, y) = \frac{1}{\sqrt{(xp + yp + 1)^2 + f^2(p^2 + q^2)}} \begin{bmatrix} fp \\ fq \\ xp + yp + 1 \end{bmatrix}.$$

1.2.3 Sources d'erreur

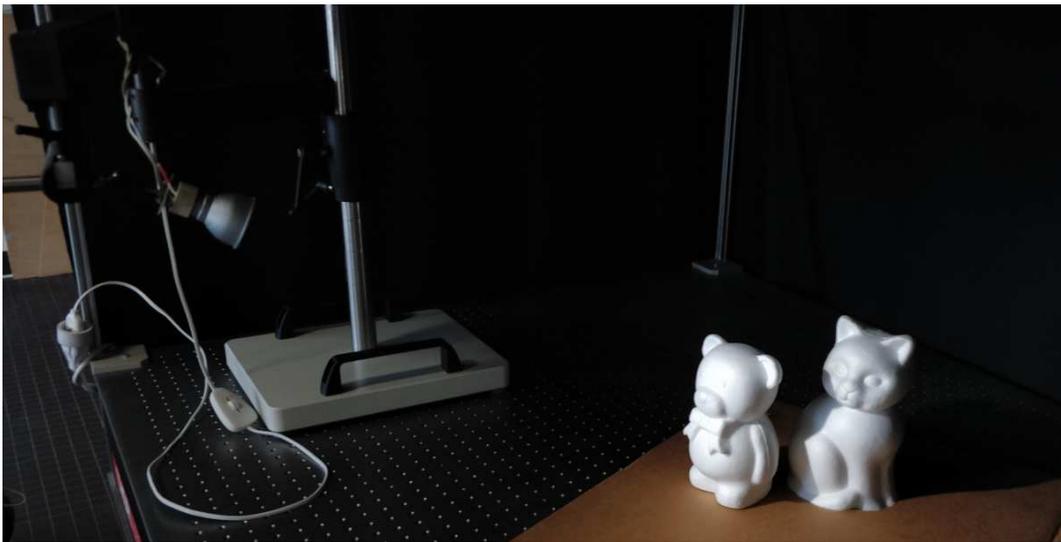
Dans cette partie, nous abordons les biais par rapport à la modélisation de l'irradiance qui peuvent apparaître, tout en respectant les hypothèses formulées sur l'éclairage et le modèle lambertien.

Ombres propres et ombres portées

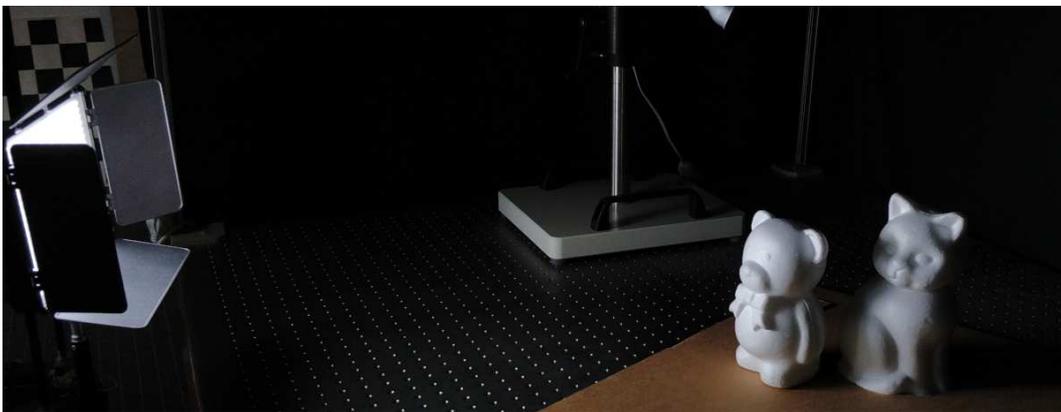
Dans les images prises en compte précédemment, nous avons supposé qu'elles ne comportent aucune ombre. D'après l'équation (1.4), les ombres propres correspondent à des valeurs nulles ou négatives du produit scalaire $\langle \mathbf{N}(u, v), \mathbf{L} \rangle$.

Toutes les ombres ne sont pas nécessairement des ombres propres. Lorsqu'un obstacle bloque la totalité des rayons lumineux, plus aucun objet placé sur la trajectoire des rayons après l'obstacle ne sera directement éclairé. Ainsi, nous retrouvons un ensemble de points \mathbf{X} pour lesquels $\langle \mathbf{N}(u, v), \mathbf{L} \rangle > 0$ mais $I(\mathbf{X}) = 0$. Cet ensemble constitue une ombre portée. La figure 1.4 présente deux situations dans lesquels deux objets, *Chat* et *Ourson*, sont illuminés successivement par une source de lumière ponctuelle (figure 1.2.3(a)) et une source de lumière étendue (figure 1.2.3(b)). Dans la première situation, le terminateur, la limite entre la zone éclairée et celle dans l'ombre (définie par $\langle \mathbf{N}(u, v), \mathbf{L} \rangle = 0$), est plus facile à distinguer par rapport au second cas. Dans les deux exemples proposés, un point \mathbf{X} considéré dans une zone d'ombre peut appartenir à une ombre propre et/ou une ombre portée. Comme ces deux types d'ombre

4. Voir les équations (5) et (6) dans [Tankus & Kiryati 2005].



(a)



(b)

FIGURE 1.4 – Ombres propres et portées observées sur les objets *Chat* et *Ourson* en utilisant une source de lumière ponctuelle (a) puis étendue (b).

produisent le même comportement sur la surface observée, il est peu aisé de distinguer leur origine. Par la suite, nous éliminerons des données utilisées les points \mathbf{X} se trouvant dans les ombres propres et portées.

Avec l'utilisation d'une source de lumière étendue, les points \mathbf{X} présents dans une zone d'ombre n'ont pas nécessairement une intensité mesurée nulle ($I(\mathbf{X}) > 0$). Ces points se trouvent dans la pénombre qui n'est pas modélisée. Pour filtrer ces points aberrants, il existe quelques approches dans la littérature qui permettent de sélectionner les "bonnes" observations (voir par exemple [Barsky & Petrou 2003, Hernández *et al.* 2010]). Toutefois, à notre connaissance, il n'y a pas de solutions efficaces et générales pour éliminer ces points aberrants.

Éclairage direct et indirect

Le modèle décrit par l'équation (1.4) est valide lorsqu'une surface convexe est observée par la caméra. Cependant, lorsque plusieurs objets sont observés simultanément ou bien, que la surface étudiée est concave, chaque point \mathbf{X} de la surface peut diffuser une partie de la lumière reçue à d'autres points de la surface avant que les rayons lumineux soient perçus par la caméra. Nous pouvons parler d'un flux lumineux principal depuis la source de lumière puis d'une somme de flux secondaires lorsque nous considérons l'ensemble des rayons lumineux diffusés par la surface, par l'intermédiaire des réflexions mutuelles. Pour anticiper les conséquences de ces inter-réflexions et en supposant la surface composée d'un ensemble de facettes, le modèle suivant peut être considéré :

$$I(u, v) = \alpha(u, v) \max \{0, \langle \mathbf{N}(u, v), \mathbf{L} \rangle\} + \sum_{\mathbf{X}'} \alpha(u, v) \max \{0, \langle \mathbf{N}(u, v), \mathbf{L}_{\mathbf{X}'}(\mathbf{X}) \rangle\},$$

où \mathbf{X}' sont les centres des facettes visibles depuis le point \mathbf{X} qui, de fait, diffusent de la lumière caractérisée par les vecteurs $\mathbf{L}_{\mathbf{X}'}(\mathbf{X})$ vers \mathbf{X} . Par conséquent, plus un point \mathbf{X} reçoit de flux lumineux secondaires, plus l'intensité $I(\mathbf{X})$ mesurée sera supérieure à celle prédite par le modèle lambertien. Nous pouvons aussi observer cette diffusion de la lumière sur des points de la surface normalement situés dans une zone d'ombre (propre ou portée) : c'est le phénomène de pénombre montré dans la figure 1.4.

Pour diminuer l'impact des réflexions, dans cette thèse, nous avons eu recours à un environnement contrôlé. Les expérimentations ont été réalisées à l'intérieur d'une structure rectangulaire et rigide, intégralement recouverte par du tissu occultant. Cette couverture permet d'absorber la lumière sans générer de reflet parasite (voir la figure 1.4). En imagerie informatique (*computational photography*), les concepts de flux lumineux principal et secondaire correspondent respectivement à un éclairage direct et indirect. En environnement contraint et contrôlé, il est possible de décomposer l'observation d'une scène selon les deux types d'éclairages, par exemple en utilisant l'approche proposée récemment dans [Kubo *et al.* 2018].

1.3 Estimation de la profondeur

Dans cette section, nous nous intéressons à la deuxième étape de la résolution du problème de stéréophotométrie. Le lien entre le vecteur normal à la surface $\mathbf{N}(u, v)$ et le gradient de profondeur $\nabla z(u, v)$ sera mis en évidence dans un premier temps. Enfin, nous présentons quelques méthodes pour réaliser numériquement l'intégration.

1.3.1 Gradients de profondeur

Le lien entre le vecteur normal $\mathbf{N}(x, y)$ et le gradient de profondeur $\nabla z(x, y)$ dépend du modèle de la caméra utilisée.

Projection orthographique

En partant de l'hypothèse d'une projection orthographique et en ignorant la seconde transformation⁵ (voir §1.1.1), nous pouvons écrire :

$$\mathbf{X}(x, y) = \begin{bmatrix} x \\ y \\ z(x, y) \end{bmatrix}.$$

Le vecteur normal $\mathbf{N}(x, y)$ décrit le plan tangent à la surface observée au point $\mathbf{X}(x, y)$. Ce plan est aussi porté par les vecteurs $\partial_x \mathbf{X}(x, y)$ et $\partial_y \mathbf{X}(x, y)$:

$$\partial_x \mathbf{X}(x, y) = \frac{\partial \mathbf{X}(x, y)}{\partial x} = \begin{bmatrix} 1 \\ 0 \\ \frac{\partial z(x, y)}{\partial x} \end{bmatrix}, \quad \partial_y \mathbf{X}(x, y) = \frac{\partial \mathbf{X}(x, y)}{\partial y} = \begin{bmatrix} 0 \\ 1 \\ \frac{\partial z(x, y)}{\partial y} \end{bmatrix}.$$

Le produit vectoriel entre les vecteurs $\partial_x \mathbf{X}(x, y)$ et $\partial_y \mathbf{X}(x, y)$ permet d'obtenir un vecteur colinéaire avec $\mathbf{N}(x, y)$:

$$\partial_x \mathbf{X}(x, y) \times \partial_y \mathbf{X}(x, y) = \begin{bmatrix} -\partial_x z(x, y) \\ -\partial_y z(x, y) \\ 1 \end{bmatrix}$$

Le produit scalaire entre deux vecteurs colinéaires est nul, ce qui nous permet d'écrire :

$$\begin{cases} N_z(x, y) \partial_x z(x, y) = -N_x(x, y) \\ N_z(x, y) \partial_y z(x, y) = -N_y(x, y) \\ N_x(x, y) \partial_y z(x, y) = N_y(x, y) \partial_x z(x, y). \end{cases} \quad (1.12)$$

Nous pouvons éliminer la troisième équation du système (1.12) en supposant $N_z \neq 0$, puisqu'il est possible de la retrouver par combinaison des deux premières équations. Ainsi, les gradients

5. Nous ignorons cette transformation puisque nous supposons que les paramètres internes de la caméra sont connus. De ce fait, la transformation est réversible.

de profondeur selon les directions x et y sont définis par :

$$\nabla z(x, y) = \begin{bmatrix} p(x, y) \\ q(x, y) \end{bmatrix} = \begin{bmatrix} \partial_x z(x, y) \\ \partial_y z(x, y) \end{bmatrix} = \begin{bmatrix} -\frac{N_x(x, y)}{N_z(x, y)} \\ -\frac{N_y(x, y)}{N_z(x, y)} \end{bmatrix}. \quad (1.13)$$

Projection perspective

Nous pouvons réaliser le même développement en partant à présent de l'hypothèse d'une projection perspective (voir §1.1.1) :

$$\mathbf{X}(x, y) = \begin{bmatrix} \frac{z(x, y)x}{f} \\ \frac{z(x, y)y}{f} \\ z(x, y) \end{bmatrix} = z(x, y) \begin{bmatrix} \frac{x}{f} \\ \frac{y}{f} \\ 1 \end{bmatrix}.$$

Le produit vectoriel de $\partial_x \mathbf{X}(x, y)$ et $\partial_y \mathbf{X}(x, y)$ devient :

$$\partial_x \mathbf{X}(x, y) \times \partial_y \mathbf{X}(x, y) = \frac{z(x, y)}{f^2} \begin{bmatrix} -f \partial_x z(x, y) \\ -f \partial_y z(x, y) \\ z(x, y) + x \partial_x z(x, y) + y \partial_y z(x, y) \end{bmatrix}.$$

Ce vecteur est à nouveau colinéaire avec le vecteur normal à la surface \mathbf{N} . Le produit vectoriel entre ces deux vecteurs est nul :

$$\begin{cases} fN_z(x, y)\partial_x z(x, y) + N_x(x, y) [z(x, y) + x\partial_x z(x, y) + y\partial_y z(x, y)] = 0 \\ fN_z(x, y)\partial_y z(x, y) + N_y(x, y) [z(x, y) + x\partial_x z(x, y) + y\partial_y z(x, y)] = 0 \\ N_y(x, y)\partial_x z(x, y) - N_x(x, y)\partial_y z(x, y) = 0. \end{cases} \quad (1.14)$$

Avant de chercher à résoudre le système d'équations (1.14), nous allons considérer le changement de variable $\check{z} = \ln z$. De cette façon, il est possible de réécrire le système d'équations (1.14) sous forme linéaire en $\partial_x \check{z}(x, y)$ et $\partial_y \check{z}(x, y)$:

$$\begin{cases} [fN_z(x, y) + xN_x(x, y)] \partial_x \check{z}(x, y) + yN_x(x, y) \partial_y \check{z}(x, y) = -N_x(x, y) \\ [fN_z(x, y) + yN_y(x, y)] \partial_y \check{z}(x, y) + xN_y(x, y) \partial_x \check{z}(x, y) = -N_y(x, y) \\ N_y(x, y) \partial_x \check{z}(x, y) - N_x(x, y) \partial_y \check{z}(x, y) = 0. \end{cases}$$

Finalement, nous pouvons exprimer les gradients de profondeur, sous l'hypothèse de la projection perspective, en considérant le changement de variable, de la manière suivante :

$$\nabla \check{z}(x, y) = \begin{bmatrix} \check{p}(x, y) \\ \check{q}(x, y) \end{bmatrix} = \begin{bmatrix} -\frac{N_x(x, y)}{xN_x(x, y) + yN_y(x, y) + fN_z(x, y)} \\ -\frac{N_y(x, y)}{xN_x(x, y) + yN_y(x, y) + fN_z(x, y)} \end{bmatrix}. \quad (1.15)$$

Dans cette situation, la phase d'intégration des gradients permet de retrouver $\check{z}(x, y)$. Après intégration, et puisque par convention $z(x, y) > 0$ (voir la figure 1.2), nous pouvons faire le changement de variable inverse $z(x, y) = \exp \check{z}(x, y)$. Ce dernier changement de variable intervient aussi lorsque nous utilisons l'approche présentée dans la partie § 1.2.2.

1.3.2 Intégration des gradients de profondeur

Pour la projection orthographique et à partir de l'équation (1.13), nous pouvons écrire une première solution pour l'intégration des gradients de profondeur afin d'obtenir $z(x, y)$:

$$z(x, y) = z(x_0, y_0) + \int_{(r,s)}^{(x,y)} [p(r, s) dr + q(r, s) ds],$$

avec $z(x_0, y_0)$ la profondeur connue au point (x_0, y_0) . Nous obtenons $z(x, y)$ à une constante additive près. Si la profondeur n'est pas connue mais que la vérité terrain est disponible, la constante peut être remplacée par une valeur numérique dans le but de réduire l'écart quadratique moyen en $z(x, y)$.

De façon similaire, pour la projection perspective et à partir de l'équation (1.15), nous pouvons à nouveau écrire une première solution pour l'intégration des gradients de profondeur afin d'obtenir $\check{z}(x, y)$:

$$\check{z}(x, y) = \check{z}(x_0, y_0) + \int_{(r,s)}^{(x,y)} [\check{p}(r, s) dr + \check{q}(r, s) ds].$$

Ce qui nous permet de déduire l'expression suivante pour retrouver $z(x, y)$:

$$z(x, y) = z(x_0, y_0) \exp \left\{ \int_{(r,s)}^{(x,y)} [\check{p}(r, s) dr + \check{q}(r, s) ds] \right\}.$$

Cette approche permet d'obtenir une estimation de z à une constante multiplicative près. Elle requiert aussi de connaître les paramètres internes de la caméra (comme pour la phase d'estimation des normales) ainsi que la profondeur au point (x_0, y_0) .

Dans [Horn & Brooks 1986], afin de fournir une solution aux équations d'intégration introduites précédemment, les auteurs proposent de résoudre l'équation suivante⁶, paramétrée dans le plan image :

$$z_{u-1,v} + z_{u+1,v} + z_{u,v-1} + z_{u,v+1} - 4z_{u,v} = \frac{p_{u+1,v} - p_{u-1,v}}{2} + \frac{q_{u,v+1} - q_{u,v-1}}{2}. \quad (1.16)$$

Il est ensuite proposé de résoudre l'équation (1.16) avec un algorithme itératif en utilisant quatre voisins les plus proches du point (u, v) considéré :

$$z_{u,v}^{(k+1)} = \frac{z_{u-1,v}^{(k)} + z_{u+1,v}^{(k)} + z_{u,v-1}^{(k)} + z_{u,v+1}^{(k)}}{4} - \frac{p_{u+1,v} - p_{u-1,v}}{8} - \frac{q_{u,v+1} - q_{u,v-1}}{8}.$$

Depuis la méthode proposée par Horn et Brooks, différentes approches ont été développées afin de réaliser l'intégration des gradients de profondeur et ainsi obtenir la géométrie de la surface (voir [Ackermann & Goesele 2015]). Dans [Quéau *et al.* 2018a], les auteurs rappellent 31 méthodes pour résoudre le problème de l'intégration des gradients et proposent une comparaison reposant sur cinq critères : la rapidité d'exécution, la robustesse au bruit dans les gradients de profondeur, la conservation des discontinuités de profondeur, la prise en compte de domaine de reconstruction non rectangulaire⁷ et l'absence ou non de contrainte au bord du domaine de reconstruction. En se concentrant cette fois sur les méthodes variationnelles, ces mêmes auteurs prolongent les comparaisons dans [Quéau *et al.* 2018b]. Ce type d'approche modélise le problème d'intégration sous la forme d'une fonction de coût qu'il faut minimiser pour retrouver la profondeur.

Par rapport aux méthodes itératives, l'approche algébrique proposée dans [Harker & O'Leary 2008] propose directement une solution unique. En considérant des gradients de profondeur perturbés par un bruit gaussien et une caméra orthographique, les auteurs proposent de résoudre le problème d'intégration au moyen d'une fonction de coût exprimée sous la forme d'une équation de Sylvester. À partir de cette équation matricielle, une unique solution au sens des moindres carrés est obtenue. Dans [Harker & O'Leary 2015], les mêmes auteurs apportent davantage de détails sur la méthode, notamment en proposant des solutions pour assouplir les contraintes au niveau du bord du domaine de reconstruction.

6. Dans le cas d'une projection perspective, p et q sont remplacés par \check{p} et \check{q} , respectivement.

7. Un domaine de reconstruction non rectangulaire peut apparaître après la segmentation de la surface observée, dans le but d'éliminer l'arrière-plan et les discontinuités de profondeur associées.

1.4 Jeux de données

Dans cette section, nous introduisons quatre jeux de données disponibles publiquement⁸ qui peuvent être employés dans l'évaluation d'une méthode en cours de développement, que ce soit pour l'estimation des normales ou pour l'intégration des gradients de profondeur.

Toutes les données présentées ont été obtenues à partir de caméras observant des objets réels, dans des conditions contrôlées (pas d'éclairage naturel ou ambiant, directions d'éclairage connues ou étalonnées).

Le jeu de données *Light Stage Data Gallery* propose 6 séquences d'observations de 4 objets différents placés au centre d'un grand dôme constitué d'un ensemble d'éclairages ponctuels (LED) pilotables individuellement. Ce dispositif, davantage détaillé dans [Chabert *et al.* 2006], permet d'obtenir une diversité importante des conditions d'illumination ainsi qu'une connaissance précise de la direction d'éclairage, par construction. Initialement destiné à la communauté *computer graphics*, ce jeu de données fournit assez d'éléments pour servir de point de comparaison qualitative, pour l'estimation des normales des surfaces des objets observés : directions des éclairages et paramètres internes de la caméra.

Gourd&Apple est le jeu de données qui accompagne la méthode proposée dans [Aldrin *et al.* 2008]. Bien que les paramètres de la caméra et la vérité terrain ne soient pas disponibles, le grand nombre d'observations des objets ainsi que les matériaux qui les composent offrent un intérêt particulier. Les surfaces observées sont peu brillantes, proches de l'hypothèse lambertienne et leur albédo n'est pas spatialement homogène.

Plus récemment, le jeu de données *DiLiGenT* a été proposé dans [Shi *et al.* 2016] puis, davantage détaillé dans [Shi *et al.* 2019]. Le principal objectif est de servir de banc d'essai pour comparer les méthodes qui s'attaquent au problème de stéréophotométrie non calibrée, c'est-à-dire, lorsque les directions des éclairages ne sont pas connues, ainsi qu'avec des surfaces qui ne respectent l'hypothèse lambertienne. En plus du grand nombre de conditions d'éclairage, la diversité des surfaces (mate, brillante, spatialement non homogène) observées permet d'évaluer les méthodes dans beaucoup de situations. Les vecteurs normaux des surfaces observées sont disponibles avec les directions d'éclairages étalonnées ainsi que les paramètres internes de la caméra employée. En accompagnement du jeu de données principal, les auteurs proposent un jeu complémentaire contenant des observations des mêmes objets depuis un point de vue différent mais sans fournir la vérité terrain. Nous utiliserons le jeu de données principal dans le chapitre 2.

8. Les liens vers les jeux de données sont dans la bibliographie, en complément des publications auxquelles ils sont rattachés.



FIGURE 1.5 – Ensemble des cinq objets observés depuis 20 points de vue différents dans le jeu de données *DiLiGenT-MV* présenté dans [Li *et al.* 2020].

Jeu de données	Light Stage Data Gallery	Gourd&Apple	DiLiGent	DiLiGent-MV
Année	2006	2008	2019	2020
# séqu. d'images	6	3	10	5
# points de vue	1/2	1	1	20
# éclairages	253	102/112	96	96
Vérité terrain	✗	✗	normales	normales, maillage 3D
Type de surface	mh	mrh	mrhH	mrH
Param. internes caméra connus	✓	✗	✓	✓

TABLEAU 1.1 – Caractéristiques de quelques jeux de données accessibles publiquement pour la comparaison des méthodes d'estimation des normales et la reconstruction 3D par stéréophotométrie. Légende de la septième ligne : m : mate, r : réfléchissante, h : homogène, H : hétérogène

Dernièrement, le jeu de données *DiGiLenT-MV* a été présenté dans [Li *et al.* 2020] et offre des observations de cinq objets différents selon 96 conditions d'illumination qui ont été répétées depuis vingt points de vue différents. L'ensemble des objets et des points de vue est visible dans la figure 1.5. L'objectif de ce jeu de données est de combiner les méthodes de reconstruction par stéréophotométrie avec celles qui exploitent un ensemble de points de vue différents (par exemple, pour guider la phase d'intégration en fournissant des points 3D, voir [Smith & Fang 2016]). La vérité terrain fournie détaille les vecteurs normaux de la surface ainsi qu'un maillage 3D qui permet de mettre en évidence les discontinuités de profondeur. Il est ainsi possible de comparer quantitativement les résultats obtenus à l'aide de différentes méthodes depuis les images de ce jeu de données.

Le Tableau 1.1 rappelle et présente quelques propriétés de ces quatre jeux de données. D'autres jeux de données accompagnant des méthodes présentes dans la littérature existent, mais nous nous sommes limités à ces quatre jeux de données puisqu'ils semblent adaptés au développement et à la comparaison de méthodes d'estimation des normales voire d'intégration des gradients. Néanmoins, toutes les images acquises ont été obtenues depuis une caméra pers-

pective. Dans la suite de ce manuscrit, nous nous intéresserons à des capteurs offrant un champ de vue plus important et, afin d'illustrer le fonctionnement des méthodes proposées à partir de données réelles, nous aurons besoin de données issues de ce type de capteur. À notre connaissance, il n'existe pas de jeu de données construit à partir d'un capteur panoramique et adapté pour la stéréophotométrie.

Conclusion

Dans ce premier chapitre, nous avons rappelé les modèles géométrique et photométrique utilisés pour la formation des images numériques à partir de caméras orthographique et perspective. Nous avons présenté la stéréophotométrie usuelle qui permet de retrouver la géométrie d'une surface statique à partir d'une série d'observations, réalisée par une caméra immobile, sous différentes conditions d'éclairages. Nous avons abordé les deux étapes successives qui composent cette technique : l'estimation des normales de la surface observée, à partir d'une séquence d'images, ainsi que l'intégration des gradients de profondeur, déduits à partir des vecteurs normaux estimés à l'étape précédente. Nous avons conclu ce chapitre par la présentation de quelques jeux de données, disponibles publiquement et présentés dans la littérature, qui facilitent la comparaison des méthodes actuelles et le développement des futures approches. Cependant, ces jeux de données complets et diversifiés ayant tous été obtenus à partir de caméras perspectives, il sera nécessaire de construire de nouveaux jeux de données acquis à partir de capteurs de vision omnidirectionnelle, afin de valider expérimentalement les contributions des chapitres 2 et 3.

Stéréophotométrie pour les capteurs à projection centrale

Contenu du chapitre

2.1	Modèle unifié pour les capteurs à projection centrale	24
2.1.1	Présentation du modèle	24
2.1.2	Représentation d'une image sphérique	25
2.2	Estimation des gradients orientés	26
2.2.1	Équation de l'irradiance pour les images sphériques	26
2.2.2	Rapports photométriques	29
2.3	Estimation des distances radiales	30
2.3.1	Résolution adaptée aux images panoramiques	30
2.4	Expérimentations	32
2.4.1	Validation en simulation	33
2.4.2	Évaluation en conditions réelles et contrôlées	38

Résumé

Ce chapitre introduit brièvement la théorie unifiée des capteurs à projection centrale. Une nouvelle équation de l'irradiance pour les images sphériques est proposée avec une chaîne de traitements dédiée. Les propriétés géométriques et les limites de la méthode originale sont étudiées à l'aide d'expérimentations reposant sur des simulations numériques et des images réelles.

2.1 Modèle unifié pour les capteurs à projection centrale

Dans cette section, nous nous concentrons sur le modèle unifié : nous détaillerons ses étapes et ses améliorations. Nous introduirons un système de coordonnées sphériques qui permet d'exploiter des images sphériques lorsque nous nous limitons à la première étape de projection.

2.1.1 Présentation du modèle

Cette partie présente le modèle unifié pour les capteurs à projection centrale tel que proposé par les auteurs dans [Geyer & Daniilidis 2000, Geyer & Daniilidis 2001]. Le théorème 1 dans [Geyer & Daniilidis 2000] lie la projection catadioptrique avec un unique point de vue à un ensemble de deux projections : une première projection sur une sphère d'un rayon connu (souvent unitaire) suivie par une projection sur un plan passant par un point. Nous considérons une sphère unitaire centrée à l'origine du repère monde $\{O; x, y, z\}$ ainsi que le plan image $z = -m$ (voir la figure 2.1). Un point 3D $X = [x, y, z]^T$ est dans un premier temps projeté aux deux antipodes Q et Q' ayant pour coordonnées $\pm X/\|X\| \triangleq \pm [x_s, y_s, z_s]^T$ sur la sphère unitaire. Pour la seconde étape de la transformation, nous déterminons la projection perspective depuis la sphère unitaire vers le plan image $z = -m$ en passant par le point $[0, 0, \ell]^T$ sur l'axe z de la sphère (qui correspond au point O' dans la figure 2.1). En se limitant au point Q , la projection

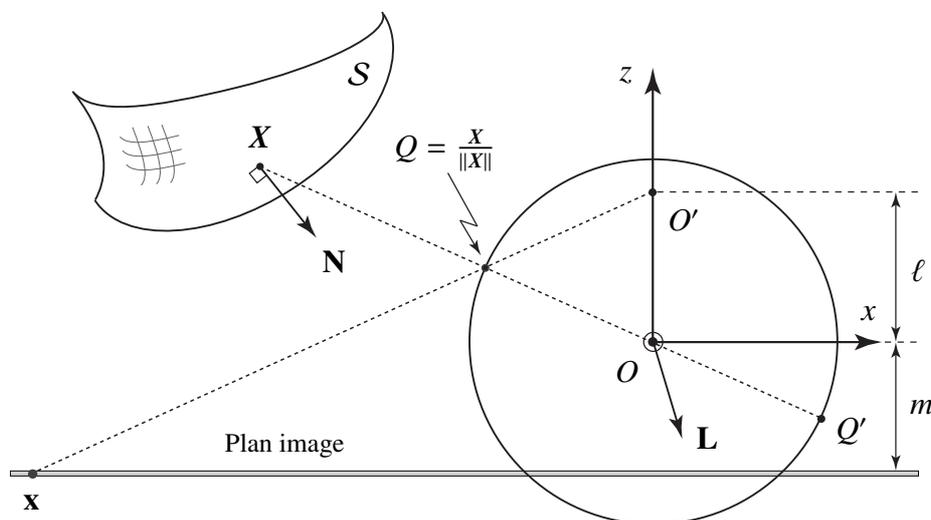


FIGURE 2.1 – Un point 3D $X = [x, y, z]^T$ est projeté aux deux antipodes $\pm X/\|X\|$ sur la sphère unitaire [Geyer & Daniilidis 2000, Sect. 2.1]. Ces points Q et Q' sont projetés sur le plan image $z = -m$ à l'aide d'une projection passant par le point O' (pour conserver la clarté du schéma, seule la projection de Q vers x sur le plan image est représentée dans la figure). Le vecteur normal à la surface S au point X est noté N . L correspond au vecteur directeur de l'éclairage.

du point X sur le plan image est définie par :

$$\mathbf{x} = \left[\frac{(\ell + m)x}{\ell \|\mathbf{X}\| - z}, \quad \frac{(\ell + m)y}{\ell \|\mathbf{X}\| - z}, \quad -m \right]^T. \quad (2.1)$$

Ces projections peuvent être décrites à l'aide de miroirs à section conique (à savoir, des miroirs paraboliques, hyperboliques et elliptiques). Dans cette situation, les paramètres ℓ et m , systématiquement positifs ou nuls, ont une signification physique : ℓ correspond à une fonction de l'excentricité de la conique, et m est une fonction à la fois de l'excentricité et de l'échelle. Lorsque le point de projection O' est situé au pôle Nord de la sphère unitaire (c'est-à-dire, lorsque $\ell = 1$ et $m = 0$), nous obtenons une projection stéréographique comme expliqué dans [Coxeter 1969, Sect. 6.9]. À l'inverse, lorsque $\ell = 0$ et $m = 1$, nous obtenons une projection perspective.

2.1.2 Représentation d'une image sphérique

Pour représenter une image sur la sphère unitaire, nous utilisons le système de coordonnées sphériques (ρ, θ, φ) [Weber & Arfken 2003, Sect. 2.5] :

$$\begin{cases} x = \rho \sin \theta \cos \varphi, \\ y = \rho \sin \theta \sin \varphi, \\ z = \rho \cos \theta, \end{cases}$$

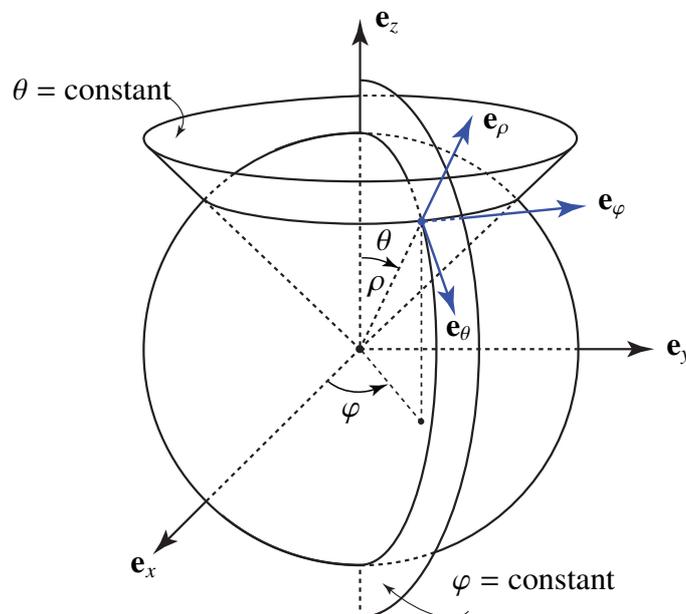


FIGURE 2.2 – Coordonnées sphériques (ρ, θ, φ) et surfaces du système de coordonnées sphériques.

dans lequel $\rho \in [0, \infty)$ est la distance radiale, $\theta \in (0, \pi)$ la colatitute et $\varphi \in [0, 2\pi)$ la longitude (voir la figure 2.2). Pour l'ensemble de points ayant $\rho = 0$, les angles θ et φ ne sont pas définis. Les surfaces dont les coordonnées θ sont constantes constituent des cônes qui partagent le même sommet situé à l'origine de la sphère unitaire, et dont l'axe de révolution correspond à l'axe z . Les surfaces dont les coordonnées φ sont constantes caractérisent des demi-plans verticaux passant par l'origine (voir la figure 2.2). En considérant $\mathbf{r} = [x, y, z]^T$, l'ensemble des vecteurs unitaires $\{\mathbf{e}_\rho, \mathbf{e}_\theta, \mathbf{e}_\varphi\}$ qui constitue la base du système de coordonnées sphériques est défini par [Weber & Arfken 2003, Sect. 2.3] :

$$\mathbf{e}_\rho = \frac{1}{h_\rho} \frac{\partial \mathbf{r}}{\partial \rho} = \begin{bmatrix} \sin \theta \cos \varphi \\ \sin \theta \sin \varphi \\ \cos \theta \end{bmatrix}, \quad \mathbf{e}_\theta = \frac{1}{h_\theta} \frac{\partial \mathbf{r}}{\partial \theta} = \begin{bmatrix} \cos \theta \cos \varphi \\ \cos \theta \sin \varphi \\ -\sin \theta \end{bmatrix}, \quad \mathbf{e}_\varphi = \frac{1}{h_\varphi} \frac{\partial \mathbf{r}}{\partial \varphi} = \begin{bmatrix} -\sin \varphi \\ \cos \varphi \\ 0 \end{bmatrix},$$

où les facteurs d'échelle (aussi appelés coefficients métriques ou de Lamé) sont :

$$(h_\rho, h_\theta, h_\varphi) = \left(\left\| \frac{\partial \mathbf{r}}{\partial \rho} \right\|, \left\| \frac{\partial \mathbf{r}}{\partial \theta} \right\|, \left\| \frac{\partial \mathbf{r}}{\partial \varphi} \right\| \right) = (1, \rho, \rho \sin \theta).$$

Nous pouvons remarquer que la direction des vecteurs unitaires \mathbf{e}_ρ , \mathbf{e}_θ et \mathbf{e}_φ varie en fonction des angles θ et φ .

2.2 Estimation des gradients orientés

Dans cette section, nous proposons une nouvelle équation de l'irradiance qui tient compte du modèle unifié présenté dans la section 2.1.

2.2.1 Équation de l'irradiance pour les images sphériques

Pour rappel, la dérivée directionnelle d'une fonction scalaire $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ selon le vecteur $\mathbf{w} \in \mathbb{R}^n$ correspond à la fonction $\nabla_{\mathbf{w}} f(\mathbf{x})$ définie par la limite $\nabla_{\mathbf{w}} f(\mathbf{x}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{w}) - f(\mathbf{x})}{h}$. Si f est différentiable en \mathbf{x} , alors la dérivée directionnelle suivant n'importe quel vecteur \mathbf{w} existe : $\nabla_{\mathbf{w}} f(\mathbf{x}) = \langle \nabla f(\mathbf{x}), \mathbf{w} \rangle$. Avec les éléments présentés depuis le début de ce chapitre, nous pouvons introduire le théorème suivant :

Théorème 2 (Équation de l'irradiance adaptée aux images sphériques). *Nous supposons que les deux vecteurs $\mathbf{v} = [v_1, v_2]^T$ et $\mathbf{w} = [w_1, w_2]^T$ appartiennent au plan tangent à la sphère \mathbb{S}^2 au point \mathbf{r} . Nous ajoutons que $\det(\mathbf{A}) > 0$ avec $\mathbf{A} = [\mathbf{v}, \mathbf{w}]$. D'après les hypothèses formulées,*

L'équation de l'irradiance adaptée aux images sphériques est,

$$I(\theta, \varphi) = \frac{\alpha(\theta, \varphi) \mathbf{L}^T}{\|\mathbf{L}\| \sqrt{(v_1 p - w_1 q)^2 + \left(\frac{w_2 q - v_2 p}{\sin \theta}\right)^2 + \det(\mathbf{A})^2}} \cdot \mathbf{R}_z(\varphi) \mathbf{R}_y(\theta) \begin{bmatrix} v_1 p - w_1 q \\ \frac{w_2 q - v_2 p}{\sin \theta} \\ -\det(\mathbf{A}) \end{bmatrix}, \quad (2.2)$$

où $\alpha(\theta, \varphi) \in [0, 1]$ est l'albedo au point (θ, φ) ,

$$p \triangleq \frac{1}{\rho} \nabla_{\mathbf{w}} \rho = \nabla_{\mathbf{w}} \ln \rho, \quad q \triangleq \frac{1}{\rho} \nabla_{\mathbf{v}} \rho = \nabla_{\mathbf{v}} \ln \rho, \quad (2.3)$$

et $\nabla_{\mathbf{w}} \rho = \langle \nabla \rho, \mathbf{w} \rangle$ correspond à la dérivée directionnelle de ρ suivant le vecteur \mathbf{w} .

Démonstration

La preuve de ce théorème suit la démarche du théorème 1 proposé dans [Tankus *et al.* 2005]. En considérant les vecteurs directeurs $\mathbf{v}, \mathbf{w} \in \mathbb{T}_{\mathbf{r}} \mathbb{S}^2$, le vecteur normal à la surface \mathbf{N} peut être estimé de la façon suivante (voir la figure 2.1) :

$$\begin{aligned} \mathbf{N}(\theta, \varphi) &= \left(v_1 \frac{\partial \mathbf{r}}{\partial \varphi} + v_2 \frac{\partial \mathbf{r}}{\partial \theta} \right) \times \left(w_1 \frac{\partial \mathbf{r}}{\partial \varphi} + w_2 \frac{\partial \mathbf{r}}{\partial \theta} \right) \\ &= \left[\left(v_1 \frac{\partial \rho}{\partial \varphi} + v_2 \frac{\partial \rho}{\partial \theta} \right) \mathbf{e}_\rho + v_1 h_\varphi \mathbf{e}_\varphi + v_2 h_\theta \mathbf{e}_\theta \right] \times \\ &\quad \left[\left(w_1 \frac{\partial \rho}{\partial \varphi} + w_2 \frac{\partial \rho}{\partial \theta} \right) \mathbf{e}_\rho + w_1 h_\varphi \mathbf{e}_\varphi + w_2 h_\theta \mathbf{e}_\theta \right] \\ &= \rho \left[\sin \theta (v_1 \nabla_{\mathbf{w}} \rho - w_1 \nabla_{\mathbf{v}} \rho) \mathbf{e}_\theta + (w_2 \nabla_{\mathbf{v}} \rho - v_2 \nabla_{\mathbf{w}} \rho) \mathbf{e}_\varphi - \rho \sin \theta \det(\mathbf{A}) \mathbf{e}_\rho \right], \end{aligned} \quad (2.4)$$

où la dernière égalité est obtenue à l'aide de la définition de dérivée directionnelle et des identités suivantes :

$$\mathbf{e}_\rho \times \mathbf{e}_\rho = \mathbf{0}, \mathbf{e}_\varphi \times \mathbf{e}_\varphi = \mathbf{0}, \mathbf{e}_\theta \times \mathbf{e}_\theta = \mathbf{0}, \mathbf{e}_\rho \times \mathbf{e}_\theta = \mathbf{e}_\varphi, \mathbf{e}_\varphi \times \mathbf{e}_\rho = \mathbf{e}_\theta, \mathbf{e}_\theta \times \mathbf{e}_\varphi = \mathbf{e}_\rho.$$

L'équation (2.4) peut être simplifiée :

$$\mathbf{N}(\theta, \varphi) = \rho \sin \theta \mathbf{R}_z(\varphi) \mathbf{R}_y(\theta) \begin{bmatrix} v_1 \nabla_{\mathbf{w}} \rho - w_1 \nabla_{\mathbf{v}} \rho \\ \frac{1}{\sin \theta} (w_2 \nabla_{\mathbf{v}} \rho - v_2 \nabla_{\mathbf{w}} \rho) \\ -\det(\mathbf{A}) \rho \end{bmatrix},$$

où $\mathbf{R}_z(\varphi) \mathbf{R}_y(\theta) = [\mathbf{e}_\theta, \mathbf{e}_\varphi, \mathbf{e}_\rho] \in \text{SO}(3)$. Pour tout vecteur défini dans $\{\mathbf{e}_\rho, \mathbf{e}_\theta, \mathbf{e}_\varphi\}$, la matrice de rotation $\mathbf{R}_z(\varphi) \mathbf{R}_y(\theta)$ nous permet de l'exprimer dans $\{\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z\}$, l'ensemble des vecteurs qui

constitue la base du système de coordonnées cartésiennes (voir la figure 2.2). Étant donné que pour $\mathbf{R} \in \text{SO}(3)$, $\|\mathbf{R}\mathbf{x}\| = \|\mathbf{x}\|$, $\forall \mathbf{x} \in \mathbb{R}^3$, nous obtenons :

$$\widehat{\mathbf{N}}(\theta, \varphi) = \frac{\mathbf{N}(\theta, \varphi)}{\|\mathbf{N}(\theta, \varphi)\|} = \frac{1}{\sqrt{(v_1 \nabla_{\mathbf{w}} \rho - w_1 \nabla_{\mathbf{v}} \rho)^2 + \left(\frac{w_2 \nabla_{\mathbf{v}} \rho - v_2 \nabla_{\mathbf{w}} \rho}{\sin \theta}\right)^2 + \det(\mathbf{A})^2 \rho^2}} \mathbf{R}_z(\varphi) \mathbf{R}_y(\theta) \begin{bmatrix} v_1 \nabla_{\mathbf{w}} \rho - w_1 \nabla_{\mathbf{v}} \rho \\ \frac{1}{\sin \theta} (w_2 \nabla_{\mathbf{v}} \rho - v_2 \nabla_{\mathbf{w}} \rho) \\ -\det(\mathbf{A}) \rho \end{bmatrix}.$$

Pour une scène lambertienne illuminée depuis la direction $\mathbf{L} = [p_d, q_d, -1]^T$ par une source ponctuelle placée à l'infini, l'équation de l'irradiance pour les vecteurs directeurs \mathbf{v} et \mathbf{w} est :

$$I(\theta, \varphi) = \alpha(\theta, \varphi) \langle \widehat{\mathbf{N}}(\theta, \varphi), \widehat{\mathbf{L}} \rangle = \frac{\alpha(\theta, \varphi) \mathbf{L}^T}{\|\mathbf{L}\| \sqrt{(v_1 \nabla_{\mathbf{w}} \rho - w_1 \nabla_{\mathbf{v}} \rho)^2 + \left(\frac{w_2 \nabla_{\mathbf{v}} \rho - v_2 \nabla_{\mathbf{w}} \rho}{\sin \theta}\right)^2 + \det(\mathbf{A})^2 \rho^2}} \mathbf{R}_z(\varphi) \mathbf{R}_y(\theta) \begin{bmatrix} v_1 \nabla_{\mathbf{w}} \rho - w_1 \nabla_{\mathbf{v}} \rho \\ \frac{1}{\sin \theta} (w_2 \nabla_{\mathbf{v}} \rho - v_2 \nabla_{\mathbf{w}} \rho) \\ -\det(\mathbf{A}) \rho \end{bmatrix}, \quad (2.5)$$

où $\alpha(\theta, \varphi)$ est l'albedo au point (θ, φ) . À partir de l'équation (2.3), nous obtenons $\nabla_{\mathbf{v}} \rho = q\rho$ et $\nabla_{\mathbf{w}} \rho = p\rho$. La distance radiale ρ est éliminée par substitution des expressions précédentes dans (2.5). Finalement, après réorganisation, l'équation (2.2) est obtenue. ■

Si $\{\mathbf{v}, \mathbf{w}\}$ est une base naturelle, c'est-à-dire $\mathbf{v} = [1, 0]^T$ et $\mathbf{w} = [0, 1]^T$, alors (2.2) se simplifie,

$$I(\theta, \varphi) = \frac{\alpha(\theta, \varphi) \mathbf{L}^T}{\|\mathbf{L}\| \sqrt{p^2 + \left(\frac{q}{\sin \theta}\right)^2 + 1}} \mathbf{R}_z(\varphi) \mathbf{R}_y(\theta) \begin{bmatrix} p \\ \frac{q}{\sin \theta} \\ -1 \end{bmatrix}, \quad (2.6)$$

où

$$p \triangleq \frac{1}{\rho} \frac{\partial \rho}{\partial \theta} = \frac{\partial \ln \rho}{\partial \theta}, \quad q \triangleq \frac{1}{\rho} \frac{\partial \rho}{\partial \varphi} = \frac{\partial \ln \rho}{\partial \varphi}. \quad (2.7)$$

L'équation de l'irradiance adaptée aux images sphériques (2.2) dépend uniquement des dérivées directionnelles de $\ln \rho(\theta, \varphi)$ en suivant les vecteurs \mathbf{v} et \mathbf{w} . Elle ne dépend pas de $\ln \rho(\theta, \varphi)$ et, de ce fait, le problème d'estimation de la distance radiale $\rho(\theta, \varphi)$ à partir de l'équation de l'irradiance se transforme en un problème d'estimation de l'expression $\ln \rho(\theta, \varphi)$ à partir de l'équation (2.2). Estimer $\ln \rho(\theta, \varphi)$ revient à estimer $\rho(\theta, \varphi) = \exp(\ln \rho(\theta, \varphi))$ puisque le logarithme népérien est une fonction bijective et $\rho(\theta, \varphi)$ est toujours strictement positif. Ce problème sera résolu dans la section 2.3.

2.2.2 Rapports photométriques

À partir de l'équation (2.2), nous cherchons à obtenir les valeurs de p et q . Les images ainsi que le vecteur de direction de la direction d'éclairage sont respectivement notés, $I_i(\theta, \varphi)$ et

$$\mathbf{L}_i \triangleq [p_{d_i}, q_{d_i}, -1]^T, \quad i \in \{0, 1, \dots, n-1\}.$$

Nous pouvons ainsi définir l'équation de l'irradiance pour la i -ième image :

$$I_i(\theta, \varphi) = \frac{\alpha(\theta, \varphi) \mathbf{L}_i^T}{\|\mathbf{L}_i\| \sqrt{(v_1 p - w_1 q)^2 + \left(\frac{w_2 q - v_2 p}{\sin \theta}\right)^2 + \det(\mathbf{A})^2}} \mathbf{R}_z(\varphi) \mathbf{R}_y(\theta) \begin{bmatrix} v_1 p - w_1 q \\ \frac{w_2 q - v_2 p}{\sin \theta} \\ -\det(\mathbf{A}) \end{bmatrix}.$$

En supposant que les intensités ne soient pas nulles, nous divisons la i -ième par la k -ième pour obtenir :

$$\frac{I_i(\theta, \varphi)}{I_k(\theta, \varphi)} = \frac{\|\mathbf{L}_k\| \mathbf{L}_i^T \mathbf{R}_z(\varphi) \mathbf{R}_y(\theta) \left[v_1 p - w_1 q, \frac{w_2 q - v_2 p}{\sin \theta}, -\det(\mathbf{A}) \right]^T}{\|\mathbf{L}_i\| \mathbf{L}_k^T \mathbf{R}_z(\varphi) \mathbf{R}_y(\theta) \left[v_1 p - w_1 q, \frac{w_2 q - v_2 p}{\sin \theta}, -\det(\mathbf{A}) \right]^T}. \quad (2.8)$$

Le rapport photométrique (2.8) est réécrit après développement du numérateur et du dénominateur. En rassemblant les termes contenant p et q , l'expression devient,

$$A_{i,k} p + B_{i,k} q + C_{i,k} = 0, \quad i, k \in \{0, 1, \dots, n-1\}, \quad (2.9)$$

où

$$A_{i,k} = \left[I_i \|\mathbf{L}_i\| \langle \mathbf{L}_k, \mathbf{e}_\theta \rangle - I_k \|\mathbf{L}_k\| \langle \mathbf{L}_i, \mathbf{e}_\theta \rangle, -\frac{1}{\sin \theta} (I_i \|\mathbf{L}_i\| \langle \mathbf{L}_k, \mathbf{e}_\varphi \rangle - I_k \|\mathbf{L}_k\| \langle \mathbf{L}_i, \mathbf{e}_\varphi \rangle) \right] \mathbf{v},$$

$$B_{i,k} = -\left[I_i \|\mathbf{L}_i\| \langle \mathbf{L}_k, \mathbf{e}_\theta \rangle - I_k \|\mathbf{L}_k\| \langle \mathbf{L}_i, \mathbf{e}_\theta \rangle, -\frac{1}{\sin \theta} (I_i \|\mathbf{L}_i\| \langle \mathbf{L}_k, \mathbf{e}_\varphi \rangle - I_k \|\mathbf{L}_k\| \langle \mathbf{L}_i, \mathbf{e}_\varphi \rangle) \right] \mathbf{w},$$

$$C_{i,k} = -\det(\mathbf{A}) (I_i \|\mathbf{L}_i\| \langle \mathbf{L}_k, \mathbf{e}_\rho \rangle - I_k \|\mathbf{L}_k\| \langle \mathbf{L}_i, \mathbf{e}_\rho \rangle),$$

et I_i est une notation compacte pour $I_i(\theta, \varphi)$. Puisque chaque équation du système nécessite deux images, nous avons besoin d'au moins trois images pour trouver une solution aux inconnues p et q (dans des conditions idéales). En utilisant plus de trois images, le nombre d'équation ne dépasse pas $N_{\text{eq}} = \binom{n}{2} = n(n-1)/2$.

Estimation de l'albedo

L'équation (2.8) ne dépend pas de l'albedo $\alpha(\theta, \varphi)$: la construction du rapport photométrique a permis de l'éliminer. L'approche proposée estime les gradients orientés (p, q) sans

tenir compte de l'albedo. En exploitant l'équation de l'irradiance (2.2), estimer l'albedo revient à résoudre un problème de rendu inversé dont une solution peut être l'équation (17) dans [Smith & Fang 2016, Sect 4.1]¹ :

$$\alpha(\theta, \varphi) = \frac{\sum_{i=0}^{n-1} \langle \widehat{\mathbf{N}}(\theta, \varphi), \widehat{\mathbf{L}}_i \rangle I_i(\theta, \varphi)}{\sum_{i=0}^{n-1} \langle \widehat{\mathbf{N}}(\theta, \varphi), \widehat{\mathbf{L}}_i \rangle^2}.$$

Pour construire le système (2.9) dans les expérimentations à venir, nous avons ignoré les intensités nulles. Cependant, il existe un ensemble d'approches dont le but est de sélectionner efficacement les intensités pour chaque pixel [Barsky & Petrou 2003, Hernández *et al.* 2010]. Elles permettent d'éliminer les observations susceptibles d'introduire des erreurs sur la base des hypothèses formulées. Nous avons choisi de ne pas appliquer ces méthodes pour conserver l'attention sur la chaîne de traitements et l'approche proposée.

2.3 Estimation des distances radiales

2.3.1 Résolution adaptée aux images panoramiques

Pour prendre en compte les contraintes géométriques présentes sur la grille (θ, φ) , nous proposons une méthode d'intégration des gradients orientés qui les traite directement. Dans la section 2.2, nous avons observé que lorsque $\{\mathbf{v}, \mathbf{w}\}$ définit une base naturelle, alors p et q sont définis dans l'équation (2.7). À l'aide de la méthode des différences finies pour approximer les deux dérivées (voir (2.7)), nous obtenons :

$$\begin{aligned} p &= \frac{\partial \ln \rho}{\partial \theta} \simeq \frac{\ln \rho(\theta + \Delta\theta, \varphi) - \ln \rho(\theta, \varphi)}{\Delta\theta}, \\ q &= \frac{\partial \ln \rho}{\partial \varphi} \simeq \frac{\ln \rho(\theta, \varphi + \Delta\varphi) - \ln \rho(\theta, \varphi)}{\Delta\varphi}, \end{aligned} \tag{2.10}$$

où $\Delta\theta$ et $\Delta\varphi$ représentent respectivement les incréments (le pas) en suivant les axes θ et φ . Un système d'équations linéaires peut être construit à partir de (2.10). Nous pouvons le résoudre avec la méthode des moindres carrés. Ce système est parcimonieux et la vitesse de résolution du problème peut être augmentée en tenant compte de cette propriété. Les deux équations dans (2.10) sont définies sur l'ensemble de la grille sauf pour les points aux frontières de cette dernière (en considérant un voisinage de von Neumann). Nous proposons d'ajouter deux contraintes qui se concentrent sur les bords de la grille. La *contrainte de circularité* (aussi appelée condition périodique à la frontière dans [Frankot & Chellappa 1988]) repose sur le fait que les distances radiales de part et d'autre des frontières azimutales sont identiques. Nous pouvons

1. L'expression a été adaptée pour conserver la cohérence des notations.

ainsi ajouter $\rho(\theta, \varphi + \Delta\varphi) = \rho(\theta, 0)$ pour $\varphi = 2\pi$ avec $\theta \in (0, \pi/2]$ (voir la figure 2.3(a)). La deuxième contrainte concerne la couronne circulaire ($\theta = \Delta\theta$) proche du pôle Nord ($\theta = 0$).

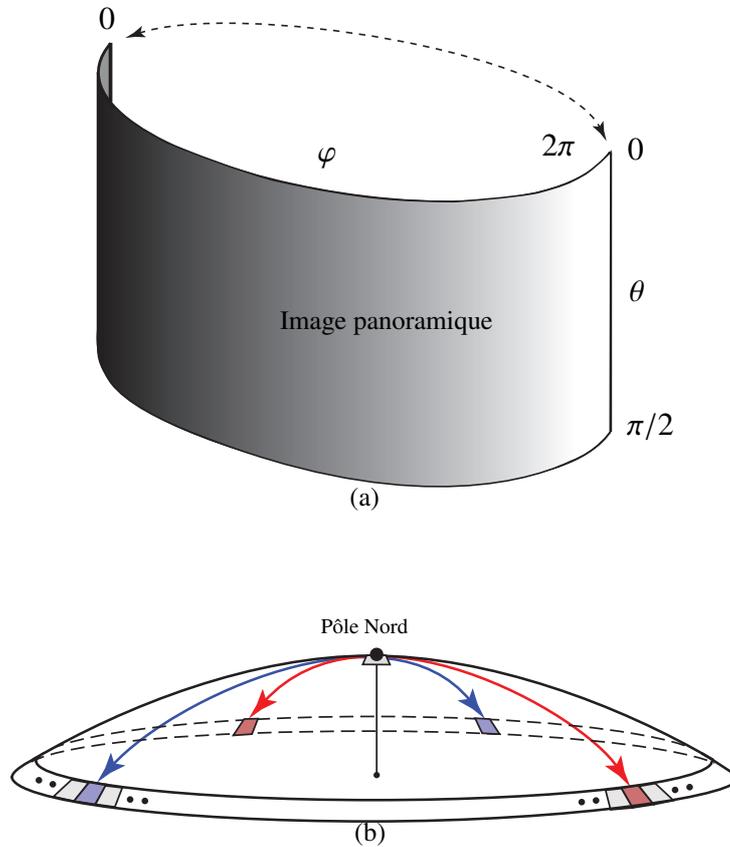


FIGURE 2.3 – Contraintes imposées par notre méthode d’intégration sans dérive des gradients orientés : (a) Contrainte de circularité et (b) Contrainte au pôle Nord. La distance entre le pôle et les voisins a été volontairement exagérée pour conserver la lisibilité de la figure.

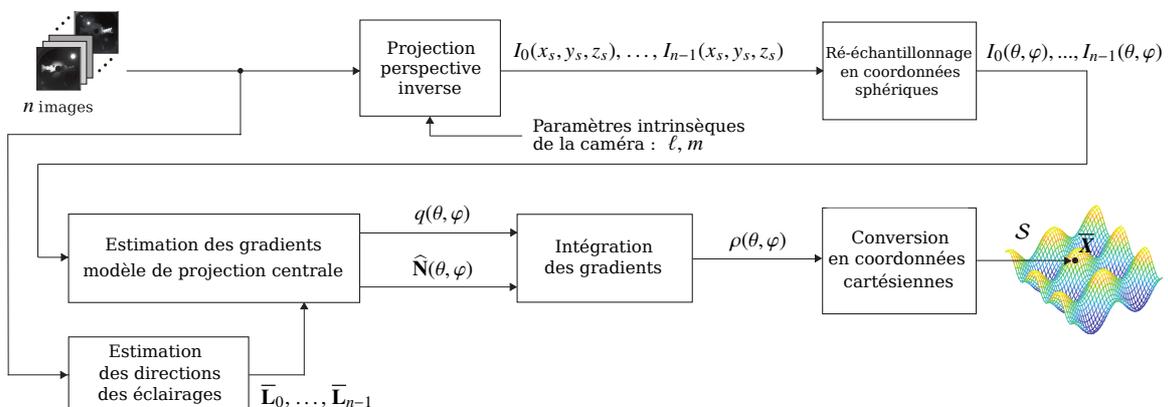


FIGURE 2.4 – Diagramme de flux de la chaîne de traitements proposée. L’algorithme a besoin de n images ainsi que des paramètres du capteur à projection centrale (obtenus par étalonnage ou imposés).

Une singularité du système de coordonnées sphériques est présente au pôle mais les éléments de la couronne circulaire peuvent être considérés comme des voisins. La *contrainte au pôle Nord* est représentée dans la figure 2.3(b). L'introduction de ces contraintes apporte un bénéfice non négligeable dans la réduction du phénomène de dérive que nous observerons dans la section 2.4.

La chaîne de traitements proposée est disponible dans la figure 2.4. Pour la compléter, la figure 2.5 apporte quelques éléments d'illustration avec l'*Escalier hélicoïdal* qui sera présentée dans la section 2.4.1. \bar{X} dans la figure 2.4 est un point générique sur la surface reconstruite à un facteur d'échelle près, dans un système de coordonnées cartésiennes (le changement de système de coordonnées est utile pour la visualisation du résultat uniquement).

2.4 Expérimentations

Cette section présente les trois scénarios mis en place pour évaluer la chaîne de traitements proposée. Les simulations numériques constituent les deux premiers et reposent sur des surfaces

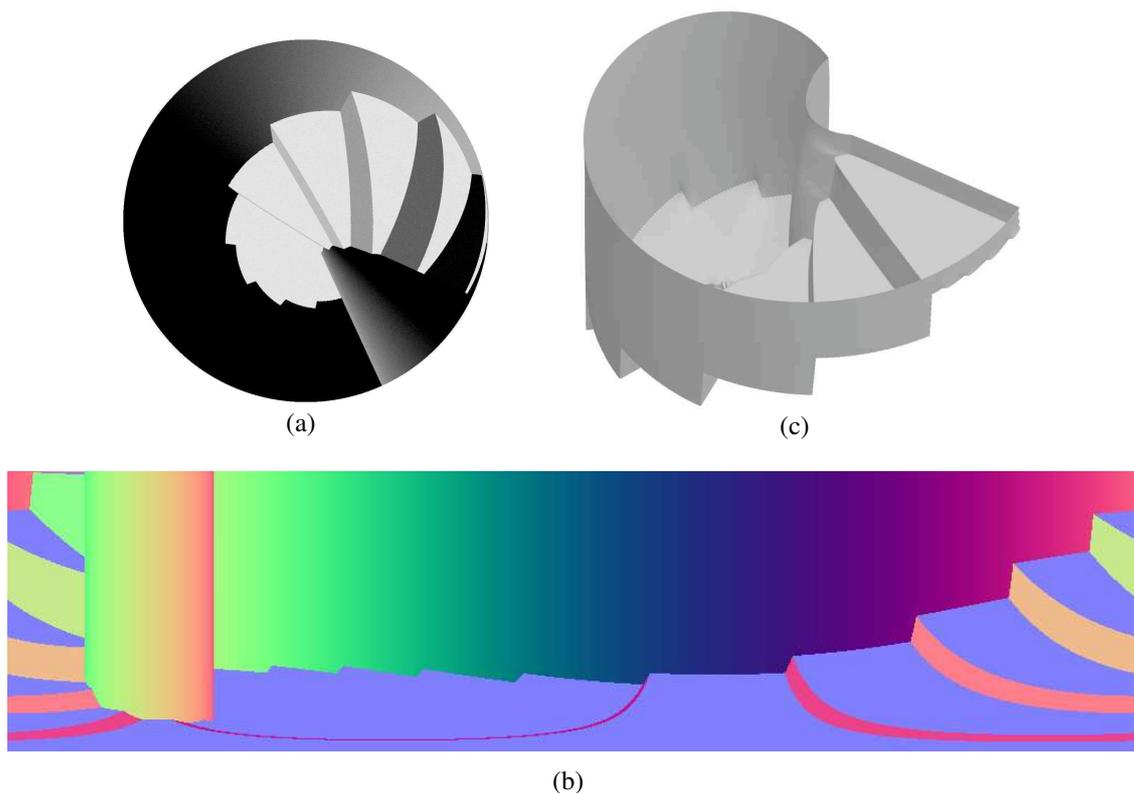


FIGURE 2.5 – Illustration de la chaîne de traitements proposée pour l'*Escalier hélicoïdal* : (a) Image panoramique d'exemple d'un escalier hélicoïdal éclairé selon huit directions d'illumination, générée à partir d'un capteur catadioptrique simulé ; (b) Carte des normales en fausses couleurs obtenue à l'aide des images (renversée verticalement pour faciliter la visualisation) ; (c) Reconstruction 3D de la scène.

analytiques (section 2.4.1) et des images de synthèse produites à partir de modèles 3D virtuels (section 2.4.1). Le dernier scénario exploite des images réelles issues d'une caméra perspective puis d'un capteur catadioptrique.

2.4.1 Validation en simulation

Surfaces analytiques

Deux surfaces de test ont été utilisées pour valider la méthode proposée ainsi que son implémentation. Leur expression mathématique est connue afin d'étudier par la suite les performances obtenues. Un ensemble d'images panoramiques de synthèse a été produit en illuminant virtuellement les deux surfaces depuis différentes directions. Les images produites sont utilisées comme données d'entrée pour la chaîne de traitements.

La première surface est appelée *Volcano* et son expression mathématique, en coordonnées sphériques, est (voir la figure 2.6(a)) :

$$\rho(\theta, \varphi) = \frac{9}{\cos \theta} + 100 \sin \theta. \quad (2.11)$$

Volcano présente une symétrie axiale qui se retrouve dans (2.11) puisqu'elle ne dépend pas du paramètre φ . La deuxième surface, appelée *Starfish* (voir la figure 2.6(d)), est décrite par l'expression suivante :

$$\rho(\theta, \varphi) = 6 + \sin\left(\frac{\theta}{10} + 5\varphi\right).$$

Pour évaluer précisément la qualité de l'estimation des gradients de la distance radiale, nous considérons le vecteur normal déduit ainsi que l'angle de déviation entre la direction estimée et la vérité-terrain connue :

$$\varepsilon(\theta, \varphi) = \arccos(\langle \mathbf{N}(\theta, \varphi), \bar{\mathbf{N}}(\theta, \varphi) \rangle), \quad (2.12)$$

avec $\theta \in \{\frac{\pi}{\lambda}, \frac{2\pi}{\lambda}, \dots, \frac{\pi}{2} - \frac{\pi}{\lambda}\}$, $\varphi \in \{0, \frac{\pi}{\lambda}, \dots, 2\pi - \frac{\pi}{\lambda}\}$, où $\bar{\mathbf{N}}(\theta, \varphi)$ correspond au vecteur normal à la surface au point (θ, φ) obtenu avec la méthode proposée et $1/\lambda$ au pas d'échantillonnage spatial (dans nos essais avec ces deux surfaces, $\lambda = 100$).

Pour les deux surfaces, les intensités dans les images $I_i(\theta, \varphi)$, $i \in \{0, 1, 2, 3\}$ ont été altérées par un bruit additif gaussien avec une variance σ^2 . Les quatre directions d'illumination utilisées pour produire les images ont été placées de façon symétrique pour illuminer uniformément les objets *Volcano* et *Starfish*.

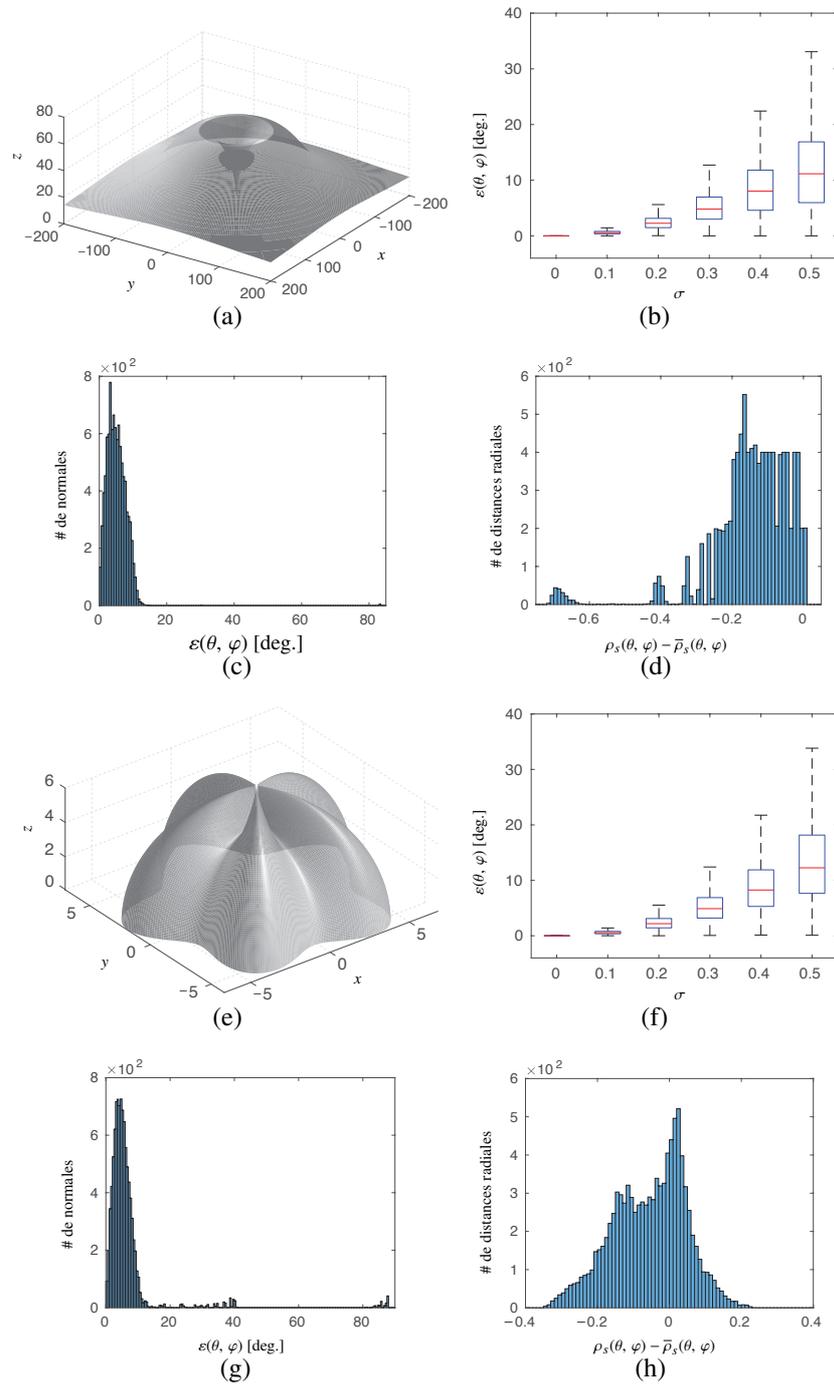


FIGURE 2.6 – Résultats obtenus pour les surfaces analytiques : reconstructions 3D de (a) *Volcano* et de (e) *Starfish* pour $\sigma = 0.3$; (b),(f) Boxplots de l'erreur $\varepsilon(\theta, \varphi)$ après ajout de bruit gaussien sur les intensités des pixels (10 tirages aléatoires, amplitude croissante); (c),(g) Distribution de l'erreur $\varepsilon(\theta, \varphi)$ (1 tirage aléatoire, largeur de colonne : 0.5°) pour $\sigma = 0.3$; (d),(h) Distribution de l'erreur sur la distance radiale $\rho_s(\theta, \varphi) - \bar{\rho}_s(\theta, \varphi)$ après mise à l'échelle (1 tirage aléatoire, largeur de colonne : 0.01) pour $\sigma = 0.3$.

Une fonction de saturation est appliquée sur les intensités perturbées pour les conserver dans l'intervalle $[0, 1]$. Pour simplifier la synthèse des images et éviter d'augmenter le nombre d'images, les ombres projetées (*self-occlusion*) n'ont pas été générées. Les figures 2.6(b) et (f) montrent les boxplots (générés avec Matlab) de l'erreur $\varepsilon(\theta, \varphi)$ (en degrés) respectivement pour *Volcano* et *Starfish*, pour un bruit d'amplitude croissante ajoutée aux images de synthèse. ($\sigma \in \{0, 0.1, \dots, 0.5\}$).

Nous pouvons remarquer que l'erreur $\varepsilon(\theta, \varphi) \neq 0$ dans le cas où aucun bruit n'est ajouté ($\sigma = 0$). Les quatre directions d'illumination considérées ne suffisent pas pour permettre d'obtenir une estimation parfaite des normales (l'erreur diminue de 25% en considérant 16 éclairages). Les figures 2.6(c) et 2.6(g) montrent la distribution de l'erreur $\varepsilon(\theta, \varphi)$ (largeur de colonne : 0.5°) pour $\sigma = 0.3$. Finalement les figures 2.6(d) et 2.6(h) présentent la distribution de l'erreur sur la distance radiale obtenue après la mise à l'échelle $\rho_s(\theta, \varphi) - \bar{\rho}_s(\theta, \varphi)$ (largeur de colonne : 0.01), à nouveau pour $\sigma = 0.3$, où,

$$\rho_s(\theta, \varphi) \triangleq \frac{\rho(\theta, \varphi)}{\max_{\theta, \varphi} \rho(\theta, \varphi)}.$$

Par définition, $\rho_s(\theta, \varphi) \in [0, 1]$. Les distances radiales $\bar{\rho}(\theta, \varphi)$ ont été obtenues après l'intégration des gradients en utilisant la méthode proposée en utilisant un pas de $\Delta\theta = \Delta\varphi = \pi/\lambda$. La mise à l'échelle des reconstructions 3D est nécessaire puisqu'elles sont obtenues à un facteur d'échelle près. Cette normalisation permet un alignement 3D-3D et introduit un biais dans le calcul de l'erreur. Dans la mesure où cette approche ne tient pas compte des valeurs aberrantes (*outliers*) de $\bar{\rho}(\theta, \varphi)$, l'alignement peut ne pas être optimal.

Images de synthèse

Afin de tester la méthode de reconstruction proposée dans des conditions davantage réalistes, nous avons conçu et développé une chaîne de traitements qui exploite le logiciel Blender [Blender Foundation 2018] pour créer une scène 3D dont nous avons extrait la vérité terrain que nous utilisons pour les évaluations quantitatives. Nous pouvons ainsi ajuster le point de vue et les paramètres intrinsèques de la caméra virtuelle. La vérité terrain contient le vecteur normal à la surface ainsi que la distance radiale entre le centre optique de la caméra et la surface, pour chaque point (θ, φ) et pour un échantillonnage spatial défini (en général, $\Delta\theta = \Delta\varphi = \pi/\lambda$ avec λ spécifié pour la scène à reconstruire). Nous avons ensuite généré des images panoramiques selon 8 directions d'illumination. Comme nous l'avons fait avec les surfaces analytiques dans la partie précédente, un bruit gaussien d'amplitude croissante a été appliqué sur les intensités. Les vecteurs d'illumination $\mathbf{L}_0, \dots, \mathbf{L}_7$ sont connus et les ombres portées ne sont pas simulées. L'effet de quantification est simulé en considérant $I_i(\theta, \varphi)$ sur 8 bits : les intensités $I_i(\theta, \varphi) \in [0, 1] \subset \mathbb{R}$ avec $i \in \{0, 1, \dots, 7\}$ sont transférées sur l'ensemble discret $\{0, 1, \dots, 255\}$.

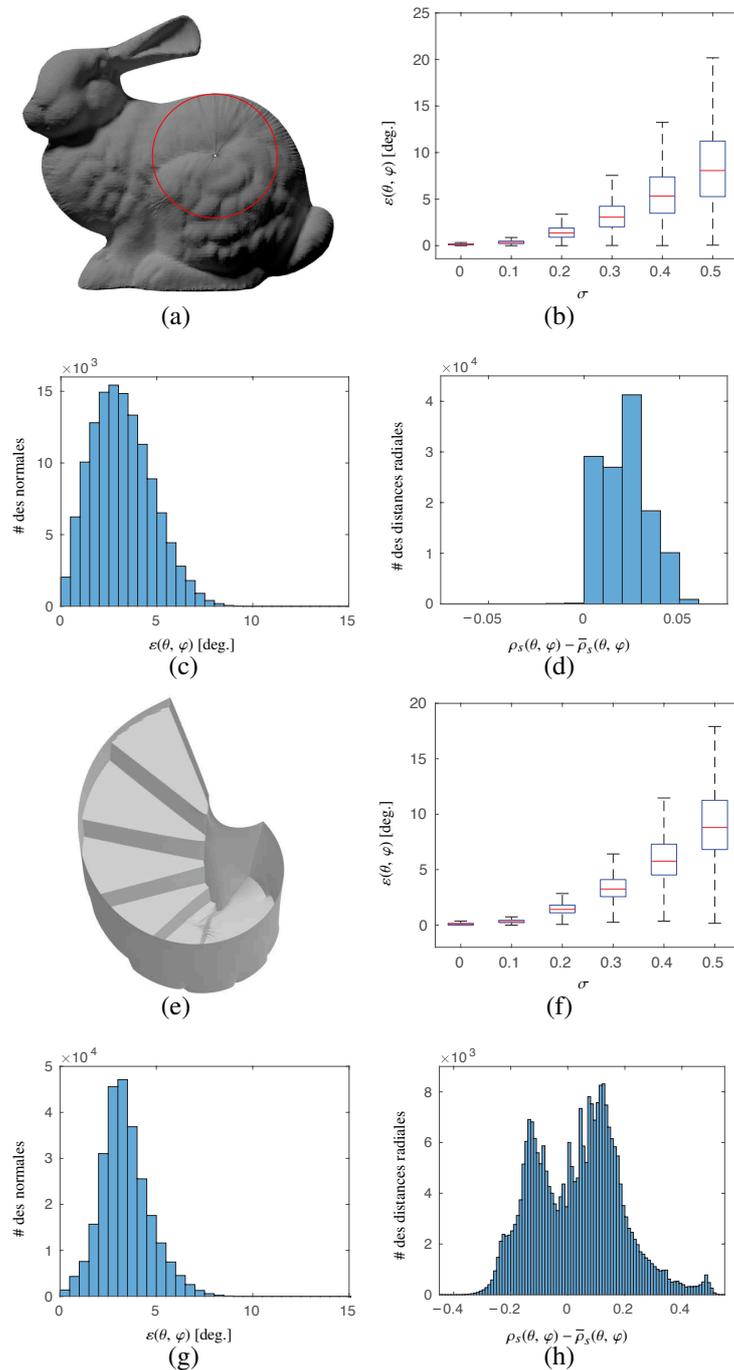


FIGURE 2.7 – Résultats obtenus pour les images de synthèse : reconstructions 3D de (a) *Stanford Bunny* et de (e) l'*Escalier hélicoïdal* pour $\sigma = 0.3$; (b),(f) Boxplot de l'erreur $\varepsilon(\theta, \varphi)$ après ajout de bruit gaussien sur les intensités des pixels (10 tirages aléatoires, amplitude croissante); (c),(g) Distribution de l'erreur $\varepsilon(\theta, \varphi)$ (1 tirage aléatoire, largeur de colonne : 0.5°) pour $\sigma = 0.3$; (d),(h) Distribution de l'erreur sur la distance radiale $\rho_s(\theta, \varphi) - \bar{\rho}_s(\theta, \varphi)$ après mise à l'échelle (1 tirage aléatoire, largeur de colonne : 0.01) pour $\sigma = 0.3$. Les erreurs mesurées pour *Stanford Bunny* correspondent à la région englobée par le cercle rouge.

Dans le premier essai de cette catégorie, nous avons utilisé le *Stanford Bunny*². La figure 2.7(a) montre la reconstruction 3D obtenue à la fin de la chaîne de traitements pour $\sigma = 0.3$. Le boxplot de l'erreur $\varepsilon(\theta, \varphi)$ en présence d'un bruit gaussien d'amplitude croissante est représenté dans la figure 2.7(b) et la distribution de cette même erreur pour $\sigma = 0.3$ est visible dans la figure 2.7(c) (largeur de colonne : 0.5°). La distribution de l'erreur $\rho_s(\theta, \varphi) - \bar{\rho}_s(\theta, \varphi)$ dans la zone délimitée par le cercle rouge visible dans la figure 2.7(a) est représentée dans la figure 2.7(d) (largeur de colonne : 0.01) pour $\sigma = 0.3$. Cette région a été délimitée pour éliminer les artefacts liés aux discontinuités de profondeur qui apparaissent sur les contours de la portion visible du modèle 3D. Cette limitation permet de réaliser une évaluation lisible et pertinente de la méthode d'intégration des gradients proposée sur le *Stanford Bunny*. Le pas d'échantillonnage considéré est $\Delta\theta = \Delta\varphi = \pi/\lambda$ avec $\lambda = 1024$.

Afin de réaliser une comparaison avec des méthodes d'intégration existantes, nous avons

2. Le modèle 3D du *Stanford Bunny* a été obtenu directement depuis le *Stanford 3D Scanning Repository*.

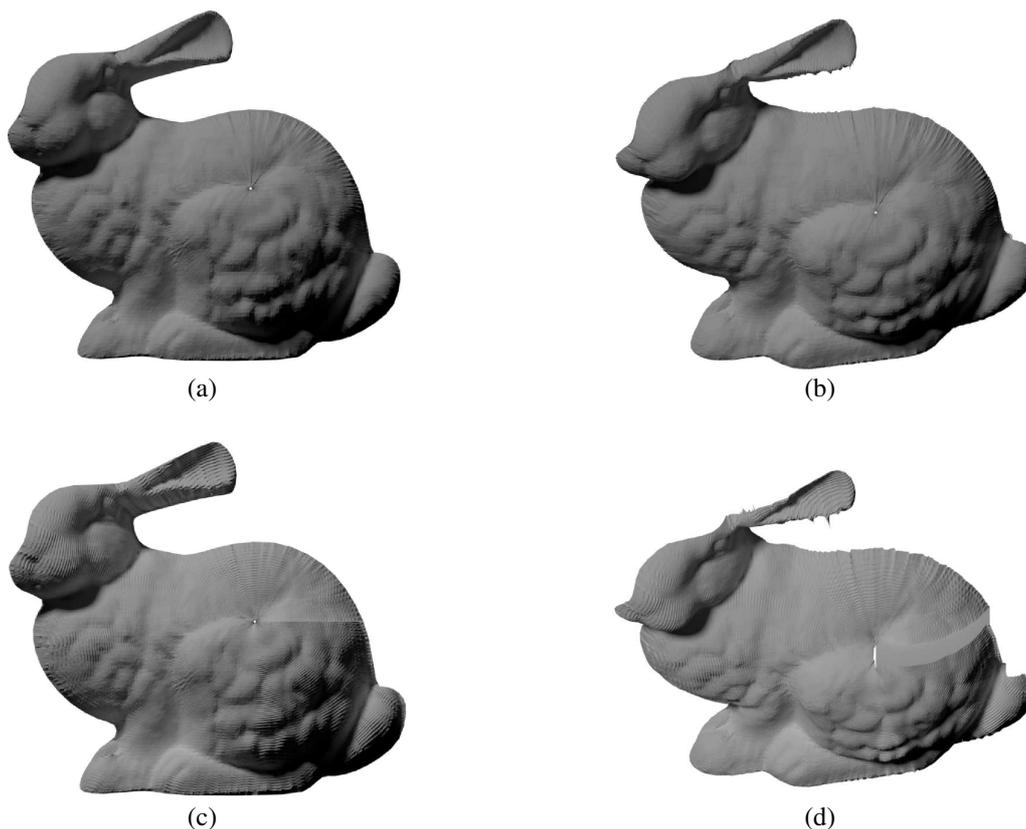


FIGURE 2.8 – Résultats obtenus pour les images de synthèse : reconstructions 3D du *Stanford Bunny* avec la méthode proposée (première ligne), et avec la méthode adaptée (deuxième ligne, voir 2.3), pour $\sigma = 0.3$; (a),(c) depuis le point de vue de la caméra, et (b),(d) depuis un point de vue incliné sur le dessus de l'objet. La dérive qui déforme l'objet reconstruit est visible dans (d).

adapté la méthode proposée dans [Harker & O’Leary 2008] pour qu’elle puisse exploiter les gradients que nous avons estimés. Précisément, nous avons substitué la grille rectangulaire (x, y) par une grille (θ, φ) . De ce fait, à la place de la profondeur $z(u, v)$, la solution de l’équation matricielle de Sylvester nous permet d’obtenir $\ln \rho(\theta, \varphi)$ puis directement la distance radiale $\rho(\theta, \varphi)$ entre le centre de la caméra O et les points 3D qui reposent sur la surface. La figure 2.8 permet de comparer les reconstructions 3D du *Stanford Bunny* selon deux points de vue différents : les contraintes de circularité et au pôle Nord permettent de réduire la dérive et les phénomènes de Gibbs observés dans la reconstruction 3D obtenue avec la méthode [Harker & O’Leary 2008].

Dans le deuxième essai, nous avons simulé un capteur catadioptrique pour produire des images panoramiques. Le capteur est placé au centre d’une scène 3D virtuelle qui représente l’intérieur d’un *Escalier hélicoïdal*. Le champ de vue de ce capteur permet de capturer en une seule image tout ce qui se trouve sous ce dernier. De façon similaire au précédent test avec le *Stanford Bunny*, les figures 2.7(e) à 2.7(h) présentent respectivement, pour l’*Escalier hélicoïdal* et pour $\sigma = 0.3$, la reconstruction 3D obtenue avec notre méthode, le boxplot de l’erreur $\varepsilon(\theta, \varphi)$ en présence d’un bruit gaussien d’amplitude croissante ainsi que les distributions des erreurs $\varepsilon(\theta, \varphi)$ et $\rho_s(\theta, \varphi) - \bar{\rho}_s(\theta, \varphi)$. De la même manière, le pas d’échantillonnage est $\Delta\theta = \Delta\varphi = \pi/\lambda$, avec $\lambda = 1024$.

2.4.2 Évaluation en conditions réelles et contrôlées

Dans le dernier scénario, nous utilisons des images réelles acquises par deux capteurs : une caméra perspective et un capteur catadioptrique. Dans cette section, nous allons d’abord présenter le matériel utilisé ainsi que les méthodes utilisées pour étalonner les capteurs et pour estimer les directions d’illumination mises en place. Ensuite, nous présenterons successivement les résultats qualitatifs obtenus avec les deux capteurs.

Matériel utilisé

Une cabine en aluminium a été construite et posée sur une table d’optique pour accueillir et maintenir une toile de tissu occultant. L’objectif du dispositif est de permettre le contrôle de l’illumination dans l’espace couvert sans être perturbé par les éclairages naturel et artificiel qui se trouvent à proximité (voir la figure 2.9). Deux petits objets en polystyrène expansé *Chat* (23 cm de haut, 14 cm au plus large) et *Ourson* (19 cm de haut, 12 cm au plus large) ont été placés dans la cabine afin de les observer. La surface blanc mat des objets suit globalement les caractéristiques d’une surface lambertienne avec un albedo non unitaire. La source de lumière à

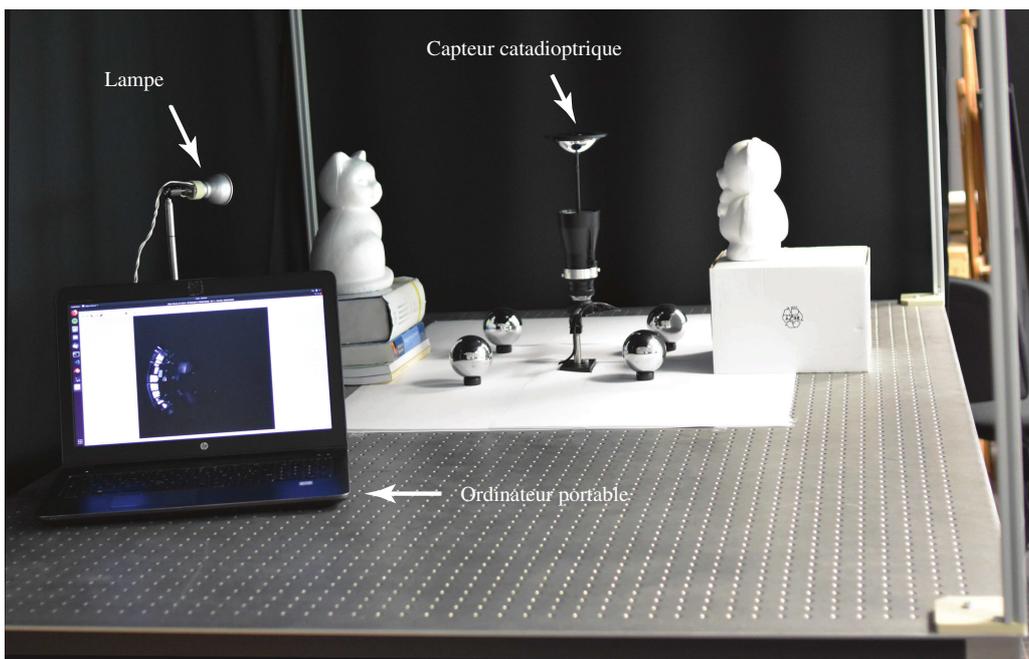


FIGURE 2.9 – Matériel dédié aux expérimentations : la lampe, les objets à reconstruire (*Chat*, à gauche, et *Ourson*, à droite) et le capteur catadioptrique. La caméra du capteur orienté vers le haut est connectée à l'ordinateur portable à l'aide d'un câble USB. Pendant les expérimentations, le rideau occultant qui recouvre la structure est fermé.

l'intérieur de la cabine est une lampe OSRAM 5-LED³ et constitue une source qui n'est ni directionnelle de par son cône de rayons lumineux ni angulairement uniforme. Les images ont été acquises avec deux boîtiers IDS UI-1245LE-C-HQ. Un objectif Tamron (8 mm, \varnothing 25.5 mm) a été ajouté au premier boîtier pour réaliser la capture des images perspectives. Un miroir parabolique RemoteReality avec un objectif télécentrique a été fixé sur le second boîtier pour obtenir les images catadioptriques. Les définitions en pixels des images perspectives et catadioptriques sont respectivement 1280 pixels \times 1024 pixels et 1024 pixels \times 1024 pixels. Les boîtiers produisent des images RVB codées sur 24 bits que nous avons converties en images monochromes codées sur 8 bits, afin de respecter le format des données d'entrée de la chaîne de traitements proposée. Afin d'estimer la fonction de réponse de la caméra (*Camera Response Function* ou CRF), nous avons capturé une séquence de 16 images avec des temps d'exposition croissants. Nous avons ensuite appliqué la méthode détaillée dans [Grossberg & Nayar 2004], cependant, les résultats n'étaient pas satisfaisants : nous avons observé une forte divergence entre la fonction estimée et la courbe obtenue en représentant les intensités mesurées en fonction des temps d'exposition. Par conséquent, comme cette dernière courbe est linéaire, nous avons suivi la mé-

3. La lampe utilisée dispose des caractéristiques suivantes : 5.3 W de puissance avec un rendement de 66.04 lm/W, indice de rendu des couleurs $R_a = 80$, température de couleur de 4000 K, angle d'ouverture du faisceau lumineux de 36° .

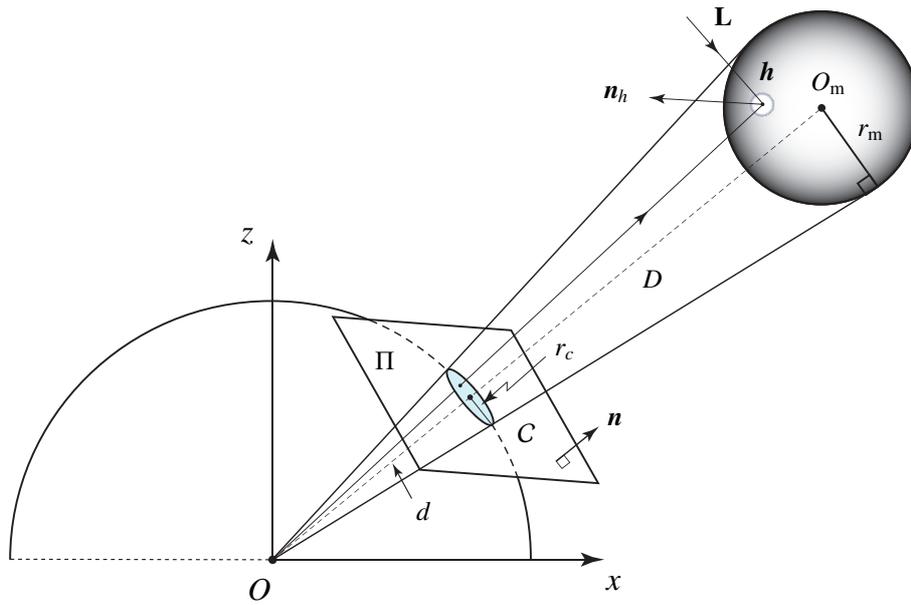


FIGURE 2.10 – Estimation de la direction d’éclairage \mathbf{L} à partir de la tache spéculaire visible sur une sphère réfléchissante de rayon r_m et de centre O_m (en haut à droite de la figure). Une partie de la sphère unitaire du modèle unifié pour les capteurs à projection centrale est représentée en bas à gauche.

thode proposée dans [Diaz & Sturm 2013] et nous avons calculé les paramètres d’une fonction affine approchant les données mesurées pour obtenir l’estimation de la fonction de réponse.

Estimation des directions d’éclairage

Pour réussir à estimer les directions d’illumination $\mathbf{L}_0, \dots, \mathbf{L}_{n-1}$, nous avons adapté la méthode proposée dans [Schnieders & Wong 2013]. Initialement adaptée aux caméras perspectives, elle a été adaptée pour prendre en compte le modèle unifié de projection. La méthode repose sur l’observation d’une sphère réfléchissante de rayon r_m connu, exposée directement au rayonnement d’une source de lumière. La sphère se trouve à une distance D du centre O de la sphère unitaire du modèle unifié (voir la figure 2.10). L’algorithme 1 résume la succession d’étapes proposée pour estimer la direction de l’illumination \mathbf{L} en utilisant la projection inverse du modèle (passage du plan de l’image numérique à la sphère unitaire virtuelle). Le cercle C correspond à l’intersection entre la sphère unitaire et le cône tangent à la sphère réfléchissante ayant pour origine O : son axe de révolution est radial à la sphère unitaire et passe par les centres des sphères unitaire et réfléchissante. La direction d’illumination \mathbf{L} , la ligne de vue du centre O jusqu’au centre de la tache spéculaire et le vecteur normal \mathbf{n}_h à la sphère réfléchissante au point h appartiennent tous à un même plan. Pour simplifier l’estimation, les coordonnées dans l’image de la tache spéculaire sont approximées par le centre de gravité de l’ensemble des pixels qui la composent.

Caméra perspective

Avant d'utiliser le matériel présenté, nous avons évalué la précision de l'estimation des gradients sur des images perspectives provenant du jeu de données 'DiLiGenT' (voir [Shi *et al.* 2019]). Nous avons utilisé les 96 images du chat produites selon autant de directions d'éclairage différentes. L'estimation des gradients repose sur un système d'équations avec $N_{\text{eq}} = 4560$ (cf. 2.2.2). Après avoir normalisé les intensités lumineuses codées sur 8 bits, nous avons éliminé les valeurs inférieures au seuil de 0.05, afin de filtrer une partie du bruit. Les figures 2.11(a) et 2.11(b) montrent, en fausses couleurs, la vérité terrain et l'estimation par notre méthode des vecteurs normaux à la surface de l'objet. À l'exception de deux régions où les ombres propres et projetées n'ont pas été filtrées correctement, les deux représentations sont très similaires. Le tableau 2.1 rappelle les résultats quantitatifs obtenus

Algorithme 1 : Estimation de la direction d'illumination \mathbf{L} à partir d'une tache spéculaire sur une sphère réfléchissante et en utilisant le modèle unifié pour les capteurs à projection centrale (voir la figure 2.10).

Données : dans une image contenant une sphère réfléchissante directement éclairée :

- a) Le contour apparent de la sphère,
- b) Les coordonnées de la tache spéculaire visible sur la sphère,
- c) Le rayon r_m connu de la sphère.

Résultat : la direction estimée de l'éclairage $\bar{\mathbf{L}}$.

- 1 Les pixels du contour apparent sont projetés sur la sphère unitaire en utilisant la projection perspective inverse afin de former le cercle C . Les coordonnées d'un point j appartenant à C sont notées $\mathbf{x}_{C_j} \in \mathbb{R}^3$ et $\mathbf{x}_{\text{cnt}} \in \mathbb{R}^3$ contient les coordonnées du barycentre des points \mathbf{x}_{C_j} , $j \in \{1, \dots, P\}$.
 - 2 Afin de réduire les effets d'un mauvais conditionnement numérique, les coordonnées des points \mathbf{x}_{C_j} sont centrées à l'origine puis mises à l'échelle pour être comprises dans l'intervalle $[-1, 1]$. Les points ainsi transformés sont notés $\mathbf{x}_{C_j}^*$.
 - 3 La décomposition en valeurs singulières de $\mathbf{M} = [\mathbf{x}_{C_1}^*, \mathbf{x}_{C_2}^*, \dots, \mathbf{x}_{C_P}^*] \in \mathbb{R}^{3 \times P}$ est notée $\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ où $\mathbf{\Sigma}$ contient les valeurs singulières de \mathbf{M} classées dans l'ordre décroissant dans la diagonale principale. Dans cette configuration, la troisième colonne de \mathbf{U} correspond au vecteur normal $\mathbf{n} = [n_x, n_y, n_z]^T$ au plan Π qui contient le cercle C .
 - 4 Si $n_z < 0$, nous conservons le vecteur opposé $-\mathbf{n}$.
 - 5 La distance d entre le plan Π et le centre de la sphère unitaire vaut $d = |\langle \hat{\mathbf{n}}, \mathbf{x}_{\text{cnt}} \rangle|$.
 - 6 Les coordonnées du centre de C sont obtenues avec $\mathbf{o}_c = -d \hat{\mathbf{n}}$ et son rayon $r_c = \sqrt{1 - d^2}$.
 - 7 Le centre O_m de la sphère réfléchissante a pour coordonnées $\mathbf{o}_m = \mathbf{o}_c r_m / r_c$.
 - 8 Le vecteur normal à la surface de la sphère réfléchissante au point \mathbf{h} est unitaire et obtenu avec $\hat{\mathbf{n}}_h = (\mathbf{h} - \mathbf{o}_m) / r_m$.
 - 9 La direction de l'éclairage \mathbf{L} est finalement estimée par $\bar{\mathbf{L}} = 2 \langle \hat{\mathbf{h}}, \hat{\mathbf{n}}_h \rangle \hat{\mathbf{n}}_h - \hat{\mathbf{h}}$.
-

TABEAU 2.1 – Résultats avec un extrait du jeu de données DiLiGenT (voir [Shi *et al.* 2019]) : Moyenne de l’erreur angulaire (MEA) en degrés de quelques méthodes de l’état de l’art et de la méthode proposée.

Algorithme	Wo80	WG10	IW12	GC10	AZ08
MEA	8.41	6.72	7.21	8.21	6.53
Algorithme	HM10	ST12	ST14	IA14	Proposé
MEA	8.40	12.34	6.12	6.74	8.36

par les auteurs du jeu de données ainsi que le résultat obtenu avec la méthode proposée. La moyenne de l’erreur angulaire (revoir (2.12)) permet de faire des comparaisons bien que les méthodes de l’état de l’art mises en compétition ne s’adressent pas spécifiquement au problème de stéréophotométrie calibrée. Les abréviations et les initiales correspondent aux méthodes suivantes : Wo80 [Woodham 1980], WG10 [Wu *et al.* 2010], IW12 [Ikehata *et al.* 2012], GC10 [Goldman *et al.* 2010], AZ08 [Alldrin *et al.* 2008], HM10 [Higo *et al.* 2010], ST12 [Shi *et al.* 2012], ST14 [Shi *et al.* 2014] et IA14 [Ikehata & Aizawa 2014]. Davantage de détails sont disponibles dans [Shi *et al.* 2019, Sect. 2]. Dans ce premier essai, toute la chaîne de traitements n’a pas été appliquée : dans la mesure où l’estimation des gradients est faite localement (par pixel ou par point sur la sphère) et que nous ne souhaitons pas estimer la reconstruction 3D de l’objet mais seulement les normales de sa surface, nous avons ignoré la phase de ré-échantillonnage spatial, ce qui permet de comparer les méthodes, sans ajouter de

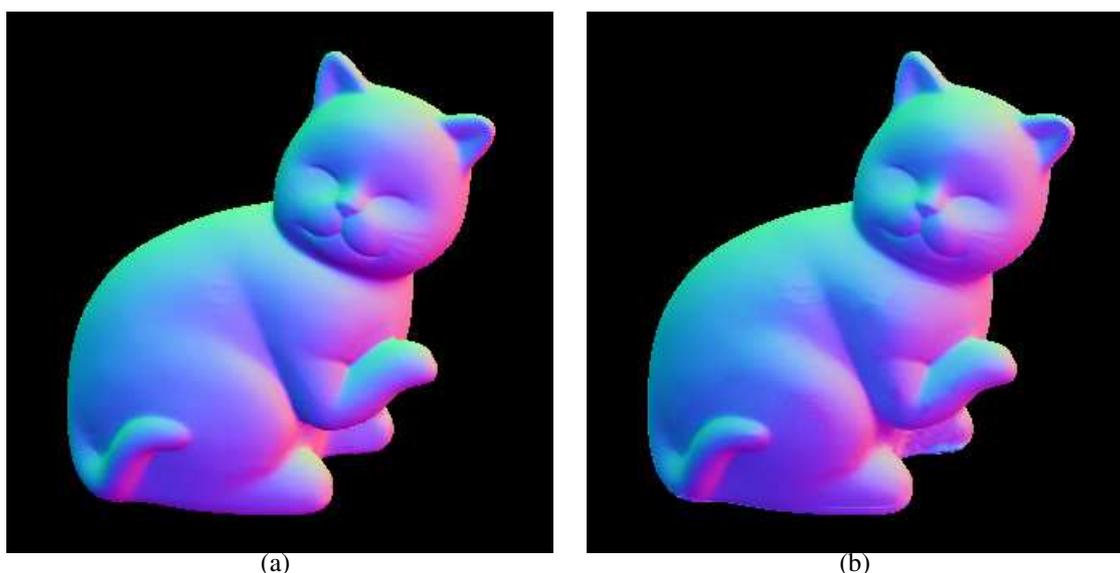


FIGURE 2.11 – Résultats avec un extrait du jeu de données DiLiGenT [Shi *et al.* 2019] : (a) Vérité terrain des normales en fausses couleurs ; (b) Normales obtenues par notre approche.

biais.

Pour le deuxième test impliquant des images perspectives, nous avons réalisé une évaluation qualitative de l'ensemble de la chaîne de traitements proposée. Pour y parvenir, nous avons utilisé 20 images du *Chat* produites par la caméra perspective (voir 2.4.2). L'étalonnage de la caméra perspective a été fait avec le logiciel HySCaS (voir [Caron & Eynard 2011b]) : cet outil exploite une version améliorée (voir [Barreto 2006]) du modèle unifié pour les capteurs à projection centrale présenté dans la section 2.1. À partir du paramètre $\xi = 0.82655$ estimé pour cette caméra, nous pouvons retrouver les paramètres l et m . Pour l'estimation des direc-

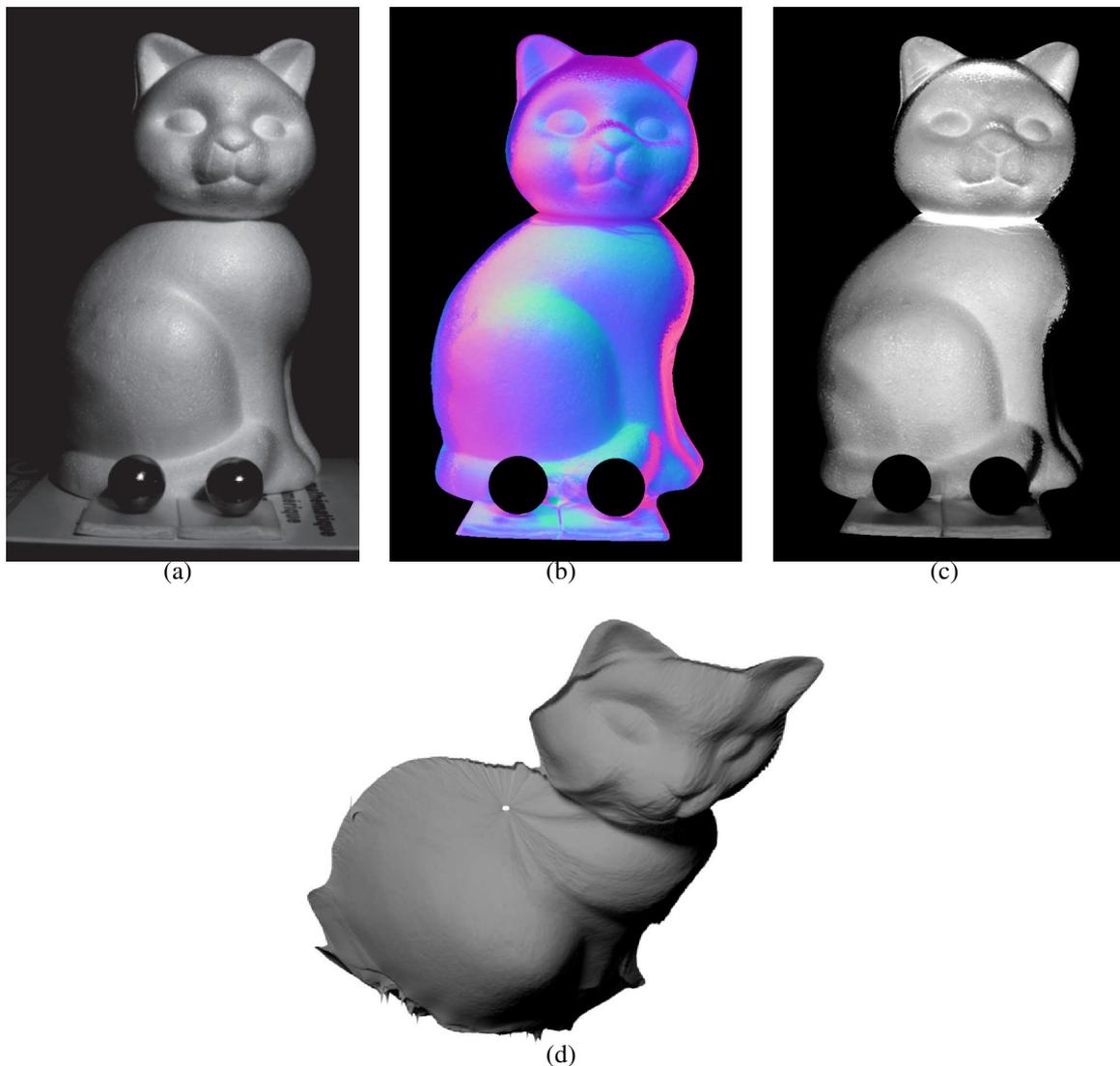


FIGURE 2.12 – Résultats pour le jeu de données réelles (caméra perspective) de l'objet *Chat* : (a) Une des 20 images utilisées ; (b) Normales (en fausses couleurs) estimées par la méthode proposée ; (c) Albedo estimé (voir (2.2.2)) et (d) Reconstruction 3D de l'objet. Les sphères réfléchissantes à la base de l'objet sont utilisées pour estimer les direction des éclairages.



FIGURE 2.13 – Expérimentations avec images réelles : le capteur catadioptrique, les objets *Chat* et *Ourson*, ainsi que les quatre sphères réfléchissantes pour l’estimation des directions d’éclairage.

tions d’éclairage, nous avons ajouté deux billes d’acier à la base de la figurine (voir la figure figure 2.12(a)). Leur rayon r_m est de 12.5 mm. La figure 2.12(b) montre le résultat de l’estimation des normales en fausses couleurs après avoir appliqué un masque pour conserver uniquement la figurine. L’estimation de l’albedo à partir des normales et des images d’entrée est affichée dans la figure 2.12(c) (voir la remarque 2.2.2 pour les détails). Pour terminer, la reconstruction 3D du *Chat* obtenue avec la méthode proposée est représentée dans la figure 2.12(d).

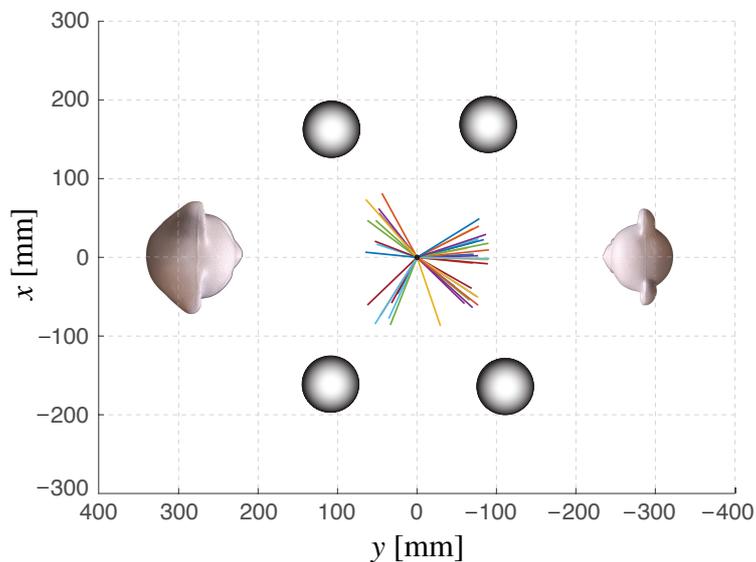


FIGURE 2.14 – Vue du dessus des directions estimées des éclairages (segments colorés) dans le repère caméra.

Capteur catadioptrique

Le dernier essai de cette série a été réalisé avec des images catadioptriques des deux figurines. Nous avons choisi d'illuminer un seul objet dans chaque image : la figurine *Chat* est observée selon 26 directions d'illumination, *Ourson* l'est selon 16 directions. La mise en place de l'expérience est visible dans la figure 2.13. Le capteur catadioptrique pointe vers le plafond de la cabine, ce qui permet d'obtenir des images de tout ce qui se trouve sur la table optique. Les quatre sphères réfléchissantes de rayon $r_m = 35.5$ mm sont placées à proximité des objets observés : elles servent à estimer la direction de l'éclairage dans le repère du capteur catadioptrique. Au moins deux sphères réfléchissantes sont illuminées par la source de lumière afin de

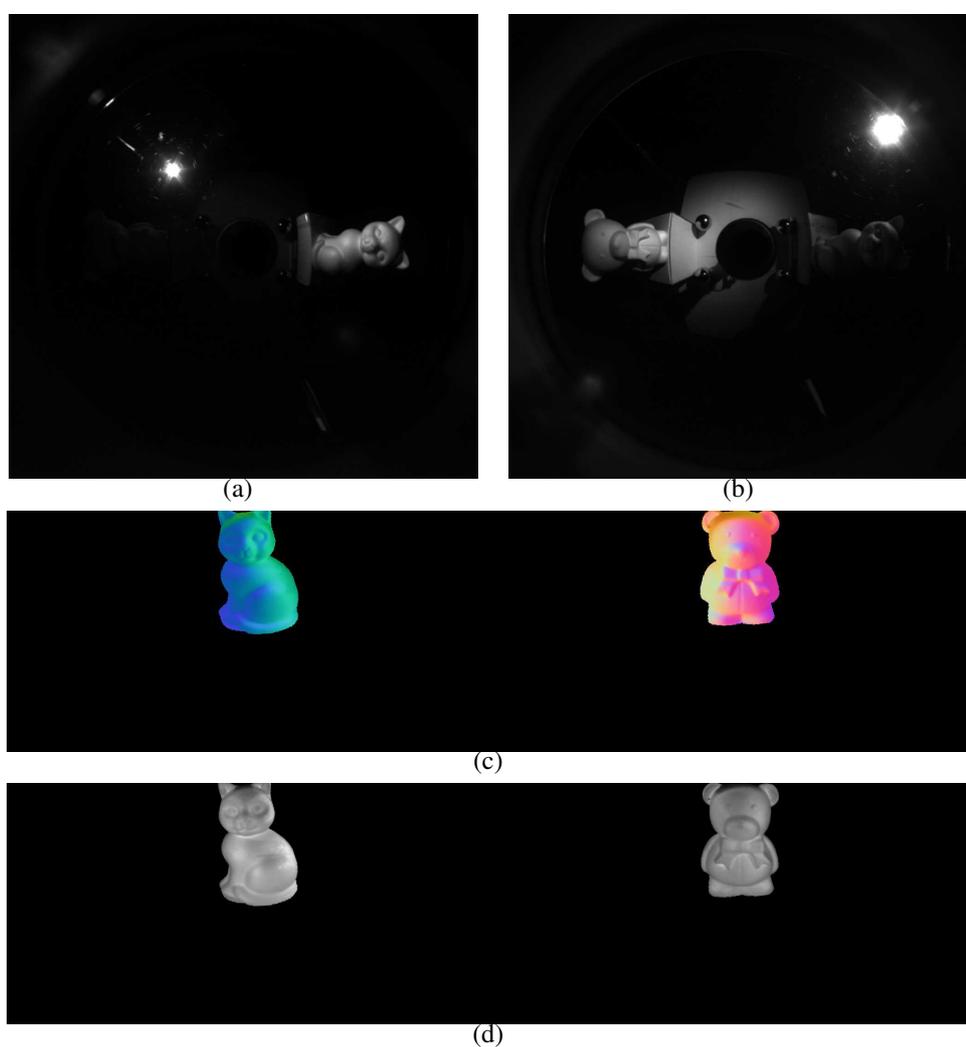


FIGURE 2.15 – Résultats pour le jeu de données réelles (capteur catadioptrique) des objets *Chat* et *Ourson* : (a),(b) Exemples des images utilisées ; (c) Normales (en fausses couleurs) estimées par notre méthode ; (d) Albedo estimé (voir (2.2.2)). Les images (c) et (d) ont subi une rotation de 180° pour améliorer la lisibilité.

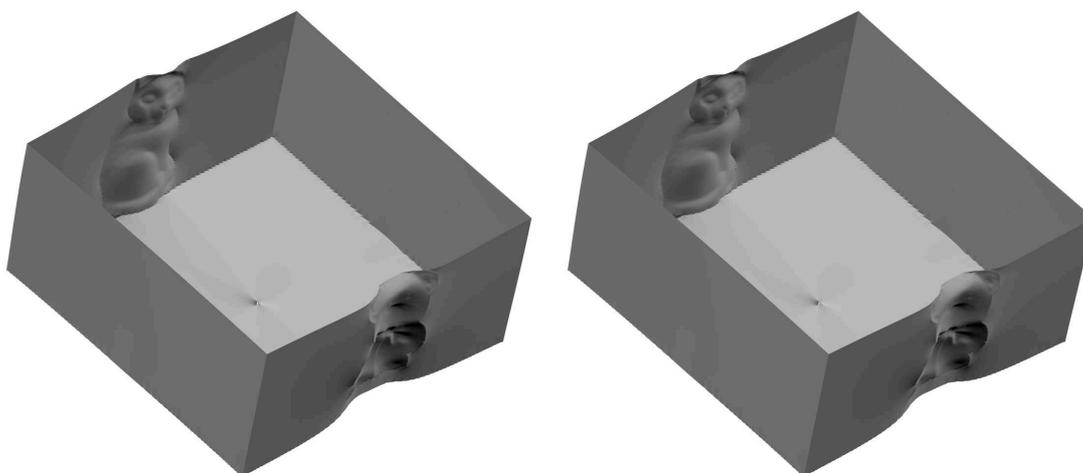


FIGURE 2.16 – Reconstruction 3D des objets *Chat* et *Ourson* observée depuis deux points de vue.

disposer d'une redondance pour estimer les directions d'éclairage. Pour information, la répartition des directions des éclairages est représentée dans la figure 2.14. Comme pour la caméra perspective, le capteur catadioptrique a été étalonné avec le logiciel HySCaS.

Les figures 2.15(a) et 2.15(b) montrent deux images catadioptriques des figurines *Chat* et *Ourson* : un seul objet est illuminé à la fois. La lampe utilisée est visible dans l'image (point lumineux à l'opposé de la figurine) et se trouve au plus à une distance de 1.7 m des objets. Les normales déduites des gradients estimés sont représentées en fausses couleurs, après application d'un masque binaire pour segmenter les objets, dans la figure 2.15(c). L'albedo estimé avec la méthode présentée dans la section 2.2.2 est disponible dans la figure 2.15(d). La reconstruction 3D finale des deux objets en utilisant la méthode d'intégration proposée est présente dans la figure 2.15(e) (pas d'échantillonnage $\Delta\theta = \Delta\varphi = \pi/\lambda$ avec $\lambda = 1024$). Pour appliquer les contraintes de circularité et au pôle Nord dans la phase d'intégration, nous avons complété les gradients dans les régions qui ne correspondent pas aux figurines *Chat* et *Ourson* : l'ajout correspond à cinq plans orthogonaux entre eux (un demi cube). Les figurines semblent avoir subi un aplatissement dans la figure 2.15(e) et restent identifiables.

Parmi les écarts par rapport aux hypothèses de départ, nous pouvons noter, concernant la lampe, qu'elle n'est pas orthotrope dans son cône de diffusion : un motif apparaît à cause des cinq diodes électroluminescentes. Placée au plus à 1.7 m des figurines, elle ne peut pas être considérée comme une source de lumière quasi-directionnelle. L'atténuation de l'éclairage n'est pas prise en compte et présente un effet qui dépend de la distance entre la lampe et les objets illuminés. Ces limitations intrinsèques à notre protocole expérimental seront reprises et étudiées dans le prochain chapitre. Les résultats présentés dans ce chapitre peuvent être retrouvés dans l'article publié [Caracotte *et al.* 2020].

Conclusion

Dans ce deuxième chapitre, nous avons présenté le modèle unifié pour les capteurs à projection centrale, afin de pouvoir introduire et détailler l'équation de l'irradiance pour les images sphériques. Nous avons ajouté la chaîne de traitements numériques adaptée à cette nouvelle équation et nous nous en sommes servis pour réaliser une batterie de tests en simulation puis en conditions réelles et contrôlées. Ces expérimentations nous ont permis de mettre en évidence le fonctionnement ainsi que certaines limites à sa mise en pratique que nous aborderons à nouveau dans le chapitre suivant.

Extension aux caméras *twin-fisheye*

Contenu du chapitre

3.1	Caméra <i>twin-fisheye</i> : applications	50
3.2	Adaptation de la chaîne de traitements	56
3.2.1	Modélisation du capteur	56
3.2.2	Fusion des données	59
3.2.3	Estimation des distances radiales	60
3.3	Expérimentations	60
3.3.1	Validation en simulation	61
3.3.2	Évaluation en conditions réelles et contrôlées	63
3.3.3	Discussion	67
3.4	Analyse des reconstructions à partir de données réelles	68
3.4.1	Éclairage non directionnel	68
3.4.2	Atténuation de l'éclairage	69

Résumé

Ce chapitre se concentre sur l'adaptation de la méthode proposée dans le chapitre précédent aux caméras *twin-fisheye*, c'est-à-dire, des caméras équipées de deux lentilles fisheye dirigées dans deux directions opposées. Alors que les caméras panoramiques sont utilisées depuis la fin des années 90, les caméras *twin-fisheye* ont fait leur apparition très récemment sur le marché et sont devenues accessibles au plus grand nombre. Elles permettent d'avoir une vue complète de l'environnement autour du capteur en une seule image. Elles augmentent la couverture des surfaces à reconstruire au détriment de la résolution spatiale qui n'est plus uniforme sur l'ensemble de l'image obtenue. Les avantages de ces caméras ont été exploités dans plusieurs travaux de recherche que nous passons en revue dans la première section de ce chapitre. Nous y découvrirons plusieurs champs de recherche impliqués tels que, par exemple, la reconstruction 3D multivues, l'exploration et l'aide à la décision en réalité virtuelle ainsi que l'estimation de déplacement et/ou d'orientation par odométrie visuelle. La suite du chapitre sera dédiée à la modélisation

géométrique de cette caméra et nous proposerons une extension de la méthode de reconstruction vue dans le chapitre 2. La modélisation détaillée dans le chapitre précédent est ajustée puis une série d'expérimentations montre le fonctionnement de l'approche. Pour terminer, nous nous intéresserons à quelques limites de l'approche que nous avons identifiées. Tout au long de ce chapitre, nous utiliserons les termes *twin-fisheye* pour faire référence au capteur de vision et *dual-fisheye* lorsqu'il s'agit des images acquises par la caméra.

3.1 Caméra *twin-fisheye* : applications

Dans cette section, nous présentons une revue rapide des travaux de recherche qui exploitent ces capteurs *twin-fisheye*. Ces caméras existent depuis une dizaine d'années et de plus en plus de domaines de recherche exploitent les images qu'elles produisent, par exemple, en robotique mobile, en reconstruction 3D, en réalité virtuelle, etc. Elles se sont progressivement miniaturisées et parmi les modèles de caméras grand public disponibles, nous pouvons citer la Samsung Gear 360, la LG 360 CAM, les différentes séries de caméras Ricoh Theta et Insta360. Elles offrent un très grand champ de vue, avec une faible zone aveugle. Outre ces avantages, ces caméras possèdent une propriété très intéressante et souvent exploitée : les points de fuite sont présents dans les images obtenues.

Les premières caméras *twin-fisheye* n'étaient pas compactes. Dans [Jamaluddin *et al.* 2017], les auteurs proposent un système qui repose sur deux caméras fisheye placées sur un banc stéréo et pointées dans deux directions opposées (figure 3.1(a)). Le modèle unifié présenté dans le chapitre précédent est employé afin de modéliser chacune des deux caméras. La fusion des images obtenues avec les deux caméras permet d'obtenir deux résultats :

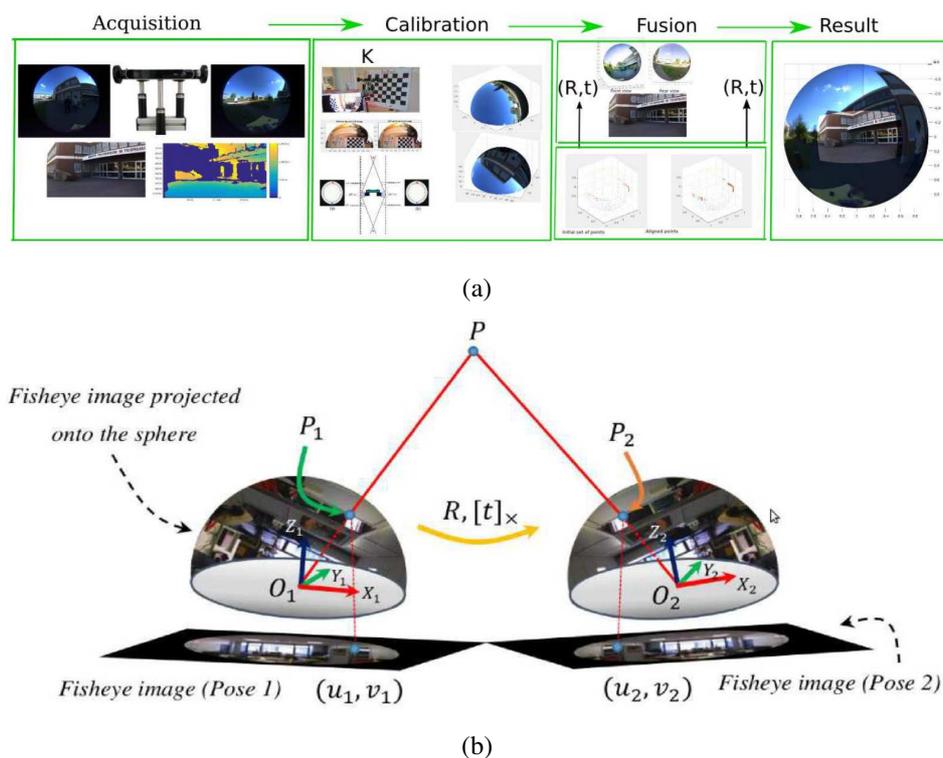


FIGURE 3.1 – Dans [Jamaluddin *et al.* 2017], les caméras fisheye permettent de produire des images sphériques et des reconstructions 3D dans la zone de recouvrement des deux capteurs : (a) Chaîne de traitements complète, (b) reconstruction 3D par géométrie épipolaire.

- une image 360° de la scène ;
- une reconstruction 3D éparsée de la zone de recouvrement entre les deux champs de vue de 185° de chaque capteur.

La figure 3.1(b) présente les images acquises par la caméra, la projection perspective inverse estimée puis la mise en correspondance des points d'une image sphérique à l'autre.

Dans [Gao & Shen 2017], deux caméras équipées de lentilles fisheye sont placées dos à dos à la verticale (figure 3.2(a)). Chaque image *dual-fisheye* (figure 3.2(b)) est découpée en quatre images perspectives afin d'obtenir des paires d'images similaires à celles d'un banc stéréo (figure 3.2(c)). Les auteurs proposent une méthode pour l'étalonnage de chaque caméra en s'appuyant sur les données optiques des lentilles, fournies par les fabricants. Une procédure automatisée pour l'estimation de la transformation rigide entre les deux caméras, ainsi qu'un algorithme pour estimer la profondeur des points visibles dans les paires d'images perspectives sont détaillés. La profondeur est estimée uniquement pour les points visibles dans les deux images.

Dans [Barazzetti *et al.* 2017], la caméra Samsung Gear 360 (figure 3.3(a)) est mise à l'épreuve afin de réaliser des reconstructions 3D d'environnements en intérieur (figure 3.3(b)). La précision métrique des reconstructions obtenues par photogrammétrie est quantifiée. Les résultats obtenus montrent qu'il est préférable d'utiliser les images fisheye brutes à la place de l'image équirectangulaire reconstruite automatiquement par le capteur. La modélisation du capteur proposée permet de réaliser une projection équirectangulaire des images fisheye avec davantage de précision. Les reconstructions obtenues avec les images équirectangulaires géné-

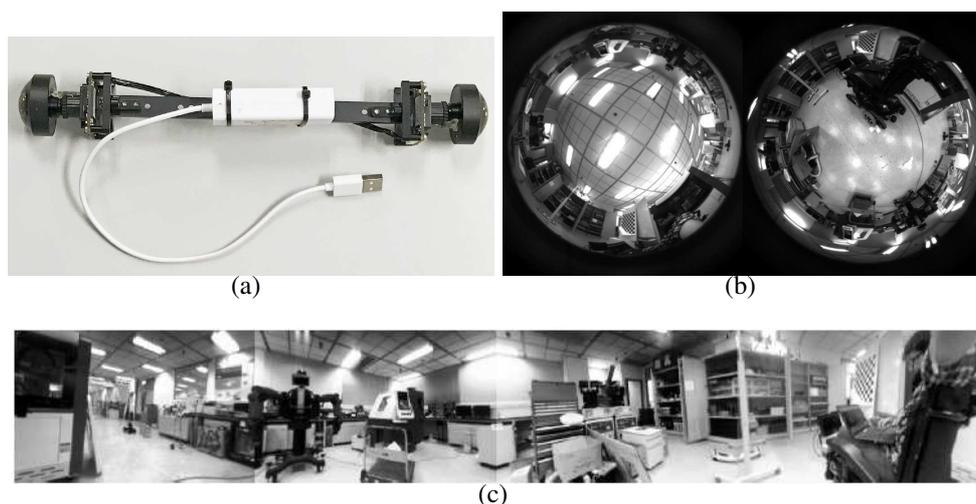


FIGURE 3.2 – Dans [Gao & Shen 2017], (a) deux caméras fisheye observant deux directions opposées sont placées verticalement. (b) Chaque image *dual-fisheye* acquise est découpée en (c) quatre images perspectives, dans l'objectif de réaliser une reconstruction 3D de la scène observée depuis les deux points de vue.

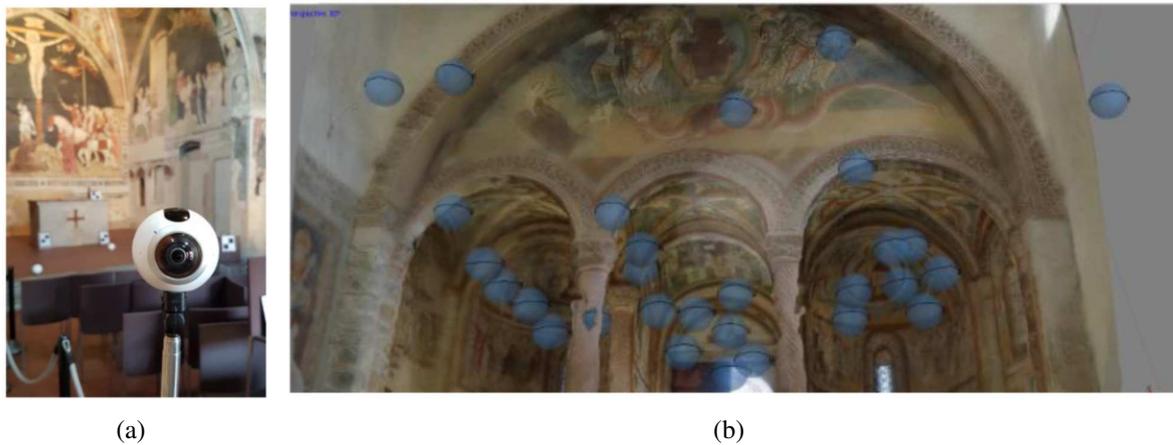


FIGURE 3.3 – Dans [Barazzetti *et al.* 2017], (a) la caméra Samsung 360° est utilisée en intérieur pour réaliser (b) des reconstructions 3D par photogrammétrie. Les sphères visibles indiquent les emplacements des prises de vues.

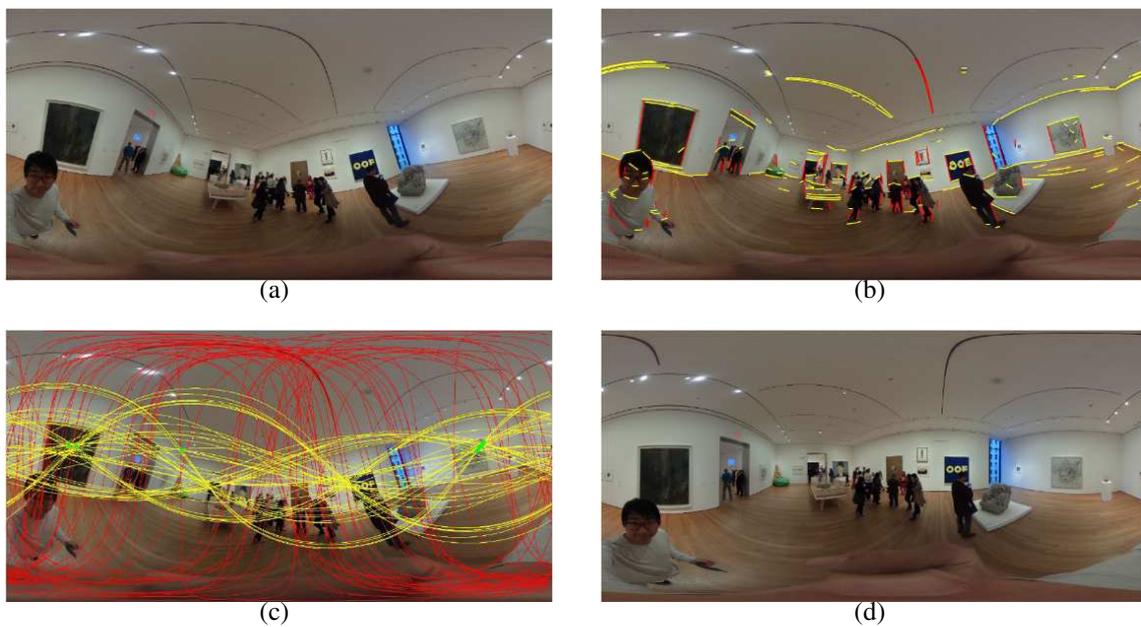


FIGURE 3.4 – Dans [Jung *et al.* 2017], (a) l'orientation d'images équirectangulaires est estimée après détection (b) des lignes horizontales et verticales puis (c) des points de fuite et des grands cercles. (d) L'orientation est finalement corrigée pour améliorer le confort d'exploration.

rées par la méthode proposée sont comparées à celles acquises par lasergrammétrie et à celles obtenues avec les images générées automatiquement par le capteur. La reconstruction par lasergrammétrie sert de référence et une réduction significative de l'erreur moyenne de reconstruction métrique est observée.

Dans [Jung *et al.* 2017], les auteurs corrigent des panoramas en projection équirectangulaire

obtenus à l'aide d'une caméra Ricoh Theta S (figure 3.4(a)). L'orientation de l'image est estimée puis exploitée pour améliorer le confort de visualisation, par exemple, en utilisant un casque de réalité virtuelle pour explorer la scène. La méthode proposée est automatique et ne repose pas sur les capteurs d'inclinaison intégrés à la caméra. Les lignes horizontales et verticales sont détectées afin d'estimer la position des points de fuite ainsi que quelques grands cercles dans les panoramas (figures 3.4(b) et 3.4(c)). À partir de ces données, la méthode permet d'estimer rapidement et avec précision l'orientation de la scène qu'il est ensuite possible de corriger, comme présenté dans la figure 3.4(d).

Dans [Pathak *et al.* 2017], les auteurs s'intéressent à l'exploration dans un casque de réalité virtuelle d'images sphériques obtenues à l'aide d'une caméra *twin-fisheye*. L'observation d'un mouvement apparent distinct pour des objets placés dans une scène à des distances croissantes par rapport à l'observateur est appelée parallaxe de mouvement. L'objectif poursuivi par les auteurs est de reproduire ce phénomène afin de permettre à l'utilisateur de percevoir facilement les changements de profondeur. Pour y parvenir, les auteurs proposent d'utiliser deux images sphériques d'une même scène, obtenues depuis deux emplacements arbitraires (même s'il est précisé qu'une translation verticale est préférable). La profondeur est ensuite obtenue à l'aide d'un algorithme de reconstruction 3D exploitant le mouvement apparent des objets entre les deux prises de vue.

Dans [Caron & Morbidi 2018], les auteurs proposent un gyroscope visuel, omnidirection-

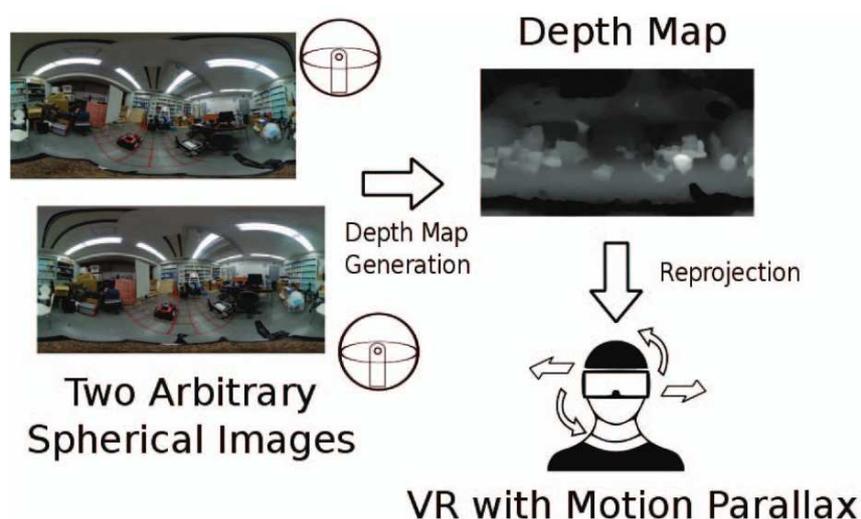


FIGURE 3.5 – Dans [Pathak *et al.* 2017], deux images fisheyes, acquises depuis deux emplacements différents, sont exploitées pour estimer la profondeur pour tous les points visibles depuis les deux points de vue. La profondeur estimée permet d'ajuster la projection du panorama dans un casque de réalité virtuelle et reproduire le mouvement apparent des objets.



FIGURE 3.6 – Dans [Caron & Morbidi 2018], (a) l’orientation d’un drone à voilure fixe est estimée à partir des images acquises par une caméra Ricoh Theta S, dans le but de (b) corriger l’image obtenue.

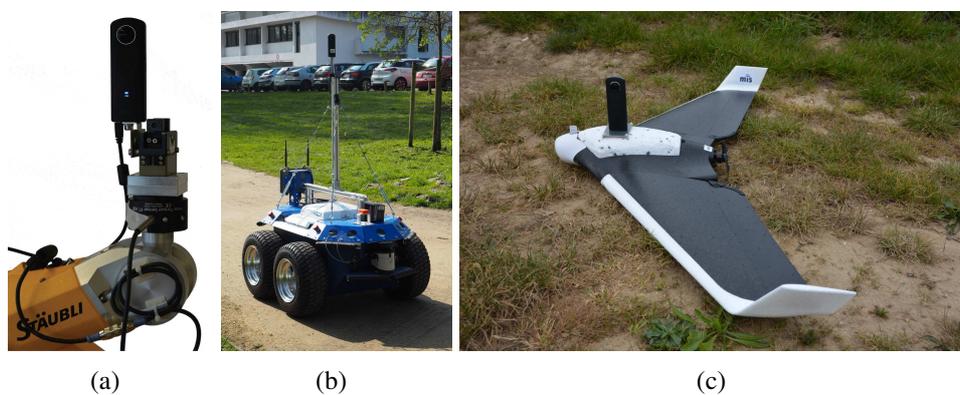


FIGURE 3.7 – Dans [Benseddik *et al.* 2020], la caméra Ricoh Theta S est fixée sur différentes plateformes robotiques ((a) bras manipulateur industriel Stäubli TX60, (b) robot mobile Seekur Jr et (c) drone à voilure fixe Disco) dans le but de produire des jeux de données pour mettre à l’épreuve les méthodes qui traitent les problèmes d’odométrie visuelle ou de SLAM, par exemple.

nel et direct qui exploite les images obtenues avec une caméra Ricoh Theta S. L’objectif de ce gyroscope est d’estimer l’orientation 3D du système sur lequel la caméra est fixée. En comparant une image courante avec une image de référence, il est possible d’estimer les changements à réaliser sur l’image courante pour atteindre la référence. La figure 3.6(a) montre une image acquise par le drone à voilure fixe sur lequel a été fixée la caméra Ricoh Theta S. Après avoir estimé l’orientation de l’appareil, l’image est finalement corrigée (figure 3.6(b)).

Dans [Benseddik *et al.* 2020], les auteurs ont construit plusieurs jeux de données rassemblés sous le nom PanoramIS. Différents capteurs avec de grands champs de vue, dont la caméra Ricoh Theta S, sont installés sur plusieurs plateformes robotiques : des robots mobiles, un drone à voilure fixe et un bras manipulateur industriel (figures 3.7(a), 3.7(b) et 3.7(c)). Les images sont collectées en intérieur et en extérieur dans le but de mettre à l’épreuve les méthodes destinées à évaluer le déplacement d’un robot dans son environnement, par exemple, de l’odométrie visuelle, du SLAM visuel (Simultaneous Localization And Mapping, Cartographie et Locali-

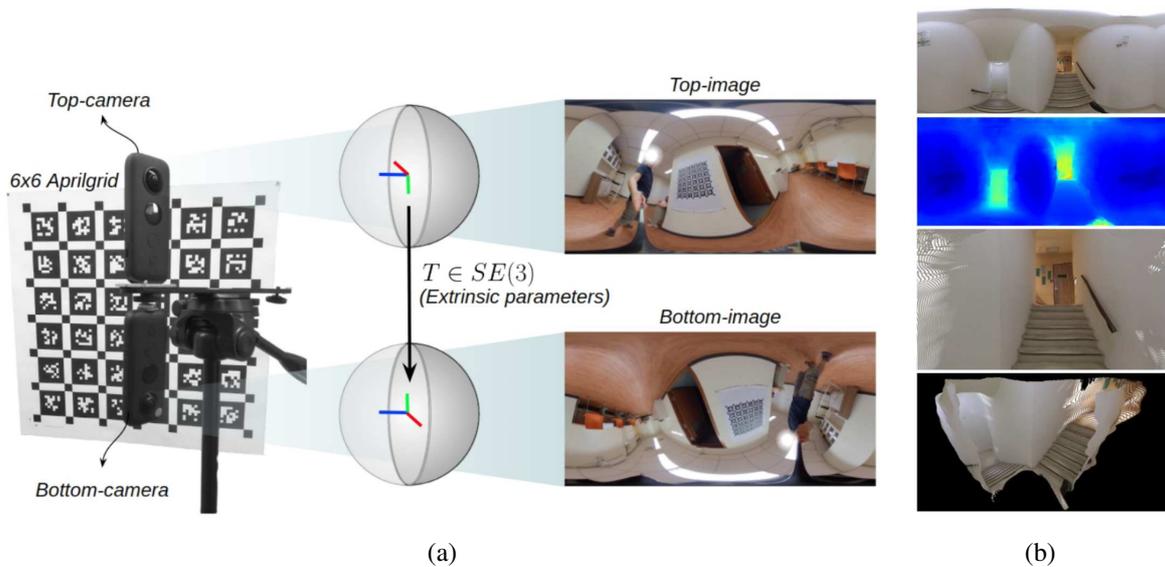


FIGURE 3.8 – Dans [Wang *et al.* 2020], deux caméras Insta360 ONE X capturent deux images équirectangulaires depuis deux points de vue différents. (b) L’estimation de la profondeur pour chaque pixel est réalisée par des réseaux de neurones profonds, spécialisés pour le traitement de ces images sphériques.

sation en Simultané) et des méthodes reposant sur l’apprentissage. L’ensemble des données est accessible publiquement afin d’offrir un élément de comparaison facile à utiliser. Un comparatif avec des jeux de données similaires est proposé afin de bien visualiser les contributions importantes de ces huit jeux de données.

Dernièrement, dans [Wang *et al.* 2020], les auteurs ont utilisé des réseaux de neurones profonds afin d’estimer la profondeur pour chaque pixel dans une paire d’images équirectangulaires obtenues avec deux caméras *twin-fisheye* (figure 3.8(a)). Les distorsions géométriques induites par la projection sont prises en charge par une architecture dédiée. Pour réaliser l’entraînement de leur algorithme, les auteurs ont utilisé exclusivement des données de synthèse. Construire un jeu de données réelles implique de capturer toute la géométrie de la scène, au moins du point de vue des caméras, puis réussir à fusionner l’ensemble, avec la plus faible erreur d’alignement possible. Des résultats quantitatifs sont disponibles sur les jeux de données de synthèse et des résultats qualitatifs sont proposés en exploitant des images réelles, obtenues à l’aide de deux caméras Insta360 ONE X (figure 3.8(b)).

Très récemment, dans [Delmas *et al.* 2021], les auteurs ont proposé un système d’aide à la conduite d’un fauteuil roulant motorisé et médicalisé (Sunrise MedicalQuickie Salsa M²), intégré dans un environnement ROS (Robot Operating System). La combinaison des images obtenues à partir d’une caméra *twin-fisheye* Ricoh Theta S ainsi que des mesures de plusieurs télémètres par temps de vol, disposés tout autour du fauteuil permet de construire une représentation en réalité augmentée de la scène autour du fauteuil, afin de faciliter la manœuvre dans

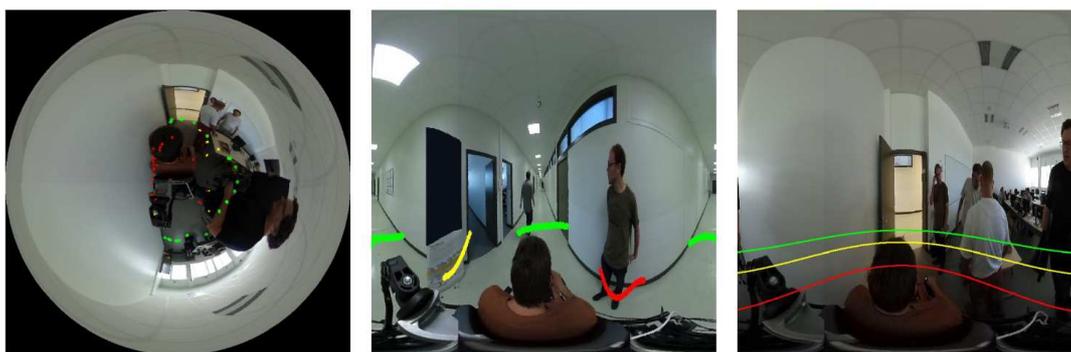


FIGURE 3.9 – Dans [Delmas *et al.* 2021], différentes représentations de l’environnement autour d’un fauteuil médicalisé sont générées à partir d’une image *dual-fisheye*, dans le but de fournir une aide à la conduite.

les environnements difficiles (marche arrière en sortie d’ascenseur, par exemple). La figure 3.9 présente quelques images augmentées avec les données collectées sur l’environnement autour du fauteuil.

3.2 Adaptation de la chaîne de traitements

Dans cette section et par la suite, nous utiliserons la caméra *twin-fisheye* Ricoh Theta V, visible dans la figure 3.10 et accompagnée d’un schéma de son système optique interne. Nous citerons quelques modélisations géométriques proposées dans les approches présentées dans la section précédente. Nous présenterons la modélisation géométrique retenue pour cette caméra *twin-fisheye*. Enfin, nous détaillerons les modifications réalisées sur la chaîne de traitements qui permet d’obtenir la reconstruction 3D, notamment pour tenir compte du large champ de vue des lentilles fisheye.

3.2.1 Modélisation du capteur

Une caméra Ricoh Theta V est composée de deux caméras fisheye qui observent dans deux directions opposées ce qui se trouve autour du capteur. Ces deux caméras internes partagent un système optique compact qui permet de rapprocher les lentilles fisheye, par rapport aux dispositifs construits dans [Jamaluddin *et al.* 2017, Gao & Shen 2017].

Dans la littérature, plusieurs modélisations géométriques spécifiques à ces caméras *twin-fisheye* ont été proposées afin de prendre en compte le large champ de vue disponible et la proximité des deux caméras internes. Ces modèles comportent différents paramètres que nous appellerons paramètres internes : ils englobent les paramètres intrinsèques propres aux deux caméras internes, ainsi que la transformation considérée entre les repères de chaque caméra.



FIGURE 3.10 – Vues de face, de côté et du dessus de la caméra *twin-fisheye* Ricoh Theta V, accompagnées du schéma du système optique interne du boîtier comportant deux lentilles fisheye, deux prismes et deux matrices photosensibles CMOS.

Parmi les applications introduites dans la section précédente, lorsqu'un modèle géométrique est explicitement présenté, nous pouvons distinguer trois approches.

Dans [Barazzetti *et al.* 2017], le modèle proposé dans [Brown 1971] est utilisé : chaque caméra interne est modélisée par huit paramètres, trois de projection et cinq pour les distorsions. La transformation rigide complète (rotation et translation) entre les repères des deux caméras internes est considérée et estimée. Dans [Gao & Shen 2017], les auteurs proposent un modèle géométrique spécifique appliqué à chaque caméra interne. Il est composé de onze paramètres, cinq de projection et six paramètres polynomiaux pour les distorsions. La transformation rigide complète est aussi considérée entre les deux repères. Dans [Jamaluddin *et al.* 2017] puis dans [Caron & Morbidi 2018], le modèle unifié pour les capteurs à projection centrale est utilisé pour modéliser chaque caméra interne. La transformation rigide entre les repères des deux caméras internes est composée de la rotation uniquement : le centre de projection des deux caméras est commun. Le modèle unifié peut être employé avec une caméra équipée de lentilles fisheye (voir [Ying & Hu 2004]), mais il est souvent complété par une modélisation des distorsions puisque ces capteurs ne correspondent pas totalement à des systèmes à projection centrale (voir [Geyer & Daniilidis 2000, Barreto 2006, Kannala & Brandt 2006, Khomutenko *et al.* 2016, Usenko *et al.* 2018]).

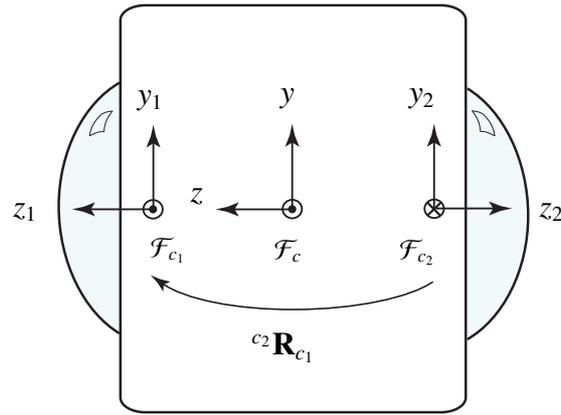


FIGURE 3.11 – En vue du dessus, \mathcal{F}_{c_1} et \mathcal{F}_{c_2} représentent les repères de coordonnées des deux images fisheye, \mathcal{F}_c correspond au repère caméra.

Nous avons choisi de reprendre la formulation présentée dans [Jamaluddin *et al.* 2017, Caron & Morbidi 2018] pour plusieurs raisons : dans [Caracotte *et al.* 2020], nous utilisons le modèle unifié pour modéliser les capteurs employés. La conception de la caméra Ricoh Theta V est très similaire à celle de la Ricoh Theta S. Enfin, l’outil *HySCaS* modifié spécialement pour les travaux présentés dans [Caron & Morbidi 2018] était disponible et a permis de réaliser rapidement l’étalonnage. D’autres modélisations du capteur peuvent fonctionner : la contrainte principale est de réussir à construire une image sphérique à partir des données acquises par le capteur.

En suivant la modélisation proposée dans [Caron & Morbidi 2018], nous associons à chaque lentille de la caméra twin-fisheye un ensemble de paramètres $\mathcal{P}_{c_j} = \{a_{u_j}, a_{v_j}, u_{0_j}, v_{0_j}, \xi_j\}$, $j \in \{1, 2\}$, où a_{u_j} et a_{v_j} correspondent respectivement aux distances focales en pixels sur les axes horizontal et vertical. Le point principal de chaque image est décrit par les coordonnées (u_{0_j}, v_{0_j}) en pixels. Le dernier paramètre de l’ensemble, ξ_j , est la distance sur z entre le centre de la sphère unitaire du modèle unifié et le centre de la deuxième projection perspective pour la lentille fisheye (voir [Barreto 2006, Fig. 2] pour plus de détails). Nous considérons que le vecteur de translation entre les deux repères caméra \mathcal{F}_{c_1} et \mathcal{F}_{c_2} est nul, afin d’imposer l’unicité du point de vue. Le repère caméra \mathcal{F}_c coïncide avec le repère \mathcal{F}_{c_1} . La figure 3.11 présente les trois repères de coordonnées en vue éclatée, pour rendre l’illustration plus accessible, cependant les trois repères ont la même origine. Pour terminer, les paramètres extrinsèques de la caméra *twin-fisheye* sont contenus dans la matrice de rotation ${}^{c_2}\mathbf{R}_{c_1}$ qui permet de passer de \mathcal{F}_{c_1} à \mathcal{F}_{c_2} . Les paramètres internes de la caméra sont contenus dans l’ensemble suivant : $\{\mathcal{P}_{c_1}, \mathcal{P}_{c_2}, {}^{c_2}\mathbf{R}_{c_1}\}$. Plus récemment, d’autres modèles unifiés adaptés aux caméras twin-fisheye ont été proposés dans [Hawary *et al.* 2020, Zhang & Huang 2021].

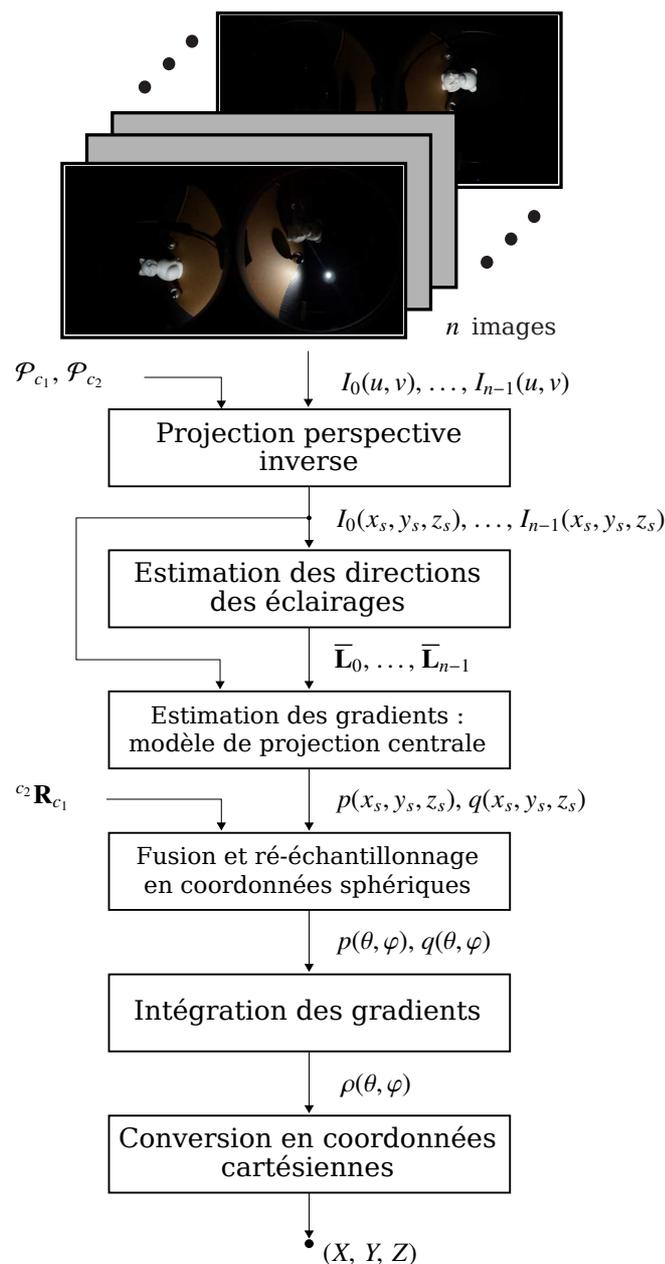


FIGURE 3.12 – Chaîne de traitements pour la reconstruction 3D par stéréophotométrie ajustée pour les capteurs *twin-fisheye*. L’algorithme prend n images *dual-fisheye* en entrée.

3.2.2 Fusion des données

La chaîne de traitements proposée, disponible dans la figure 3.12, tient compte des spécificités de la caméra *twin-fisheye*. Dans un premier temps, les différentes étapes depuis la projection perspective inverse jusqu’à l’estimation des gradients orientés sont réalisées séparément à partir des deux sous-images qui composent l’image *dual-fisheye*. Nous obtenons une estimation des gradients p et q pour chaque pixel utile dans \mathcal{F}_{c_1} et \mathcal{F}_{c_2} . Un pixel utile correspond à un élément

de la scène, par opposition aux pixels qui correspondent au boîtier de la caméra *twin-fisheye*, ou à la région autour de l'image fisheye. Pendant l'étape de projection perspective inverse, chaque intensité $I_i(u, v)$ dans le plan image numérique est mise en correspondance avec son équivalence $I_i(x_s, y_s, z_s)$ puis $I_i(\theta, \varphi)$ sur la sphère unitaire, avec $i \in \{0, 1, \dots, n-1\}$. Pour estimer les gradients orientés, nous utilisons l'équation d'irradiance simplifiée (voir 2.6). Ces gradients sont définis dans trois repères : dans le plan image numérique avec $(p(u, v), q(u, v))$ puis, sur la sphère unitaire avec $(p(x_s, y_s, z_s), q(x_s, y_s, z_s))$ et $(p(\theta, \varphi), q(\theta, \varphi))$ qui correspondent à un même point 3D, décrit en coordonnées cartésiennes puis sphériques. L'étape de ré-échantillonnage spatial est modifiée pour inclure la fusion des données provenant des deux sous-images qui constituent l'image *dual-fisheye* : les gradients orientés sont mélangés sur une seule et unique sphère unitaire, chaque hémisphère correspondant à chacune des deux caméras internes. Pour tenir compte de la région de l'équateur, nous définissons,

$$p(\pi/2, \varphi) = \frac{1}{2} [p(\pi - \pi/\lambda, \varphi) + p(\pi + \pi/\lambda, \varphi)],$$

$$q(\pi/2, \varphi) = \frac{1}{2} [q(\pi - \pi/\lambda, \varphi) + q(\pi + \pi/\lambda, \varphi)],$$

où $\varphi \in [0, 2\pi)$ et $1/\lambda$ correspond au pas d'échantillonnage spatial.

3.2.3 Estimation des distances radiales

Après la fusion, les gradients orientés $(p(\theta, \varphi), q(\theta, \varphi))$ sont définis sur l'entièreté de la sphère unitaire, sauf aux deux pôles qui constituent des singularités. La *contrainte de circularité* définie par $\rho(\theta, \varphi + \Delta\varphi) = \rho(\theta, 0)$ pour $\varphi = 2\pi$ avec $\theta \in (0, \pi)$ est étendue afin de couvrir toute la frontière azimutale ($\Delta\varphi$ correspond au pas d'échantillonnage spatial sur l'axe φ). La *contrainte au Pôle Nord* est dupliquée au second pôle de la sphère afin de tenir compte de la deuxième singularité du système de coordonnées sphériques. Après l'intégration des gradients fusionnés incluant les contraintes supplémentaires, nous obtenons $\ln \rho(\theta, \varphi)$ et finalement, les distances radiales $\rho(\theta, \varphi)$.

3.3 Expérimentations

Dans cette section, nous présentons les résultats de nos expérimentations réalisées dans un premier temps avec des images de synthèse, puis avec des images réelles issues d'une caméra *twin-fisheye*.

3.3.1 Validation en simulation

Cette première partie se concentre sur les images de synthèse, afin de vérifier le bon fonctionnement de l'approche proposée lorsque toutes les hypothèses sur l'éclairage ainsi que sur les propriétés et la géométrie de la surface observée sont respectées.

Le logiciel Blender¹ [Blender Foundation 2018] a été utilisé pour produire deux scènes 3D virtuelles : *Polyèdre Géodésique* et *Chambre*. La vérité terrain contient à nouveau le vecteur normal à la surface observée ainsi que la distance radiale entre le centre optique de la caméra et la surface, pour chaque point (θ, φ) et pour un échantillonnage spatial défini (en général, $\Delta\theta = \Delta\varphi = \pi/\lambda$ avec λ spécifié pour la scène à reconstruire). Nous avons utilisé les paramètres internes de la caméra Ricoh Theta V pour générer des images *dual-fisheye* aussi similaires que possible à celles que la caméra *twin-fisheye* nous permet de capturer (voir la prochaine partie 3.3.2 pour plus de détails). Les éclairages \mathbf{L}_i , $i \in \{0, 1, \dots, 13\}$ ont été positionnés afin que tout point observable de la scène soit éclairé par au moins trois sources différentes. Pour simplifier la simulation et réduire le nombre d'images à considérer, les ombres portées ne sont pas prises en compte. L'effet de quantification est à nouveau simulé en considérant $I_i(\theta, \varphi)$ sur 8 bits : les intensités $I_i(\theta, \varphi) \in [0, 1] \subset \mathbb{R}$ avec $i \in \{0, 1, \dots, 7\}$ sont projetées sur l'ensemble discret $\{0, 1, \dots, 255\}$.

Dans le premier essai, nous avons considéré une géode par triangulation, subdivisée à deux reprises et nommée *Polyèdre Géodésique*. La figure 3.13(a) montre une des 14 images *dual-fisheye* générées en tenant compte de la quantification. Les normales en fausses couleurs visibles dans la figure 3.13(b) ont été déduites en utilisant l'expression suivante (cf. équation (2.6)) :

$$\bar{\mathbf{N}}(\theta, \varphi) = -\frac{1}{\sqrt{p^2 + \left(\frac{q}{\sin \theta}\right)^2 + 1}} \mathbf{R}_z(\varphi) \mathbf{R}_y(\theta) \begin{bmatrix} p \\ \frac{q}{\sin \theta} \\ 1 \end{bmatrix}.$$

Pour finir, la figure 3.13(c) présente la surface externe du *Polyèdre Géodésique* reconstruit avec la méthode présentée dans le chapitre précédent (cf. § 2.3.1).

Dans notre deuxième et troisième essais, nous nous sommes concentrés sur deux phénomènes qui apparaissent lorsque les images sont acquises depuis une caméra réelle : l'impact de la correction gamma et l'impact d'un éclairage non-directionnel sur la qualité de la reconstruction 3D.

La deuxième ligne du Tableau 3.1 présente la moyenne de l'erreur angulaire (MEA) pour une valeur croissante de γ , lorsque nous appliquons (colonne grise) ou non la correction gamma sur les intensités des images. Dans le même tableau, la quatrième ligne montre la moyenne

1. Blender 2.79b : 3D Computer Graphics Software, www.blender.org

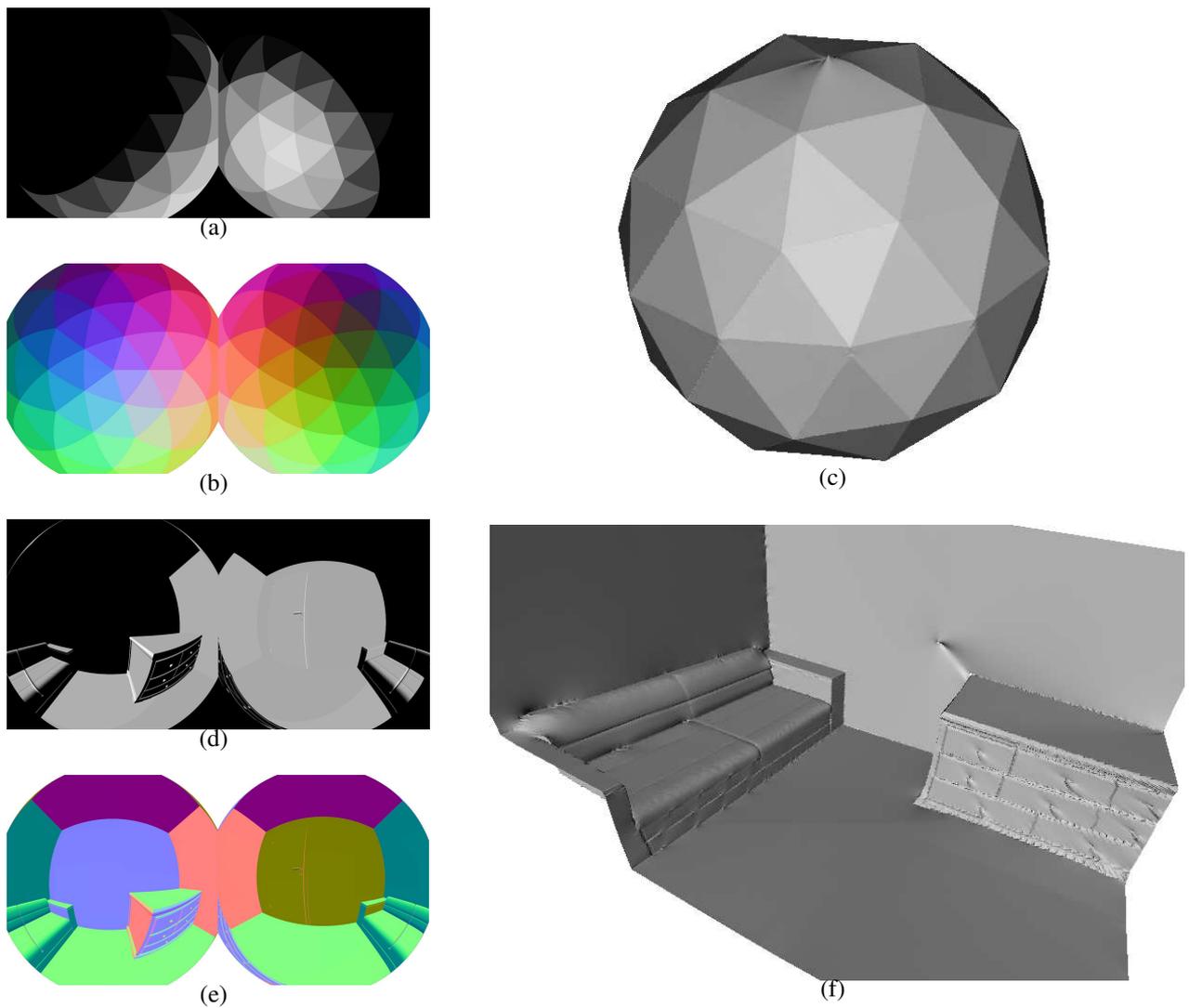


FIGURE 3.13 – Résultats obtenus avec les images de synthèse : *Polyèdre Géodésique* sur la première ligne et *Chambre* sur la seconde. (a),(d) Exemples d’images *dual-fisheye* utilisées en entrée de la chaîne de traitements ; (b),(e) Représentations en fausses couleurs des normales déduites depuis les gradients estimés ; (c),(f) Reconstructions 3D obtenues. Dans (c) *Polyèdre Géodésique* est affiché depuis l’extérieur pour faciliter la visualisation de la reconstruction. Une partie de la reconstruction 3D de *Chambre* dans (f) a été masquée pour distinguer plus facilement l’intérieur de la scène.

de l’erreur angulaire obtenue, en simulation, pour une distance croissante entre une source de lumière ponctuelle et le centre de *Polyèdre Géodésique*. Pour mesurer cette distance, nous avons considéré des multiples entiers du rayon r de la sphère englobant l’objet.

Pour la dernière simulation, nous avons utilisé une scène 3D plus complexe, appelée *Chambre*, qui comporte plusieurs discontinuités de profondeur. Les images *dual-fisheye* de cette scène ont été générées en utilisant le même protocole que la scène *Polyèdre Géodésique*. Une des 14 images *dual-fisheye* de synthèse est visible dans la figure 3.13(d). Les normales

TABLEAU 3.1 – *Polyèdre Géodésique* : pour les lignes 1 et 2, moyenne de l’erreur angulaire (MEA) en degrés pour une valeur croissante de γ : les colonnes grises présentent les résultats obtenus en appliquant la correction gamma. Pour les lignes 3 et 4, MEA pour une distance croissante entre la source de lumière ponctuelle et l’objet observé.

γ	1.2	1.2	1.5	1.5	2.2	2.2
MEA [deg.]	0.75	0.36	2.02	0.37	4.22	0.41
Distance	2 r	5 r	10 r	15 r	20 r	25 r
MEA [deg.]	21.53	3.67	1.54	1.14	0.89	0.70

en fausses couleurs et la reconstruction 3D finale sont représentées respectivement dans les figures 3.13(e) et 3.13(f) : un seul côté de l’intérieur de la scène *Chambre* est montré dans la figure 3.13(f), pour faciliter la visualisation du résultat obtenu.

3.3.2 Évaluation en conditions réelles et contrôlées

Dans cette partie, nous expliquons la méthode utilisée pour réaliser l’étalonnage du capteur et obtenir les paramètres internes de la caméra Ricoh Theta V. Nous présentons ensuite l’algorithme employé pour estimer les directions des éclairages, à partir de miroirs sphériques. Pour conclure, nous présentons puis discutons les résultats obtenus à partir d’images réelles acquises par la caméra Ricoh Theta V.

Matériel utilisé

Pour nos expérimentations, nous avons à nouveau utilisé la chambre noire montée sur la table optique, décrite dans le chapitre précédent. Les figurines, la lampe et les sphères réfléchissantes sont identiques aux précédents tests en conditions réelles. Dans cette série d’expérimentations, la caméra utilisée est la Ricoh Theta V.

Étalonnage du capteur

Avec son système optique interne, la caméra Ricoh Theta V peut capturer tout ce qui se trouve autour d’elle en une image (figure 3.10). Une image *dual-fisheye* générée par la caméra contient deux images (2896×2896 pixels chacune) produites par les deux capteurs photosensibles, placées côte à côte.

Nativement, la caméra enregistre les images *dual-fisheye* uniquement dans les modes de fonctionnement *live* et *video*, avec une définition plus faible. Dans les autres modes de fonctionnement, la caméra produit par défaut une image équirectangulaire en sortie. Pour contour-

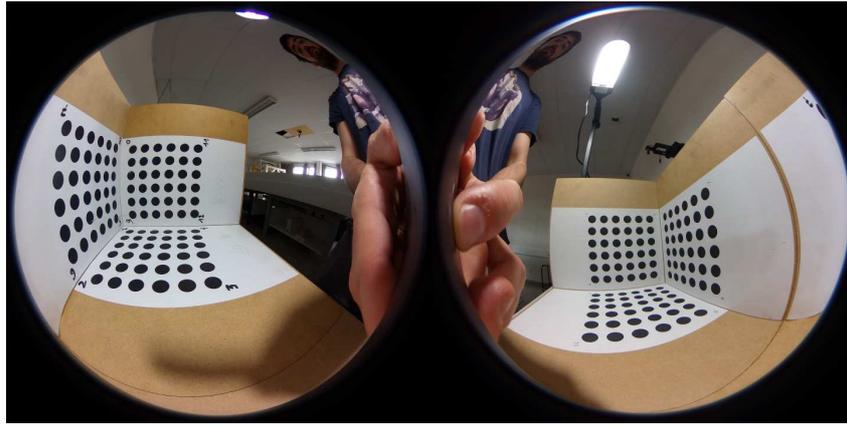


FIGURE 3.14 – Exemple d’image *dual-fisheye* utilisée pour l’étalonnage de la caméra Ricoh Theta V. Les six mires placées sur trois plans orthogonaux sont composées de 36 disques chacune. L’image a subi une rotation de 180° pour améliorer sa lisibilité.

ner ces problèmes, nous avons développé un plugin dédié qui permet de capturer des images *dual-fisheye* en utilisant la définition la plus grande possible (5792×2896 pixels). Le plugin permet aussi d’ajuster des paramètres inaccessibles depuis l’application officielle pour téléphone², disponible au moment des expérimentations, comme le temps d’exposition, l’ouverture du diaphragme, la sensibilité ISO et le facteur de compression JPEG des images enregistrées.

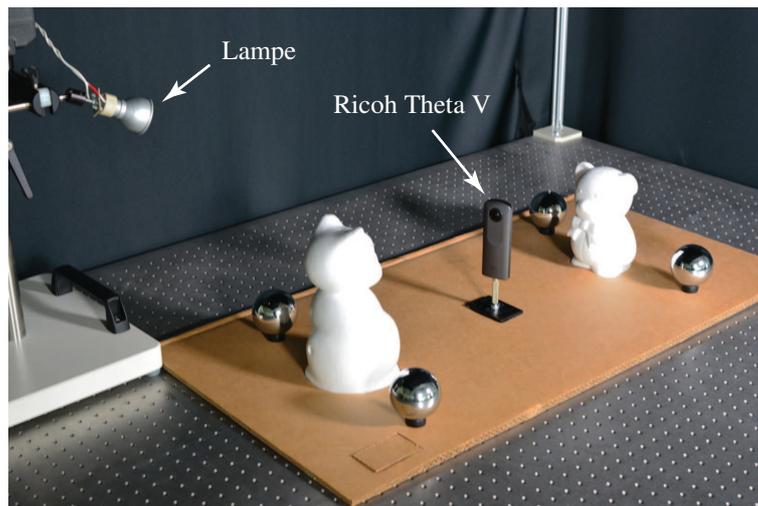
Nous avons modélisé la caméra Ricoh Theta V comme un système stéréo composé de deux caméras fisheye partageant le même centre de projection (le vecteur de translation entre les repères \mathcal{F}_{c_1} et \mathcal{F}_{c_2} est considéré nul). De la même manière que dans [Caron & Morbidi 2018], pour l’étalonnage de la caméra, nous avons utilisé une structure comportant six mires d’étalonnage placées dans deux demi-cubes. Trois images de la structure ont été acquises, caméra tenue en main, dont une est montrée dans la figure 3.14. Après l’étalonnage avec le logiciel *HySCaS* [Caron & Eynard 2011a], nous avons obtenu les paramètres suivants $\mathcal{P}_{c_1} = \{2492.8, 2496.0, 4311.0, 1464.1, 1.866\}$, $\mathcal{P}_{c_2} = \{2484.1, 2505.5, 1457.2, 1466.3, 1.848\}$ et

$${}^{c_2}\mathbf{R}_{c_1} = \begin{bmatrix} -1.000 & -0.001 & 0.025 \\ 0.000 & 1.000 & 0.031 \\ -0.025 & 0.031 & -0.999 \end{bmatrix} \simeq \mathbf{R}_y(\pi).$$

Estimation des directions d’éclairage

Afin d’estimer les directions d’éclairage, nous avons utilisé la méthode présentée dans le chapitre précédent : dans chaque image *dual-fisheye*, nous disposons de quatre sphères réfléchissantes (deux par sous-image fisheye).

2. L’application Android [RICOH THETA](#).



(a)



(b)

FIGURE 3.15 – (a) Matériel utilisé pour les expérimentations : la source de lumière, la caméra Ricoh Theta V, les sphères réfléchissantes dédiées à l’estimation des directions des éclairages ainsi que les objets à reconstruire (*Chat* et *Ourson*) ; (b) Une des images *dual-fisheye* acquises par la caméra et utilisées pour la reconstruction 3D.

Résultats en conditions contrôlées

Un masque binaire appliqué sur les images *dual-fisheye* du *Chat* et de l’*Ourson* permet de concentrer les traitements sur les points dans l’image qui correspondent à ces deux objets. Nous avons exploité les intensités du canal rouge de 11 images de l’*Ourson* et de 12 images du *Chat* pour estimer les gradients et réaliser la reconstruction 3D. Le choix de ce canal s’explique par l’absence de saturation dans les mesures faites par les capteurs concernés.

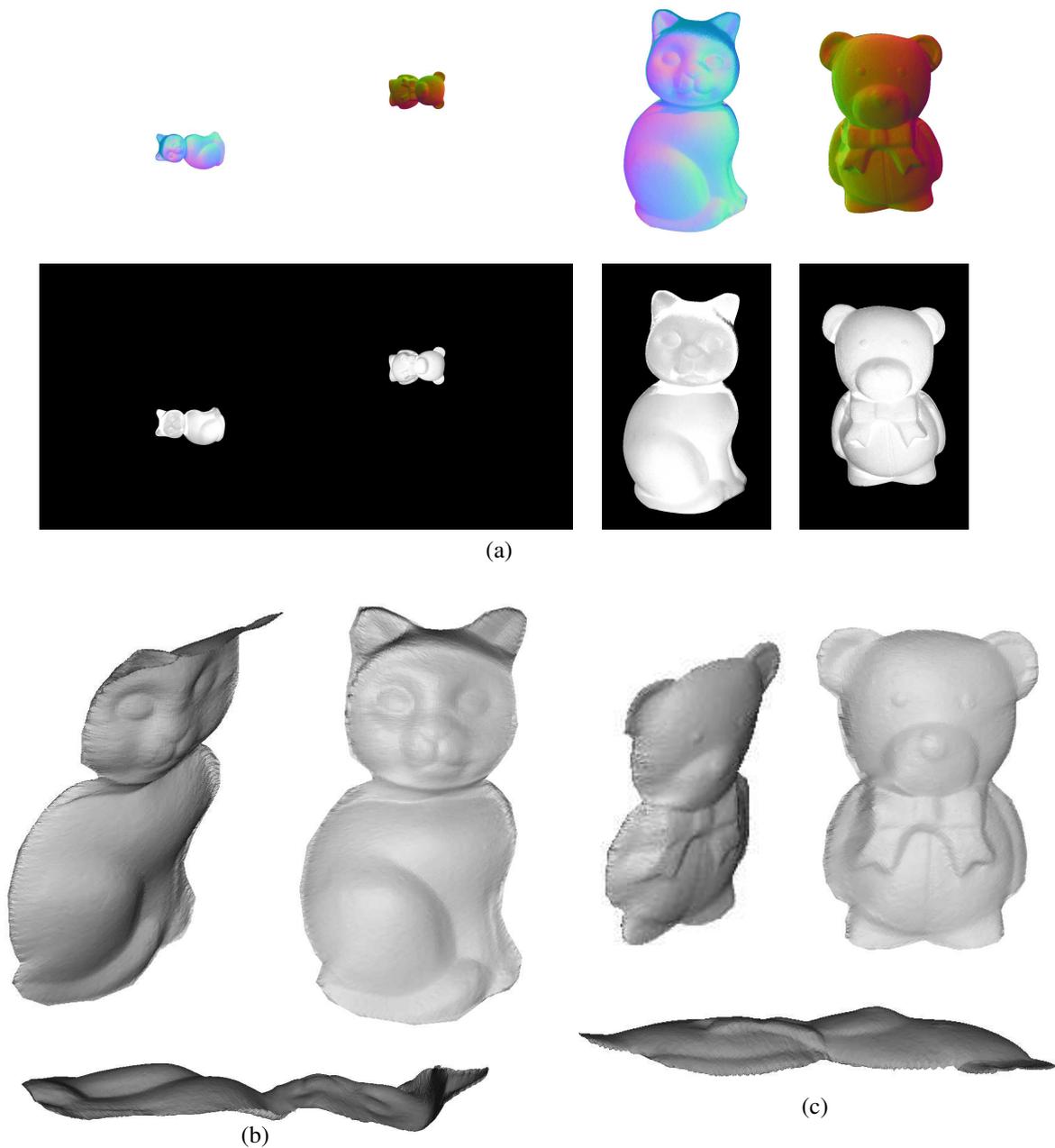


FIGURE 3.16 – Résultats obtenus avec les images réelles : (a) Représentation en fausses couleurs des normales déduites des gradients (en haut) et albedo estimé (en bas). Les résultats sont au format *dual-fisheye* à gauche et une vue rapprochée est proposée à droite. (b),(c) Différentes vues des reconstructions 3D obtenues pour le *Chat* et l’*Ourson*.

Le matériel dédié aux expérimentations, sa mise en place et une des images obtenues avec la caméra Ricoh Theta V sont respectivement visibles dans les figures 3.15(a) et 3.15(b). Les normales en fausses couleurs ainsi que l’albedo estimé (cf. § 2.2.2) des deux objets sont disponibles dans les figures 3.16(a). Les reconstructions 3D obtenues des objets *Ourson* et *Chat*

sont proposées dans les figures 3.16(c) et 3.16(b). Nous pouvons remarquer une déformation au niveau du haut de la tête du *Chat* : les vecteurs normaux estimés $\bar{\mathbf{N}}(\theta, \varphi)$ sont orientés dans la direction opposée, ce qui correspond à une erreur d'estimation des gradients $(p(\theta, \varphi), q(\theta, \varphi))$ dans cette région.

3.3.3 Discussion

Dans cette partie, nous revenons sur la chaîne de traitements et les résultats expérimentaux obtenus précédemment, en particulier à partir de données réelles. Pour finir, nous nous concentrons sur les déformations qui apparaissent sur les reconstructions 3D.

Comme détaillé dans la section 3.2, les gradients orientés obtenus séparément à partir des deux sous-images sont fusionnés, en considérant une rotation pure ${}^{c_2}\mathbf{R}_{c_1}$ entre les repères \mathcal{F}_{c_1} et \mathcal{F}_{c_2} . En principe, cette procédure peut être appliquée à d'autres systèmes multi-caméras, tant que ces deux conditions sont réunies :

- le modèle unifié s'applique à chaque caméra ;
- les paramètres internes du système sont connus ou estimés.

Dans la chaîne de traitements adaptée aux images *dual-fisheye*, l'interpolation des données intervient plus tardivement par rapport à celle proposée dans [Caracotte et al. 2020]. En la réalisant à l'étape "Fusion et ré-échantillonnage spatial" (figure 3.12), nous obtenons au préalable une estimation des gradients orientés $(p(\theta, \varphi), q(\theta, \varphi))$ pour chaque pixel utile. Dans de futurs travaux, nous avons prévu d'exploiter davantage les vecteurs directeurs \mathbf{v} et \mathbf{w} , introduits dans [Caracotte et al. 2020, Th. 2], afin d'atténuer les conséquences du ré-échantillonnage spatial voire de supprimer complètement cette étape.

Les expérimentations réalisées à partir des images de synthèse ont permis de valider le fonctionnement du modèle et de la chaîne de traitements. Les résultats obtenus à partir d'un faible nombre d'images réelles apportent davantage d'éléments en ce sens. Néanmoins, les reconstructions 3D obtenues présentent peu de relief. Au cours des tests avec des images de synthèse, la fonction de réponse de la caméra virtuelle est connue et l'hypothèse d'un éclairage directionnel est pleinement satisfaite. La situation est différente lorsque nous utilisons la caméra Ricoh Theta V : au cours du deuxième test dans la section 3.3.1, en essayant de reproduire virtuellement le comportement d'une caméra réelle, nous avons étudié les effets d'une correction gamma appliquée aux images avant leur prise en compte par la chaîne de traitements.

Les erreurs de reconstruction observées dans les modèles 3D estimés du *Chat* et de l'*Ourson* peuvent être imputées à différents facteurs comme les incertitudes dans l'estimation des éclairages, les ombres portées, etc. Cependant, les principales sources de ces erreurs semblent être l'atténuation de la lumière qui n'est pas prise en compte, l'éclairage qui n'est pas idéal³, la

3. L'éclairage utilisé n'est ni directionnel ni homogène.

fonction de réponse de la caméra Ricoh Theta V qui n'est pas connue et le ré-échantillonnage spatial des gradients orientés. Les résultats reportés dans le Tableau 3.1 et les tests de la section à venir renforcent cette conclusion.

Dans cette partie, nous avons rapidement abordé plusieurs facteurs potentiellement responsables des déformations observées sur les reconstructions 3D du *Chat* et de l'*Ourson*. Dans la prochaine section, nous nous concentrerons davantage sur l'étude de deux phénomènes qui concernent l'éclairage utilisé.

3.4 Analyse des reconstructions à partir de données réelles

Dans cette section, nous nous intéressons spécifiquement aux résultats obtenus à partir de données réelles. Nous proposons deux hypothèses pour expliquer le phénomène d'aplatissement observé sur les reconstructions 3D estimées. Elles seront concentrées sur l'éclairage utilisé pendant les expérimentations. Pour chacune d'elles, nous réaliserons des simulations numériques pour vérifier leur rôle.

3.4.1 Éclairage non directionnel

Dans nos expérimentations exploitant des images réelles, l'hypothèse de l'éclairage directionnel n'est pas respectée : la lampe utilisée ne se trouve pas à une distance suffisante pour obtenir des rayons lumineux parallèles. Pour vérifier l'importance de cette hypothèse, nous avons réalisé deux tests.

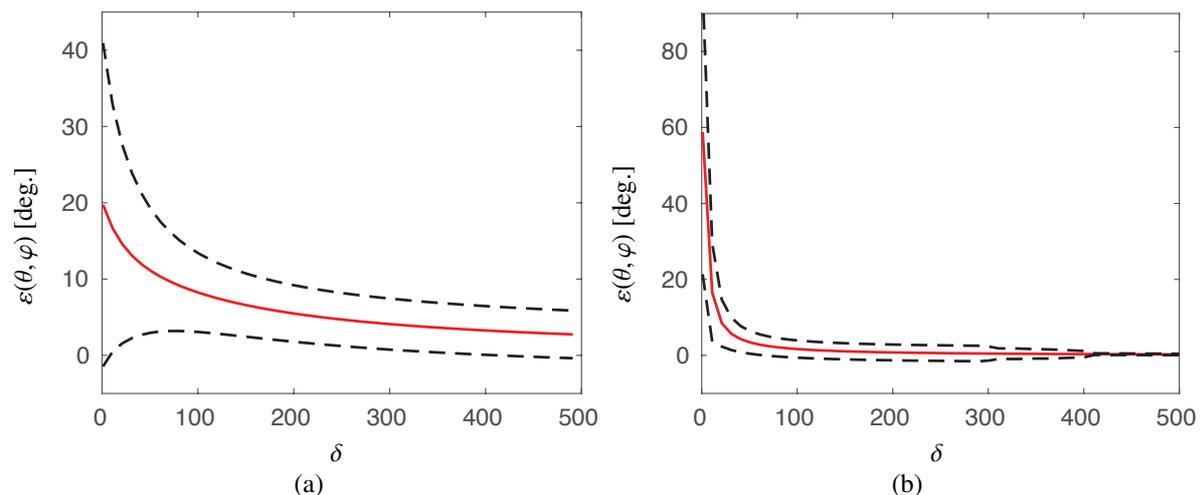


FIGURE 3.17 – Effet de la distance δ (entre la source de lumière ponctuelle et l'objet statique observé) sur l'erreur d'estimation des gradients orientés : la moyenne (ligne rouge) et l'écart-type (± 1 sigma, tirets noirs) de l'erreur $\varepsilon(\theta, \varphi)$, pour (a) le *Stanford Bunny* (taille de l'objet : 100 unités) avec la caméra perspective, et (b) l'*Escalier Hélicoïdal* (taille de l'objet : 30 unités) avec le capteur catadioptrique.

Dans le premier test, nous avons simulé le déplacement progressif d'une source ponctuelle de lumière qui illumine une scène virtuelle, dans le but de quantifier l'erreur d'estimation de $(p(\theta, \varphi), q(\theta, \varphi))$. La caméra virtuelle exploite les paramètres internes estimés de la caméra Ricoh Theta V et la scène observée est la *Chambre*. D'après les résultats numériques reportés dans le Tableau 3.1, plus la source de lumière est éloignée, moins la moyenne de l'erreur angulaire est élevée. Dans le contexte expérimental de la figure 3.15(a), la distance entre l'éclairage et les objets à reconstruire varie entre 6.1 r et 8.7 r, où $r = 23$ cm correspond à la moitié de la distance la plus courte entre le *Chat* et l'*Ourson*.

Dans le deuxième test, nous avons simulé les deux capteurs du chapitre 2 : une caméra perspective et un capteur catadioptrique, tous deux simulés à partir des paramètres intrinsèques estimés dans le chapitre 2. Le même protocole est appliqué sur l'éclairage. Nous avons exploité à nouveau le *Stanford Bunny* avec la caméra perspective et l'*Escalier Hélicoïdal* avec le capteur catadioptrique. Les figures 3.17(a),(b) montrent la moyenne (en rouge) et l'écart-type ($\pm 1\sigma$, tirets noirs) de l'erreur $\varepsilon(\theta, \varphi)$ respectivement pour le *Stanford Bunny* et l'*Escalier Hélicoïdal*, en fonction de la distance δ entre l'objet observé et la source de lumière ponctuelle. Dans ce test, la diagonale de la boîte englobante du premier objet mesure 100 unités de longueur et la diagonale pour le deuxième objet vaut 30 unités. D'après la figure 3.17(b), nous pouvons observer qu'à partir d'une distance $\delta > 450$ (c'est-à-dire, 15 fois la diagonale de la boîte englobante de l'*Escalier Hélicoïdal*), la source ponctuelle de lumière peut être considérée comme à l'infini : à partir de cette distance, il n'y a plus d'effet notable sur l'erreur $\varepsilon(\theta, \varphi)$.

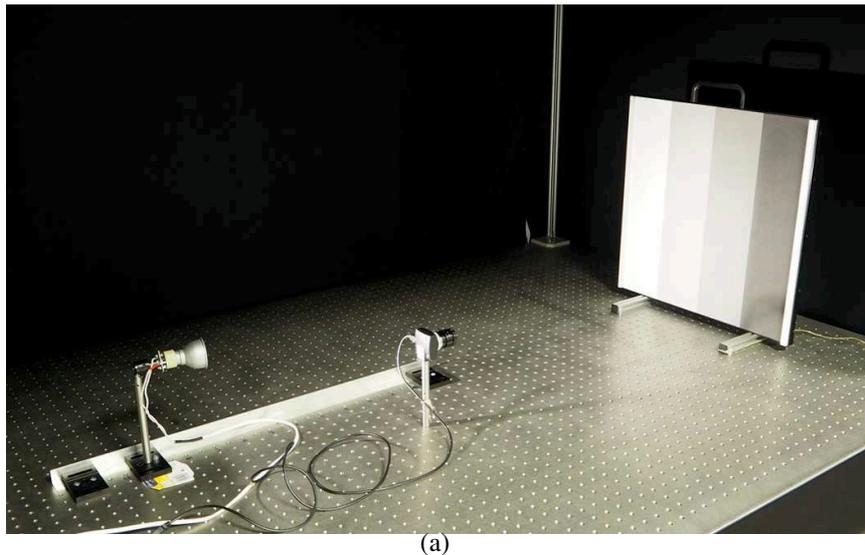
Par la suite, nous avons prévu de nous intéresser à la résolution du problème de stéréophotométrie impliquant des éclairages non directionnels ("near-field lights"), en adaptant les travaux existants aux systèmes de vision omnidirectionnelle (voir [Mecca *et al.* 2014, Logothetis *et al.* 2017]).

3.4.2 Atténuation de l'éclairage

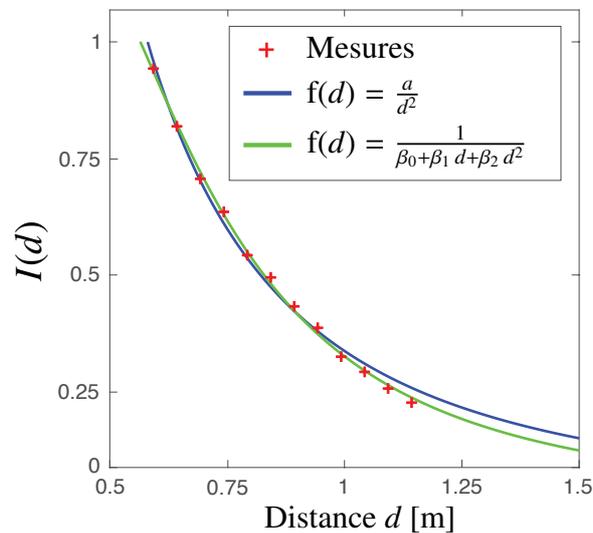
Dans les images de synthèse générées pour valider le fonctionnement de l'approche, nous n'avons pas tenu compte de l'atténuation de la lumière. Au cours des expérimentations de la section 3.3.2, la source de lumière a été placée à une distance d'environ 1 m des objets à reconstruire. Nous pouvons aussi noter que cette source de lumière n'est pas spatialement homogène : elle émet un cône de lumière avec un angle d'ouverture de 36° qui produit un disque central davantage lumineux sur les objets éclairés.

Pour quantifier l'atténuation de la lumière dans le contexte de nos expérimentations, nous avons placé une caméra perspective IDS UI-1245LE-C-HQ dans la chambre noire, de manière à ce qu'elle observe un panneau vertical, éclairé par la lampe (figure 3.18(a)). L'axe optique de la caméra et la direction pointée par la hauteur du cône lumineux sont orthogonaux avec le plan dé-

crit par le panneau. Nous avons déplacé la lampe le long d'un rail gradué avec un pas de 50 mm, dans l'intervalle de distance $d \in [0.593, 1.143]$ m. Pour chacune des 12 graduations, nous avons capturé une image du panneau éclairé. À partir d'une région rectangulaire définie arbitrairement dans les images acquises, nous avons calculé l'intensité moyenne des pixels concernés, afin de filtrer le bruit. La région est contenue dans une des bandes du panneau. Les intensités moyennes et normalisées entre 0 et 1 sont représentées par des croix rouges dans la figure 3.18(b), en fon-



(a)



(b)

FIGURE 3.18 – (a) Matériel utilisé : à gauche, la lampe coulisse sur un rail gradué, au centre, la caméra perspective fixée à la table optique et à droite, le panneau. (b) Fonctions de l'atténuation de la lumière (éq. (3.1) en bleu, éq. (3.2) en vert) estimées à partir de la série de mesures (approximation au sens des moindres carrés). Les douze mesures d'intensité dans l'image en fonction de la distance sont représentées en rouge. Pour renforcer la lisibilité, les intensités dans l'image définies dans l'intervalle $[0, 255]$ ont été normalisées.

tion de la distance d croissante. Dans le cadre de nos expérimentations dans la section 3.3.2, la lampe se trouvait à une distance comprise dans l'intervalle $[0.75, 1.1]$ m. De ce fait et d'après la courbe obtenue, l'effet lié à l'atténuation de la lumière n'est pas négligeable (réduction de l'intensité supérieure à 25%).

Le modèle mathématique précédemment proposé peut être modifié pour intégrer l'atténuation de la lumière. Pour rappel (voir la partie 2.2.2), l'équation de l'irradiance pour les images sphériques selon les vecteurs directeurs $\mathbf{v} = [v_1, v_2]^T$ et $\mathbf{w} = [w_1, w_2]^T$ est :

$$I(\theta, \varphi) = \frac{\alpha(\theta, \varphi) \mathbf{L}^T}{\|\mathbf{L}\| \sqrt{(v_1 p - w_1 q)^2 + \left(\frac{w_2 q - v_2 p}{\sin \theta}\right)^2 + \det(\mathbf{A})^2}} \mathbf{R}_z(\varphi) \mathbf{R}_y(\theta) \begin{bmatrix} v_1 p - w_1 q \\ \frac{w_2 q - v_2 p}{\sin \theta} \\ -\det(\mathbf{A}) \end{bmatrix},$$

où \mathbf{L} correspond à la direction de l'éclairage et $\alpha(\theta, \varphi) \in [0, 1]$ est l'albedo au point (θ, φ) . Pour tenir compte de l'atténuation de la lumière, nous supposons que l'intensité réelle $I^f(\theta, \varphi)$ au point (θ, φ) est,

$$I^f(\theta, \varphi) = f(d) I(\theta, \varphi),$$

où $f(d)$ est une *fonction de l'atténuation de la lumière* et d est la distance entre la source de lumière et un point sur la surface \mathcal{S} définie dans la partie 2.1.1. Plusieurs choix pour $f(d)$ sont possibles : dans [Prados & Faugeras 2005, Liao *et al.* 2007], les auteurs considèrent une loi en carré inverse, c'est-à-dire,

$$f(d) = \frac{a}{d^2}, \quad (3.1)$$

où a est une constante positive (en général, $a = 1$). Dans la figure 3.18(b), l'estimation du paramètre a vaut 0.3329, après approximation au sens des moindres carrés. Un autre exemple, issu du monde de l'infographie (cf. le modèle d'illumination dans la spécification d'OpenGL), nous donne :

$$f(d) = \frac{1}{\beta_0 + \beta_1 d + \beta_2 d^2}, \quad (3.2)$$

où β_0, β_1 et β_2 sont trois paramètres positifs ou nuls. Le paramètre β_0 vaut généralement 1 afin de garantir un dénominateur supérieur ou égal à 1. Au niveau de la source de lumière ($d = 0$), l'atténuation de la lumière n'a aucun effet ($f(d) = 1$). Dans la figure 3.18(b), l'estimation des paramètres nous donne $\beta_0 = 1.434$, $\beta_1 = -3.971$ et $\beta_2 = 5.671$. Parmi les deux modèles présentés, la figure 3.18(b) montre que la fonction (3.1) (représentée en vert) est la plus ajustée aux mesures réalisées.

D'autres fonctions d'atténuation en fonction de la distance d existent. D'après la loi empirique de Beer-Lambert-Bouguer, lorsque la lumière se déplace dans un milieu homogène et diffusant, elle est uniformément atténuée selon une loi exponentielle décroissante de la forme

$\exp(-\mu d)$, où μ est le coefficient d'atténuation du milieu traversé, exprimé en quantité de lumière absorbée par unité de distance (voir [Fan *et al.* 2016]).

En suivant le développement présenté dans la section 2.2.1, nous considérons n éclairages directionnels $\mathbf{L}_0, \mathbf{L}_1, \dots, \mathbf{L}_{n-1}$ placés aux distances respectives d_0, d_1, \dots, d_{n-1} de la surface \mathcal{S} . En utilisant l'intensité réelle expérimentée précédemment, nous notons les intensités $I_0^f(\theta, \varphi), I_1^f(\theta, \varphi), \dots, I_{n-1}^f(\theta, \varphi)$, en fonction de l'éclairage. En divisant les intensités de la i -ème image par celle de la k -ème, et en excluant toutes les valeurs nulles, nous obtenons :

$$\frac{I_i^f(\theta, \varphi)}{I_k^f(\theta, \varphi)} = \frac{f(d_i) I_i(\theta, \varphi)}{f(d_k) I_k(\theta, \varphi)} = \frac{f(d_i)}{f(d_k)} \frac{\|\mathbf{L}_k\| \mathbf{L}_i^T \mathbf{R}_z(\varphi) \mathbf{R}_y(\theta) \left[v_1 p - w_1 q, \frac{w_2 q - v_2 p}{\sin \theta}, -\det(\mathbf{A}) \right]^T}{\|\mathbf{L}_i\| \mathbf{L}_k^T \mathbf{R}_z(\varphi) \mathbf{R}_y(\theta) \left[v_1 p - w_1 q, \frac{w_2 q - v_2 p}{\sin \theta}, -\det(\mathbf{A}) \right]^T}.$$

En développant le numérateur et le dénominateur, puis en factorisant par p et q , nous pouvons ré-écrire le rapport photométrique (cf. équation (2.8)),

$$A_{i,k}^f p + B_{i,k}^f q + C_{i,k}^f = 0, \quad i, k \in \{0, 1, \dots, n-1\},$$

où

$$\begin{aligned} A_{i,k}^f &= \left[I_i^f f^{-1}(d_i) \|\mathbf{L}_i\| \langle \mathbf{L}_k, \mathbf{e}_\theta \rangle - I_k^f f^{-1}(d_k) \|\mathbf{L}_k\| \langle \mathbf{L}_i, \mathbf{e}_\theta \rangle, \right. \\ &\quad \left. - \frac{1}{\sin \theta} \left(I_i^f f^{-1}(d_i) \|\mathbf{L}_i\| \langle \mathbf{L}_k, \mathbf{e}_\varphi \rangle - I_k^f f^{-1}(d_k) \|\mathbf{L}_k\| \langle \mathbf{L}_i, \mathbf{e}_\varphi \rangle \right) \right] \mathbf{v}, \\ B_{i,k}^f &= - \left[I_i^f f^{-1}(d_i) \|\mathbf{L}_i\| \langle \mathbf{L}_k, \mathbf{e}_\theta \rangle - I_k^f f^{-1}(d_k) \|\mathbf{L}_k\| \langle \mathbf{L}_i, \mathbf{e}_\theta \rangle, \right. \\ &\quad \left. - \frac{1}{\sin \theta} \left(I_i^f f^{-1}(d_i) \|\mathbf{L}_i\| \langle \mathbf{L}_k, \mathbf{e}_\varphi \rangle - I_k^f f^{-1}(d_k) \|\mathbf{L}_k\| \langle \mathbf{L}_i, \mathbf{e}_\varphi \rangle \right) \right] \mathbf{w}, \\ C_{i,k}^f &= -\det(\mathbf{A}) \left(I_i^f f^{-1}(d_i) \|\mathbf{L}_i\| \langle \mathbf{L}_k, \mathbf{e}_\rho \rangle - I_k^f f^{-1}(d_k) \|\mathbf{L}_k\| \langle \mathbf{L}_i, \mathbf{e}_\rho \rangle \right). \end{aligned}$$

Dans les trois précédentes équations, $f^{-1}(d_i) = 1/f(d_i)$ et $I_i^f(\theta, \varphi)$ simplement noté I_i^f permettent d'alléger les notations. Dans l'hypothèse où les distances d_i et d_k sont parfaitement connues, il est possible de compenser l'effet de l'atténuation : les termes $A_{i,k}^f, B_{i,k}^f$ et $C_{i,k}^f$ seraient réduits à leur forme respective $A_{i,k}, B_{i,k}$ et $C_{i,k}$. Nous obtiendrions alors l'équation (2.9). Cependant, ces distances d_i et d_k ne sont pas connues en pratique et leur estimation robuste n'est pas simple.

Afin de quantifier l'effet de l'atténuation de la lumière sur l'estimation des gradients orientés p et q , et par propagation, sur la reconstruction 3D finale, nous avons réalisé quelques simulations. Pour simplifier l'expérience, notre objectif est de reconstruire la surface d'une sphère unitaire placée sur l'axe optique d'une caméra perspective virtuelle. Huit sources directionnelles de lumière ont été placées afin d'éclairer uniformément la partie visible de la sphère. La distance entre la surface et l'éclairage a été générée aléatoirement : pour chaque source, nous avons perturbé une distance nominale d^* avec un bruit gaussien, dont la moyenne est nulle et

la variance est définie par σ^2 . Plusieurs valeurs de d^* ont été testées et nous présentons dans la figure 3.19 les résultats pour la valeur $d^* = 0.8$. L'atténuation a été simulée en utilisant la fonc-

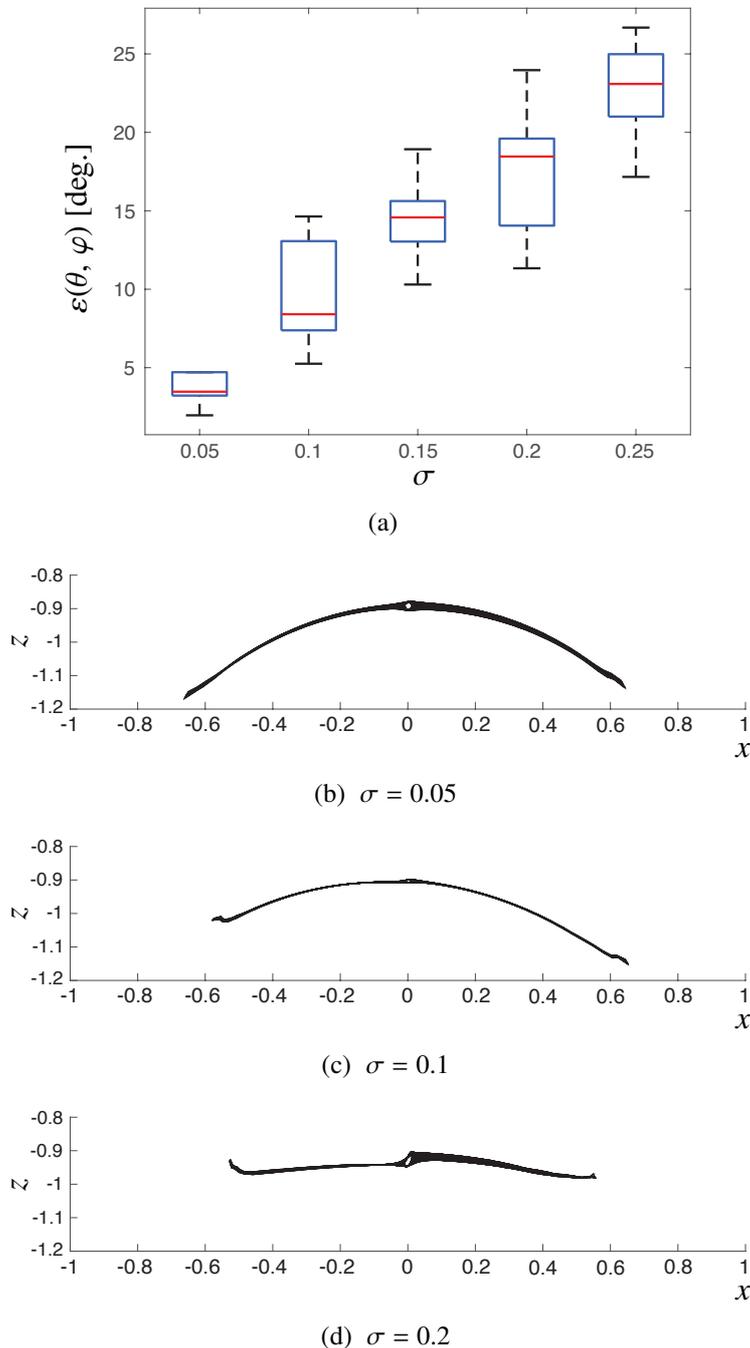


FIGURE 3.19 – Effet de l'atténuation de l'éclairage sur la reconstruction 3D d'une sphère unitaire virtuelle (résultats numériques pour l'équation (3.2)) : (a) Boxplot de la MEA $\varepsilon(\theta, \varphi)$ (en degrés), pour un bruit σ d'amplitude croissante sur la distance entre la sphère et les sources de lumière : 10 tirages aléatoires ont été générés pour chaque valeur de σ ; (b)-(e) Vue en coupe selon le plan xz d'une portion de la reconstruction 3D de la sphère unitaire, pour un bruit σ d'amplitude croissante.

tion (3.2) ainsi que les coefficients β_0 , β_1 et β_2 estimés à partir des mesures. La figure 3.19(a) montre la moyenne de l'erreur angulaire en fonction du bruit appliqué sur la distance (en ajustant σ). Les indicateurs statistiques ont été calculés à partir de 10 essais, pour chaque valeur de σ . Les boxplots montrent que plus la dispersion spatiale des éclairages autour de d^* est grande, plus l'erreur d'estimation des gradients orientés p et q est importante. Nous observons aussi cet effet sur les reconstructions 3D visibles dans les figures 3.19(b)-(d) : chaque figure présente une coupe au niveau du plan (xz) de la surface reconstruite après un tirage aléatoire pour des valeurs croissantes de σ . En effet, plus le bruit est important, plus la surface semble se déformer et s'aplatir.

L'atténuation peut expliquer en partie l'aplatissement observé dans les reconstructions du *Chat* et de l'*Ourson*, dans les figures 3.16(b) et 3.16(c). Bien que la modélisation mathématique de l'atténuation de la lumière semble relativement simple à intégrer, obtenir une estimation précise des distances entre la source de lumière et chaque point de la surface éclairée reste un problème à part entière. Au cours de nos expérimentations, nous avons placé au moins deux des sphères réfléchissantes. Initialement prévues pour assurer une redondance en cas d'occultation par ombre portée, ces sphères peuvent servir à estimer la position de la source de lumière dans le repère caméra, par triangulation.

Dans de futurs travaux, nous comptons nous inspirer de méthodes existantes, par exemple [Liao *et al.* 2007], afin de tenir compte de l'atténuation sans recourir à une estimation précise de la distance entre la surface observée et l'éclairage. Un de nos objectifs finaux est de parvenir à utiliser cette méthode de reconstruction 3D dans le cadre de la sauvegarde du patrimoine culturel [Mouaddib *et al.* 2019].

Conclusion

Dans ce troisième chapitre, nous avons présenté une extension de la méthode introduite dans le chapitre 2 afin d'exploiter les données d'une caméra *twin-fisheye*. Les modifications apportées à la modélisation géométrique ainsi qu'à la chaîne de traitements ont été détaillées. Des améliorations ont été proposées, notamment pour retarder la phase d'interpolation et pour bénéficier le plus possible des images *dual-fisheye* (contrainte de circularité, contraintes polaires détaillées dans la partie 3.2). Un ensemble d'expérimentations à partir d'images de synthèse a permis de valider le fonctionnement de l'extension proposée. Un essai à partir d'images réelles provenant d'une caméra Ricoh Theta V a donné un aperçu des résultats qu'il est possible d'obtenir à partir de ces données. Les déformations et l'aplatissement apparent des deux figurines de test ont été étudiés. Pour expliquer ces phénomènes, deux hypothèses ont été formulées et testées : l'utilisation d'éclairages non directionnels ainsi que l'atténuation de la lumière, qui semble être la source principale des erreurs de reconstruction observées. Dans la conclusion générale, nous nous concentrerons sur quelques améliorations de la méthode en cours d'étude ainsi que d'autres approches connexes.

Conclusion générale

Tout au long de ce manuscrit, nous avons présenté plusieurs contributions qui visent à résoudre le problème de stéréophotométrie en considérant le modèle unifié d'une caméra à projection centrale, ainsi que différents scénarios impliquant des capteurs simulés et réels avec un champ de vue de plus en plus grand (caméra perspective, capteur catadioptrique, caméra *twin-fisheye*).

Au cours du premier chapitre, nous avons présenté les notions nécessaires permettant de comprendre la formalisation du problème de stéréophotométrie, ainsi que quelques méthodes qui tentent de le résoudre afin d'obtenir une reconstruction 3D d'une surface observée. Dans le deuxième chapitre, nous avons présenté une nouvelle équation de l'irradiance adaptée aux images sphériques que nous avons obtenu à l'aide du modèle unifié pour les capteurs à projection centrale. Nous avons détaillé une chaîne de traitements modulaire et dimensionnée pour ces capteurs omnidirectionnels. La validation expérimentale à partir d'images de synthèse a permis de vérifier le fonctionnement de l'approche, puis les essais à partir d'images réelles ont permis de mettre à l'épreuve la méthode et d'identifier ses limites. Dans le troisième chapitre, nous avons exploité l'équation de l'irradiance proposée ainsi que la nouvelle chaîne de traitements afin de prendre en charge les images acquises par les caméras *twin-fisheye*. Pour finir, nous nous sommes concentrés sur les limites de l'approche et les causes des artéfacts observés dans les reconstructions à partir d'images réelles.

Afin de perfectionner les travaux présentés et, en lien avec d'autres approches apparues dans la littérature pendant la même période, nous proposons ici un ensemble de pistes d'amélioration réparties autour de trois axes : l'enrichissement du modèle proposé, l'amélioration de la validation expérimentale ainsi que quelques approches qui prennent davantage de recul par rapport aux éléments présentés.

Améliorations directes du modèle proposé

Suppression du ré-échantillonnage spatial

Dans la chaîne de traitements introduite dans le chapitre 2, pendant l'étape de ré-échantillonnage spatial, nous avons choisi un pas constant au niveau des angles d'azimuth et d'élévation ($\Delta\theta$ et $\Delta\varphi$), ce qui nous a permis de construire une grille rectangulaire attendue par l'algorithme d'intégration des normales. Le passage par cette étape requiert une interpolation qui altère les mesures d'intensité ou les gradients (voir chapitre 3). De plus, la répartition des points sur la sphère n'est pas spatialement uniforme, d'un point de vue géodésique. Dans de futurs travaux, nous prévoyons d'exploiter davantage les vecteurs \mathbf{v} et \mathbf{w} , introduits dans

le Théorème 2, afin de définir des grilles plus proches des données utilisées et ainsi permettre la suppression de la phase d'interpolation. De façon analogue, nous pourrions définir une répartition uniforme des points autour de la sphère en exploitant une des nombreuses méthodes disponibles dans la littérature comme, par exemple, la construction d'un *Quaternary Triangular Mesh* (voir [Horn 1984, Sect. V] et [Guan & Smith 2017]), ou d'une grille géodésique (voir [Zhao *et al.* 2015, Adarve & Mahony 2017]).

D'autres systèmes de coordonnées curvilignes

Dans le chapitre 2, avant de présenter l'équation d'irradiance adaptée aux images sphériques, nous avons introduit les coordonnées sphériques sur lequel s'appuient tous les développements réalisés. Ce système de coordonnées est adapté au modèle de projection centrale. Dans l'idée de pouvoir prendre en charge les caméras catadioptriques à projection non-centrale ou quasi-centrale [Gonçalves & Araújo 2009, Schönbein *et al.* 2014] ou les caméras fisheye à très forte distorsion [Khomutenko *et al.* 2016, Usenko *et al.* 2018], nous avons essayé de répéter la même démarche en utilisant d'autres systèmes de coordonnées orthogonales (les coordonnées sphériques étirées [Happel & Brenner 1983, Appendix A.17]). Les premiers développements mathématiques sont prometteurs mais nécessitent davantage d'efforts avant de pouvoir être implémentés.

Modélisation de l'éclairage

Dans le modèle présenté, nous supposons que l'éclairage est *directionnel* et *uniforme*. Cette hypothèse n'est pas respectée dans les expérimentations réalisées à partir d'images réelles, malgré la distance entre la source de lumière et les objets à reconstruire ainsi que l'étalonnage réalisé. Afin de pallier cet inconvénient, nous comptons nous inspirer des méthodes de stéréophotométrie non calibrée, comme, par exemple, [Mecca *et al.* 2014, Quéau *et al.* 2017, Logothetis *et al.* 2017].

Enrichissement de la validation expérimentale

Images de synthèse photoréalistes

Les images de synthèse générées dans les chapitres 2 et 3, bien que très simples, permettent de valider le fonctionnement des approches présentées en travaillant dans des conditions idéales. Ces données de synthèse nous permettent aussi de quantifier les performances de la méthode en calculant l'erreur angulaire lorsque les normales sont estimées. Il n'est pas aisé de construire un jeu de données offrant une vérité-terrain fiable et, de ce fait, l'évaluation des résultats de nos expérimentations à partir de données réelles n'est que qualitative. Afin de s'approcher des

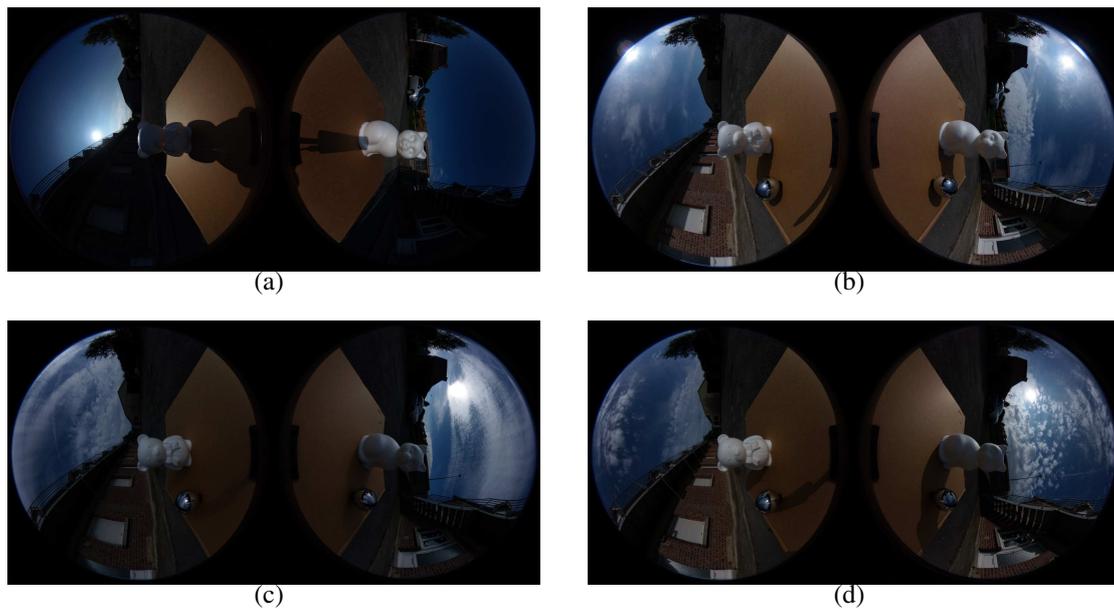


FIGURE 3.20 – Prises de vue de la scène en extérieur à différents moments de la journée : fin de matinée (a), (b) après-midi avec le disque solaire visible dans les deux images fisheye, (c) après-midi avec occultation partielle du disque solaire et (d) fin de l’après-midi.

conditions réelles tout en conservant une vérité-terrain fiable, nous envisageons d’utiliser le moteur de rendu *OmniSCV*, introduit dans [Berenguel-Baeta *et al.* 2020], pour générer des images *dual-fisheye* photoréalistes.

Reconstruction d’un environnement extérieur

En utilisant la caméra Ricoh Theta V, un jeu de données a été construit à partir d’une scène en extérieur sur une période de sept heures (de 10h00 à 17h00). La figure 3.20 présente quatre images à des moments différents de la journée. L’exploitation de ce jeu de données présente de nouveaux défis par rapport aux expérimentations en conditions contrôlées, liés (entre autre) à la lumière ambiante, l’éclairage variable (présence ou non de nuages) ou les matériaux non lambertiens qui composent la scène. Toutefois, les sphères réfléchissantes et la présence du disque solaire dans l’image peuvent être utilisés, en principe, pour estimer la direction principale de l’éclairage. Pour traiter ces données, nous pensons nous appuyer sur les méthodes existantes exploitant des données, acquises à partir de webcams en extérieur (voir [Abrams *et al.* 2012, Ackermann *et al.* 2012]).

Approches alternatives

Méthodes globales

Plutôt que d’estimer les gradients p et q puis d’estimer la distance radiale après intégration, il est envisageable de transformer le système (2.9) en un unique système linéaire en $\ln \rho$, afin de pouvoir trouver une solution au sens des moindres carrés. Ce système serait très grand mais il a l’avantage d’être épars. Avec cette approche, nous pourrions retrouver la géométrie de la surface directement à partir des intensités, sans passer par une quelconque étape intermédiaire (la phase d’intégration). Cette approche dite “globale” a déjà été considérée dans la littérature, par exemple, dans [Logothetis *et al.* 2016], ou encore [Smith & Fang 2016]. Nous supposons qu’une telle approche réduirait la propagation d’erreur et, de fait, la production d’artéfacts dans la reconstruction finale.

Équation cumulative de l’irradiance

Dans l’ensemble du manuscrit, nous avons considéré que les pixels correspondent à des points 3D sur la surface observée. Néanmoins, et d’autant plus avec la résolution spatiale hétérogène des capteurs à très grand champ de vue, ces pixels correspondent en réalité à des portions de la surface. En exploitant à nouveau la grille de répartition⁴, nous avons essayé de récrire l’équation de l’irradiance sous la forme d’une fonction d’énergie appliquée à une portion de la surface étudiée. Cette portion, appelée “patch”, serait délimitée par des bornes, $[\theta_{\min}, \theta_{\max}]$ et $[\varphi_{\min}, \varphi_{\max}]$, en conservant la grille utilisée dans les expérimentations. Nous estimons pouvoir utiliser cette reformulation pour tenir compte de l’hétérogénéité de la taille des “patches”, voire, travailler à des échelles différentes.

Séparation de l’illumination directe et indirecte

Lorsque la scène observée comporte des surfaces concaves, le phénomène d’inter-réflexions se produit et conduit à une illumination plus importante des points de la surface, suite au rebond des rayons lumineux (voir la section §1.2.3). En s’inspirant des travaux développés dans le OMILab⁵ découverts au cours d’un stage au Japon de deux mois, nous pourrions envisager de synchroniser la source de lumière avec la caméra, dans un environnement contrôlé, afin d’appliquer une des techniques permettant de distinguer l’éclairage direct et indirect (par exemple, la méthode dans [Kubo *et al.* 2018]). Cependant, en considérant la complexité de l’équipement utilisé dans [Kubo *et al.* 2018], l’acquisition des images risque d’être bien plus lente et laborieuse.

4. Voir le paragraphe sur la suppression du ré-échantillonnage spatial.

5. Optical Media Interface Lab (Mukaigawa Lab), Division of Information Science, Graduate School of Science and Technology, Nara Institute of Science and Technology, Japan

Approches reposant sur l'apprentissage

Dans les situations où l'utilisateur est en mesure d'acquérir de grands volumes d'images, les techniques d'estimation des normales basées sur l'apprentissage profond (réseaux de neurones) peuvent être considérées. En effet, de plus en plus d'approches ont fait leur apparition dans la littérature récente, afin de résoudre le problème de stéréophotométrie non calibré. Par exemple, dans [Logothetis *et al.* 2020] et [Chen *et al.* 2020], les réseaux de neurones permettent d'estimer les normales de la surface observée, puis, un algorithme d'intégration exploite cette estimation afin de retrouver la géométrie de la surface. Enfin, des approches, dites "end-to-end", permettant de reconstruire la surface 3D directement à partir des images, sans passer par la phase d'intégration, sont actuellement à l'étude.

Bibliographie

- [Abrams *et al.* 2012] A. Abrams, C. Hawley et R. Pless. *Heliometric Stereo : Shape from Sun Position*. In Proc. Eur. Conf. Comp. Vis., pages 357–370, 2012. (cité en pages 1 et 78)
- [Abrams *et al.* 2013] A. Abrams, K. Miskell et R. Pless. *The Epipolar Constraint : Monocular Shape from Shadow Correspondence*. In Proc. IEEE Conf. Comp. Vis. Pattern Recogn., pages 1407–1414, 2013. (cité en page 1)
- [Ackermann & Goesele 2015] J. Ackermann et M. Goesele. *A Survey of Photometric Stereo Techniques*. Found. Trends Comput. Graphics and Vision, vol. 9, no. 3–4, pages 149–254, 2015. (cité en pages 1, 9 et 19)
- [Ackermann *et al.* 2012] J. Ackermann, F. Langguth, S. Fuhrmann et M. Goesele. *Photometric Stereo for Outdoor Webcams*. In Proc. IEEE Conf. Comp. Vis. Pattern Recogn., pages 262–269, 2012. (cité en pages 1 et 78)
- [Adarve & Mahony 2017] J.D. Adarve et R. Mahony. *Spherepix : a Data Structure for Spherical Image Processing*. IEEE Robot. Autonom. Lett., vol. 2, no. 2, pages 483–490, 2017. (cité en page 77)
- [Alldrin *et al.* 2008] N.G. Alldrin, T. Zickler et D.J. Kriegman. *Photometric stereo with non-parametric and spatially-varying reflectance*. In Proc. IEEE Conf. Comp. Vis. Pattern Recogn., 2008. [web] <http://vision.ucsd.edu/~nalldrin/research/cvpr08/datasets/>. (cité en pages 20 et 42)
- [Barazzetti *et al.* 2017] L. Barazzetti, M. Previtali et F. Roncoroni. *3D modelling with the Samsung Gear 360*. In Proc. 3D Virtual Reconstruction and Visualization of Complex Architectures, volume XLII-2/W3, pages 85–90, 2017. (cité en pages 51, 52 et 57)
- [Barreto 2006] J. Barreto. *A unifying geometric representation for central projection systems*. Comput. Vis. Image Und., vol. 103, no. 3, pages 208–217, 2006. (cité en pages 43, 57 et 58)
- [Barsky & Petrou 2003] S. Barsky et M. Petrou. *The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows*. IEEE Trans. Pattern Anal., vol. 25, no. 10, pages 1239–1252, 2003. (cité en pages 14 et 30)
- [Basri *et al.* 2007] R. Basri, D. Jacobs et I. Kemelmacher. *Photometric Stereo with General, Unknown Lighting*. Int. J. Comput. Vision, vol. 72, no. 3, pages 239–257, 2007. (cité en page 1)

- [Benseddik *et al.* 2020] H.-E. Benseddik, F. Morbidi et G. Caron. *PanoraMIS : An ultra-wide field of view image dataset for vision-based robot-motion estimation*. The International Journal of Robotics Research, vol. 39, no. 9, pages 1037–1051, 2020. (cité en page 54)
- [Berenguel-Baeta *et al.* 2020] B. Berenguel-Baeta, J. Bermudez-Cameo et J.J. Guerrero. *Omniscv : An Omnidirectional Synthetic Image Generator for Computer Vision*. Sensors, vol. 20, no. 7, page 2066, 2020. (cité en page 78)
- [Blender Foundation 2018] Blender Foundation. *Blender 2.79b : 3D Computer Graphics Software*, 2018. www.blender.org. (cité en pages 35 et 61)
- [Brown 1971] D. Brown. *Close-range camera calibration*. Photogramm. Eng, vol. 37, no. 8, pages 855–866, 1971. (cité en page 57)
- [Caracotte *et al.* 2020] J. Caracotte, F. Morbidi et E.M. Mouaddib. *Photometric stereo with central panoramic cameras*. Comput. Vis. Image Und., vol. 201, page 103080, 2020. (cité en pages 4, 46, 58 et 67)
- [Caracotte *et al.* 2021] J. Caracotte, F. Morbidi et E.M. Mouaddib. *Photometric Stereo with Twin-Fisheye Cameras*. In Proc. IEEE Int. Conf. Pattern Recogn., pages 5270–5277, 2021. (cité en page 4)
- [Caron & Eynard 2011a] G. Caron et D. Eynard. *Hybrid stereoscopic calibration*, 2011. SPIE Newsroom, DOI : 10.1117/2.1201106.003738. (cité en page 64)
- [Caron & Eynard 2011b] G. Caron et D. Eynard. *Multiple Camera Types Simultaneous Stereo Calibration*. In Proc. IEEE Int. Conf. Robot. Automat., pages 2933–2938, 2011. (cité en page 43)
- [Caron & Morbidi 2018] G. Caron et F. Morbidi. *Spherical Visual Gyroscope for Autonomous Robots using the Mixture of Photometric Potentials*. In Proc. IEEE Int. Conf. Robot. Automat., pages 820–827, 2018. (cité en pages 53, 54, 57, 58 et 64)
- [Chabert *et al.* 2006] C.-F. Chabert, P. Einarsson, A. Jones, B. Lamond, W.-C. Ma, S. Sylvan, T. Hawkins et P. Debevec. *Relighting human locomotion with flowed reflectance fields*. In ACM SIGGRAPH Sketches, pages 76–es. 2006. [web] <https://vgl.ict.usc.edu/Data/LightStage/>. (cité en page 20)
- [Chen *et al.* 2020] G. Chen, K. Han, B. Shi, Y. Matsushita et K.-W.K. Wong. *Deep photometric stereo for non-Lambertian surfaces*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020. (cité en page 80)
- [Coxeter 1969] H.S.M. Coxeter. *Introduction to Geometry*. John Wiley & Sons, 2nd édition, 1969. (cité en page 25)
- [Delmas *et al.* 2021] S. Delmas, F. Morbidi, G. Caron, J. Albrand, M. Jeanne-Rose, L. Devigne et M. Babel. *SpheriCol : A Driving Assistance System for Power Wheelchairs Based on*

- Spherical Vision and Range Measurements*. In Proc. IEEE/SICE International Symposium on System Integration, pages 505–510, 2021. (cité en pages 55 et 56)
- [Diaz & Sturm 2013] M. Diaz et P. Sturm. *Estimating Photometric Properties from Image Collections*. J. Math. Imaging Vis., vol. 47, no. 1–2, pages 93–107, 2013. (cité en page 40)
- [Fan et al. 2016] H. Fan, Y. Luo, L. Qi, N. Wang, J. Dong et H. Yu. *Robust Photometric Stereo in a Scattering Medium via Low-Rank Matrix Completion and Recovery*. In Proc. 9th Int. Conf. Human Syst. Interactions, pages 323–329, 2016. (cité en page 72)
- [Forsyth & Ponce 2012] D.A. Forsyth et J. Ponce. *Computer vision : a modern approach*. Pearson, 2012. (cité en page 9)
- [Frankot & Chellappa 1988] R.T. Frankot et R. Chellappa. *A Method for Enforcing Integrability in Shape from Shading Algorithms*. IEEE Trans. Pattern Anal., vol. 10, no. 4, pages 439–451, 1988. (cité en page 30)
- [Furukawa & Hernández 2015] Y. Furukawa et C. Hernández. *Multi-View Stereo : A Tutorial*. Found. Trends Comput. Graphics and Vision, vol. 9, no. 1–2, pages 1–148, 2015. (cité en page 1)
- [Gao & Shen 2017] W. Gao et S. Shen. *Dual-fisheye omnidirectional stereo*. In Proc. IEEE/RSJ Int. Conf. Intel. Robots Syst., pages 6715–6722, 2017. (cité en pages 51, 56 et 57)
- [Geyer & Daniilidis 2000] C. Geyer et K. Daniilidis. *A Unifying Theory for Central Panoramic Systems and Practical Implications*. In Proc. Eur. Conf. Comp. Vis., pages 445–461, 2000. (cité en pages 24 et 57)
- [Geyer & Daniilidis 2001] C. Geyer et K. Daniilidis. *Catadioptric Projective Geometry*. Int. J. Comput. Vision, vol. 45, no. 3, pages 223–243, 2001. (cité en page 24)
- [Goldman et al. 2010] D.B. Goldman, B. Curless, A. Hertzmann et S.M. Seitz. *Shape and Spatially-Varying BRDFs from Photometric Stereo*. IEEE Trans. Pattern Anal., vol. 32, no. 6, pages 1060–1071, 2010. (cité en page 42)
- [Gonçalves & Araújo 2009] N. Gonçalves et H. Araújo. *Estimating parameters of noncentral catadioptric systems using bundle adjustment*. Comput. Vis. Image Und., vol. 113, no. 1, pages 11–28, 2009. (cité en page 77)
- [Grossberg & Nayar 2004] M.D. Grossberg et S.K. Nayar. *Modeling the space of camera response functions*. IEEE Trans. Pattern Anal., vol. 26, no. 10, pages 1272–1282, 2004. (cité en page 39)
- [Guan & Smith 2017] H. Guan et W.P. Smith. *BRISKS : Binary Features for Spherical Images on a Geodesic Grid*. In Proc. IEEE Conf. Comp. Vis. Pattern Recogn., pages 4516–4524, 2017. (cité en page 77)

- [Hajjami *et al.* 2020] J. Hajjami, J. Caracotte, G. Caron et T. Napoleon. *ArUcOmni : detection of highly reliable fiducial markers in panoramic images*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pages 634–635, 2020. (cité en page 4)
- [Happel & Brenner 1983] J. Happel et H. Brenner. *Low Reynolds number hydrodynamics : with special applications to particulate media*. Springer, 1983. (cité en page 77)
- [Harker & O’Leary 2008] M. Harker et P. O’Leary. *Least Squares Surface Reconstruction from Measured Gradient Fields*. In Proc. IEEE Conf. Comp. Vis. Pattern Recogn., pages 1–7, 2008. (cité en pages 19 et 38)
- [Harker & O’Leary 2015] M. Harker et P. O’Leary. *Regularized Reconstruction of a Surface from its Measured Gradient Field : Algorithms for Spectral, Tikhonov, Constrained, and Weighted Regularization*. *J. Math. Imaging Vis.*, vol. 51, no. 1, pages 46–70, 2015. (cité en page 19)
- [Hawary *et al.* 2020] F. Hawary, T. Maugey et C. Guillemot. *Sphere Mapping for Feature Extraction From 360° Fish-Eye Captures*. In Proc. IEEE 22nd Int. Workshop Multimedia Signal Proc., 2020. (cité en page 58)
- [Hayakawa 1994] H. Hayakawa. *Photometric stereo under a light source with arbitrary motion*. *J. Opt. Soc. Am. A*, vol. 11, no. 11, pages 3079–3089, 1994. (cité en page 1)
- [Hernández *et al.* 2008] C. Hernández, G. Vogiatzis et R. Cipolla. *Multiview Photometric Stereo*. *IEEE Trans. Pattern Anal.*, vol. 30, no. 3, pages 548–554, 2008. (cité en page 2)
- [Hernández *et al.* 2010] C. Hernández, G. Vogiatzis et R. Cipolla. *Overcoming Shadows in 3-Source Photometric Stereo*. *IEEE Trans. Pattern Anal.*, vol. 33, no. 2, pages 419–426, 2010. (cité en pages 14 et 30)
- [Higo *et al.* 2010] T. Higo, Y. Matsushita et K. Ikeuchi. *Consensus photometric stereo*. In Proc. IEEE Conf. Comp. Vis. Pattern Recogn., pages 1157–1164, 2010. (cité en page 42)
- [Horn & Brooks 1986] B.K.P. Horn et M.J. Brooks. *The variational approach to shape from shading*. *Computer Vision, Graphics, and Image Processing*, vol. 33, no. 2, pages 174–208, 1986. (cité en page 19)
- [Horn 1984] B.K. Horn. *Extended Gaussian Images*. *Proc. of the IEEE*, vol. 72, no. 12, pages 1671–1686, 1984. (cité en page 77)
- [Ikehata & Aizawa 2014] S. Ikehata et K. Aizawa. *Photometric Stereo using Constrained Bivariate Regression for General Isotropic Surfaces*. In Proc. IEEE Conf. Comp. Vis. Pattern Recogn., pages 2187–2194, 2014. (cité en page 42)

- [Ikehata *et al.* 2012] S. Ikehata, D. Wipf, Y. Matsushita et K. Aizawa. *Robust photometric stereo using sparse regression*. In Proc. IEEE Conf. Comp. Vis. Pattern Recogn., pages 318–325, 2012. (cité en page 42)
- [Jamaluddin *et al.* 2017] A.Z. Jamaluddin, C. Jiang, O. Morel, R. Seulin et D. Fofi. *3D Reconstruction from Specialized Wide Field of View Camera System Using Unified Spherical Model*. In Proc. Int. Conf. Image Anal. Process., pages 495–506, 2017. (cité en pages 50, 56, 57 et 58)
- [Jung *et al.* 2015] J. Jung, J.-Y. Lee et I. Kweon. *One-day outdoor photometric stereo via skylight estimation*. In Proc. IEEE Conf. Comp. Vis. Pattern Recogn., pages 4521–4529, 2015. (cité en page 1)
- [Jung *et al.* 2017] J. Jung, B. Kim, J.Y. Lee, B. Kim et S. Lee. *Robust upright adjustment of 360 spherical panoramas*. Vis. Comput., vol. 33, no. 6, pages 737–747, 2017. (cité en page 52)
- [Kannala & Brandt 2006] J. Kannala et S.S. Brandt. *A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses*. IEEE Trans. Pattern Anal., vol. 28, no. 8, pages 1335–1340, 2006. (cité en page 57)
- [Khomutenko *et al.* 2016] B. Khomutenko, G. Garcia et P. Martinet. *An Enhanced Unified Camera Model*. IEEE Robot. Autonom. Lett., vol. 1, no. 1, pages 137–144, 2016. (cité en pages 57 et 77)
- [Kondo *et al.* 2008] K. Kondo, Y. Mukaigawa et Y. Yagi. *Free-Form Mirror Design Inspired by Photometric Stereo*. In Proc. 8th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras, 2008. (cité en page 2)
- [Kubo *et al.* 2018] H. Kubo, S. Jayasuriya, T. Iwaguchi, T. Funatomi, Y. Mukaigawa et S.G. Narasimhan. *Acquiring and characterizing plane-to-ray indirect light transport*. In 2018 IEEE International Conference on Computational Photography (ICCP), pages 1–10, 2018. (cité en pages 15 et 79)
- [Li *et al.* 2020] M. Li, Z. Zhou, Z. Wu, B. Shi, C. Diao et P. Tan. *Multi-view photometric stereo : a robust solution and benchmark dataset for spatially varying isotropic materials*. IEEE Transactions on Image Processing, vol. 29, pages 4159–4173, 2020. [web] <https://sites.google.com/site/photometricstereodata/mv>. (cité en page 21)
- [Liao *et al.* 2007] M. Liao, L. Wang et M. Gong. *Light Fall-off Stereo*. In Proc. IEEE Conf. Comp. Vis. Pattern Recogn., 2007. (cité en pages 71 et 74)
- [Logothetis *et al.* 2016] F. Logothetis, R. Mecca, Y. Quéau et R. Cipolla. *Near-Field Photometric Stereo in Ambient Light*. In Proc. 27th British Machine Vision Conf., pages 1–12, 2016. (cité en page 79)

- [Logothetis *et al.* 2017] F. Logothetis, R. Mecca et R. Cipolla. *Semi-calibrated near field photometric stereo*. In Proc. IEEE Conf. Comp. Vis. Pattern Recogn., pages 941–950, 2017. (cité en pages 1, 69 et 77)
- [Logothetis *et al.* 2019] F. Logothetis, R. Mecca et R. Cipolla. *A Differential Volumetric Approach to Multi-View Photometric Stereo*. In Proc. IEEE Int. Conf. Comp. Vis., pages 1052–1061, 2019. (cité en page 2)
- [Logothetis *et al.* 2020] F. Logothetis, I. Budvytis, R. Mecca et R. Cipolla. *A CNN Based Approach for the Near-Field Photometric Stereo Problem*. In Proc. 31st British Machine Vision Conf., 2020. (cité en page 80)
- [Ma *et al.* 2004] Y. Ma, S. Soatto, J. Košecská et S. Shankar Sastry. *An Invitation to 3D Computer Vision : From Images to Geometric Models*. Springer, 2004. (cité en page 8)
- [Mecca *et al.* 2014] R. Mecca, A. Wetzler, A.M. Brustein et R. Kimmel. *Near Field Photometric Stereo with Point Light Sources*. SIAM J. Imag. Sci., vol. 7, no. 4, pages 2732–2770, 2014. (cité en pages 1, 69 et 77)
- [Mouaddib *et al.* 2019] E. Mouaddib, G. Caron, D. Lecllet-Groux et F. Morbidi. *Le patrimoine in silico. Exemple de la cathédrale d’Amiens*. In Situ. Revue des patrimoines, no. 39, 2019. Numéro spécial sur “Imagerie numérique et patrimoine culturel : enjeux scientifiques et opérationnels”. (cité en page 74)
- [Mouaddib *et al.* 2025] E. Mouaddib, G. Caron, D. Lecllet-Groux et F. Morbidi. *E-Cathedral : In Silico Monumental Heritage, 2015-2025*. [web] <https://mis.u-picardie.fr/e-cathedrale>. (cité en page 3)
- [Nicodemus *et al.* 1992] F.E. Nicodemus, J.C. Richmond, J.J. Hsia, I.W. Ginsberg et T. Limperis. *Geometrical considerations and nomenclature for reflectance*. NBS monograph, vol. 160, page 4, 1992. (cité en page 9)
- [Papadhimitri & Favaro 2013] T. Papadhimitri et P. Favaro. *A New Perspective on Uncalibrated Photometric Stereo*. In Proc. IEEE Conf. Comp. Vis. Pattern Recogn., pages 1474–1481, 2013. (cité en page 1)
- [Pathak *et al.* 2017] S. Pathak, A. Moro, H. Fujii, A. Yamashita et H. Asama. *Virtual reality with motion parallax by dense optical flow-based depth generation from two spherical images*. In Proc. IEEE/SICE International Symposium on System Integration), pages 887–892, 2017. (cité en page 53)
- [Prados & Faugeras 2005] E. Prados et O. Faugeras. *Shape From Shading : a well-posed problem?* In Proc. IEEE Conf. Comp. Vis. Pattern Recogn., volume 2, pages 870–877, 2005. (cité en page 71)

- [Prados & Faugeras 2006] E. Prados et O.D. Faugeras. *Shape From Shading*. In N. Paragios, Y. Chen et O.D. Faugeras, éditeurs, *Handbook of Mathematical Models in Computer Vision*, chapitre 23, pages 375–388. Springer, 2006. (cité en page 1)
- [Quéau *et al.* 2015] Y. Quéau, F. Lauze et J.-D. Durou. *Solving Uncalibrated Photometric Stereo Using Total Variation*. *J. Math. Imaging Vis.*, vol. 52, no. 1, pages 87–107, 2015. (cité en page 1)
- [Quéau *et al.* 2017] Y. Quéau, T. Wu, F. Lauze, J.-D. Durou et D. Cremers. *A Non-Convex Variational Approach to Photometric Stereo under Inaccurate Lighting*. In *Proc. IEEE Conf. Comp. Vis. Pattern Recogn.*, pages 350–359, 2017. (cité en pages 1 et 77)
- [Quéau *et al.* 2018a] Y. Quéau, J.-D. Durou et J.-F. Aujol. *Normal Integration : A Survey*. *J. Math. Imaging Vis.*, vol. 60, no. 4, pages 576–593, 2018. (cité en page 19)
- [Quéau *et al.* 2018b] Y. Quéau, J.-D. Durou et J.-F. Aujol. *Variational Methods for Normal Integration*. *J. Math. Imaging Vis.*, vol. 60, no. 4, pages 609–632, 2018. (cité en page 19)
- [Quéau 2015] Y. Quéau. *Reconstruction tridimensionnelle par stéréophotométrie*. PhD thesis, Institut National Polytechnique de Toulouse, 2015. (cité en page 9)
- [Schnieders & Wong 2013] D. Schnieders et K.-Y. Wong. *Camera and light calibration from reflections on a sphere*. *Comput. Vis. Image Und.*, vol. 117, no. 10, pages 1536–1547, 2013. (cité en pages 3 et 40)
- [Schönbein *et al.* 2014] M. Schönbein, T. Strauß et A. Geiger. *Calibrating and Centering Quasi-Central Catadioptric Cameras*. In *Proc. IEEE Int. Conf. Robot. Automat.*, pages 4443–4450, 2014. (cité en page 77)
- [Shi *et al.* 2012] B. Shi, P. Tan, Y. Matsushita et K. Ikeuchi. *Elevation Angle from Reflectance Monotonicity : Photometric Stereo for General Isotropic Reflectances*. In *Proc. Eur. Conf. Comp. Vis.*, pages 455–468, 2012. (cité en page 42)
- [Shi *et al.* 2014] B. Shi, P. Tan, Y. Matsushita et K. Ikeuchi. *Bi-Polynomial Modeling of Low-Frequency Reflectances*. *IEEE Trans. Pattern Anal.*, vol. 36, no. 6, pages 1078–1091, 2014. (cité en page 42)
- [Shi *et al.* 2016] B. Shi, Z. Wu, Z. Mo, D. Duan, S.-K. Yeung et P. Tan. *A Benchmark Dataset and Evaluation for Non-Lambertian and Uncalibrated Photometric Stereo*. In *Proc. IEEE Conf. Comp. Vis. Pattern Recogn.*, pages 3707–3716, 2016. (cité en page 20)
- [Shi *et al.* 2019] B. Shi, Z. Mo, Z. Wu, D. Duan, S.-K. Yeung et P. Tan. *A Benchmark Dataset and Evaluation for Non-Lambertian and Uncalibrated Photometric Stereo*. *IEEE Trans. Pattern Anal.*, vol. 41, no. 2, pages 271–284, 2019. [web] <https://sites.google.com/site/photometricstereodata/single>. (cité en pages 20, 41 et 42)

- [Smith & Fang 2016] W. Smith et F. Fang. *Height from photometric ratio with model-based light source selection*. *Comput. Vis. Image Und.*, vol. 145, pages 128–138, 2016. (cité en pages 21, 30 et 79)
- [Tankus & Kiryati 2005] A. Tankus et N. Kiryati. *Photometric Stereo under Perspective Projection*. In *Proc. IEEE Int. Conf. Comp. Vis.*, volume 1, pages 611–616, 2005. (cité en pages 2, 11, 12 et 13)
- [Tankus *et al.* 2005] A. Tankus, N. Sochen et Y. Yeshurun. *Shape-from-Shading Under Perspective Projection*. *Int. J. Comput. Vision*, vol. 63, no. 1, pages 21–43, 2005. (cité en pages 13 et 27)
- [Usenko *et al.* 2018] V. Usenko, N. Demmel et D. Cremers. *The Double Sphere Camera Model*. In *Proc. Int. Conf. 3D Vision*, pages 552–560, 2018. (cité en pages 57 et 77)
- [Vlasic *et al.* 2009] D. Vlasic, P. Peers, I. Baran, P. Debevec, J. Popović, S. Rusinkiewicz et W. Matusik. *Dynamic Shape Capture using Multi-View Photometric Stereo*. *ACM Trans. Graphic*, vol. 28, no. 5, 2009. Article n. 174. (cité en page 2)
- [Wang *et al.* 2020] N.-H. Wang, B. Solarte, Y.-H. Tsai, W.-C. Chiu et M. Sun. *360SD-Net : 360° Stereo Depth Estimation with Learnable Cost Volume*. In *Proc. IEEE Int. Conf. Robot. Automat.*, pages 582–588, 2020. (cité en page 55)
- [Weber & Arfken 2003] H.J. Weber et G.B. Arfken. *Essential Mathematical Methods for Physicists*. Academic Press, 6th édition, 2003. (cité en pages 25 et 26)
- [Woodham 1980] R.J. Woodham. *Photometric method for determining surface orientation from multiple images*. *Opt. Eng.*, vol. 19, no. 1, pages 139–144, 1980. (cité en pages 1, 10 et 42)
- [Wu *et al.* 2010] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang et Y. Ma. *Robust Photometric Stereo via Low-Rank Matrix Completion and Recovery*. In *Proc. Asian Conf. Comp. Vis.*, pages 703–717, 2010. (cité en page 42)
- [Ying & Hu 2004] X. Ying et Z. Hu. *Can We Consider Central Catadioptric Cameras and Fisheye Cameras within a Unified Imaging Model?* In *Proc. Eur. Conf. Comp. Vis.*, pages 442–455, 2004. (cité en page 57)
- [Zhang & Huang 2021] Y. Zhang et F. Huang. *Panoramic Visual SLAM Technology for Spherical Images*. *Sensors*, vol. 21, no. 3, page 705, 2021. (cité en page 58)
- [Zhao *et al.* 2015] Q. Zhao, W. Feng, L. Wan et J. Zhang. *SPHORB : A Fast and Robust Binary Feature on the Sphere*. *Int. J. Comput. Vis.*, vol. 113, no. 2, pages 143–159, 2015. (cité en page 77)

Résumé : Cette thèse s'intéresse à la reconstruction 3D par stéréophotométrie et à la vision omnidirectionnelle. La stéréophotométrie repose sur l'exploitation de plusieurs photographies d'une scène, capturées sous différents éclairages par un même appareil immobile. La vision omnidirectionnelle rassemble les capteurs ainsi que les assemblages de caméras qui permettent de faire l'acquisition en une image d'une très grande portion de l'environnement autour de l'appareil. En s'appuyant sur quatre décennies de travaux en stéréophotométrie et en utilisant le modèle unifié pour la projection centrale, nous tentons de réunir ces deux domaines de recherche. Dans un premier temps, nous nous intéressons aux techniques pour l'estimation des normales à la surface ainsi qu'à l'intégration des gradients de profondeur, dans le but de retrouver la géométrie de la scène observée. Nous poursuivons en introduisant une nouvelle équation de l'irradiance ainsi qu'une chaîne de traitements modulaire, toutes deux adaptées aux capteurs à projection centrale. Le fonctionnement de la méthode est vérifié en utilisant une caméra perspective et un capteur catadioptrique. Nous détaillons ensuite une extension afin de réaliser des reconstructions 3D à partir d'images issues de caméras twin-fisheye. Enfin, nous nous intéressons aux limites des approches proposées ainsi qu'aux pistes d'amélioration envisagées.

Mots clés : reconstruction 3D, stéréophotométrie, vision omnidirectionnelle, estimation de normales

Abstract : This thesis focuses on the photometric stereo problem and omnidirectional vision. The photometric stereo problem is a 3D-reconstruction technique that requires multiple images of a surface under different lighting conditions from a single point of view. Omnidirectional vision encompasses the devices and camera platforms that capture much of the environment around them in a single image. Following four decades of research in the photometric stereo literature and using the unified model for central projection cameras, we try to merge these research fields. We first focus on techniques for estimating surface normals, and integrating depth gradients to retrieve the shape. Next, we introduce a new spherical irradiance equation which we use to solve the photometric stereo problem using two central projection cameras. The approach is validated using synthetic and real images from a perspective camera and a catadioptric imaging device. We later extend the approach to perform 3d-reconstruction by photometric stereo using twin-fisheye cameras. Finally, we study some limitations of the approaches and we discuss the ways to overcome these limits.

Keywords : 3D reconstruction, photometric stereo, omnidirectional vision, normal estimation
