Fusion d'approches photométriques et géométriques pour la création de modèles 3D

Jean Mélou

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Institut National Polytechnique de Toulouse (INP Toulouse)*

Présentée et soutenue le *12 mars 2020* **par :**

Jean MÉLOU

Fusion d'approches photométriques et géométriques pour la création de modèles 3D

JURY

Alfred BRUCKSTEIN	Technion, Haïfa	Rapporteur
Fabien CASTAN	Mikros Image, Paris	Invité
Vincent CHARVILLAT	IRIT, Toulouse	Président
Jean-Denis DUROU	IRIT, Toulouse	Membre
Maurizio FALCONE	La Sapienza, Rome	Rapporteur
Noura FARAJ	LIRMM, Montpellier	Membre
Pierre GURDJOS	IRIT, Toulouse	Membre
François LAUZE	DIKU, Copenhague	Membre
Benoît MAUJEAN	Mikros Image, Paris	Invité
Yvain QUÉAU	GREYC, Caen	Membre

École doctorale et spécialité :
MITT : Image, Information, Hypermédia
Unité de Recherche :
IRIT, UMR CNRS 5505
Directeur de Thèse :
Jean-Denis DUROU
Rapporteurs :
Alfred BRUCKSTEIN et *Maurizio* FALCONE

Remerciements

Je tiens tout d'abord à remercier les membres du jury d'évaluation. En premier lieu Alfred Bruckstein et Maurizio Falcone pour avoir accepté de rédiger un rapport sur le manuscrit, pour leur bienveillance ainsi que pour leurs critiques pertinentes. Un grand merci à Noura Faraj, pour avoir apporté un autre point de vue sur certains aspects de la thèse, ouvrant ainsi de nouvelles pistes de recherche. Un très grand merci à François Lauze dont l'ombre a plané sur ces années de doctorat, de son accueil à Kolding jusqu'à sa présence au sein du jury. Ses remarques se sont toujours révélées précieuses. Je tiens également à remercier Pierre Gurdjos pour avoir accepté de participer au jury. C'est notamment sous sa direction que j'ai fait mes premières armes au laboratoire. Mon disquaire tient également à le remercier pour mes nombreux achats suites à ses recommandations avisées. Enfin, merci à Vincent Charvillat, à la fois pour avoir accepté de présider le jury de thèse, et pour l'énergie certaine qu'il met dans la gestion de notre équipe de recherche.

Je souhaite ensuite remercier les personnes que j'ai pu côtoyer au sein de Mikros Image. Un immense merci à Benoit Maujean pour son accueil au sein de l'équipe R&D. L'enthousiasme dont il fait preuve concernant les projets de son équipe reste des plus admirables. Un très grand merci à mes collègues Mikrosiens et Mikrosiennes. Je pense tout d'abord à ceux avec qui j'ai pu travailler sur le projet AliceVision, mais également aux membres des autres équipes de Dev/TD. J'ai beaucoup appris auprès d'eux, au cours de ces années de doctorat.

Je remercie chaleureusement les membres de l'équipe RÉVA qui entretiennent un environnement des plus favorables au bon déroulement d'une thèse. Merci donc aux permanents, doctorants, ingénieurs et stagiaires pour la bonne humeur apportée au laboratoire, leur soutien, et leurs précieux conseils.

Merci aux stagiaires que j'ai pu encadrer : ils m'ont permis de mettre au clair bien des points sur lesquels je ne me serais peut-être jamais penché. Je tiens à remercier également les chercheurs avec qui j'ai pu échanger au cours de ce doctorat. Merci / vielen Dank / grazie mille aux membres des équipes au sein desquelles j'ai pu faire des séjours brefs mais toujours enrichissants.

Un très grand merci à Sylvie, Muriel et Annabelle, qui nous sauvent plus souvent qu'à leur tour, et ce avec une efficacité redoutable. Puissent-elles toujours nous pardonner nos écarts administratifs. Merci également aux enseignants que j'ai croisés sur ma route, dévoués et passionnés, qui savent transmettre à leurs élèves l'envie d'apprendre et d'être curieux.

Un très grand merci à mes amis qui ne se doutent pas du rôle important qu'ils ont joué au cours des dernières années. La diaspora bretonne, Huscule, les escales Hoëdicaises, et les wagons n°4 se sont révélés être des sas de décompression efficaces. Je tiens, en particulier, à remercier mes colocataires sans qui les séjours parisiens auraient eu une toute autre saveur. Bien entendu, certains collègues évoqués précédemment auraient tout à fait leur place ici.

Je tiens à exprimer mon immense gratitude envers ma famille, qui , sans surprise, fut un soutien inflexible. Un très grand merci à mes parents pour avoir été présents, attentifs

et aimants, et pour m'avoir permis de rester encore étudiant un peu plus longtemps. Merci à ma soeur pour ses petites attentions fraternelles et ma collection de chaussettes. Merci à mes grand-mères pour avoir bichonné leur petit fils. Merci à Lucien pour les après-midi à Castanet avec Denise, pour qui je garde une pensée particulière : ils ont été très présents et attentionnés tout au long de mes études toulousaines.

Un immense merci à Élodie pour sa patience et son soutien. La compréhension et l'attention dont elle a fait preuve, en particulier lors de ces derniers mois de doctorat, m'ont été d'une aide précieuse.

Enfin, je ne saurai jamais assez remercier les personnes ayant encadré cette thèse. J'ai eu la chance d'être entouré de personnes passionnées, d'une grande richesse professionnelle et humaine. Je leur suis extrêmement reconnaissant, et ne cacherai pas une certaine fierté d'avoir été leur thésard. Un immense merci à Fabien Castan, dont la curiosité scientifique n'a d'égal que son enthousiasme envers les projets qu'il porte. Il a été, tout au long de ce doctorat, un interlocuteur de qualité. Par sa situation d'industriel, il a apporté un point de vue radicalement différent, mais toujours pertinent. Un immense merci à Yvain Quéau. Il est peu dire qu'il a été un atout solide. Il m'a aidé bien plus souvent qu'à son tour, et s'est toujours montré d'une grande clairvoyance, et d'une grande disponibilité pour parler recherche et bien plus encore. Sans lui, mon expérience de doctorant n'aurait sûrement pas été la même. Un immense merci à Jean-Denis Durou, pour la manière dont il a encadré ces années de doctorat. Sa conception de la recherche, de par son caractère profondément humain, m'a permis de vivre plus sereinement ces années de doctorat. Il m'a, entre autre, convaincu de l'importance d'une certaine exigence pédagogique. En ce sens, je le remercie sincèrement pour sa patience face à mes nombreuses questions et pour les longues heures passées à la relecture, dans un souci de cohérence et de clarté du discours. Il m'a aussi convaincu de l'importance de bien choisir son réseau : la recherche avance également autour d'une table, entre amis, ou en famille. J'espère avoir assimilé une partie de ses nombreux et précieux conseils (scientifiques ou non) distillés au fil de nos discussions, et de nos pauses café.

Table des matières

Introduction

La captation du réel sous forme de représentation numérique en trois dimensions constitue une aide précieuse dans de nombreux domaines. Une grande variété de systèmes ont été mis au point pour atteindre cet objectif : IRM de diffusion, sonars, palpeurs mécaniques, télémètres laser, etc. Les progrès technologiques récents des appareils photographiques numériques ont fait d'eux une alternative, à la fois crédible et accessible au grand public, à ces systèmes d'acquisition généralement coûteux. La **reconstruction 3D photographique**, mieux connue sous le nom de **photogrammétrie**, regroupe un ensemble de techniques dont le but est de retrouver la dimension perdue lors de la prise de vue. Ces techniques intéressent de nombreux terrains d'application, comme par exemple la **post-production cinématographique**, domaine dans lequel s'inscrit cette thèse.

Modélisation 3D

Un **modèle 3D** est constitué d'un ensemble d'informations **géométriques** (position, orientation, etc.) et **photométriques** (couleur, texture, etc.) caractérisant une scène. À partir de ces informations, la **synthèse d'images** permet de créer une image, en simulant à la fois l'éclairage et l'appareil photographique. Un **moteur de rendu** [10, 133] calcule la trajectoire des rayons lumineux, depuis les sources lumineuses jusqu'à la surface photosensible de l'appareil photographique, après réflexions (éventuellement multiples) sur la surface des objets constituant la scène. Inversement, le **scannage 3D** vise à retrouver les caractéristiques géométriques et photométriques d'une scène à partir de photographies.

L'obtention d'un modèle 3D se révèle utile, voire nécessaire, à de nombreux domaines d'application [117] comme l'archéologie [87], l'aide au diagnostic médical [158], l'architecture [39] ou encore l'inspection visuelle de surfaces [147]. Quant à la post-production cinématographique, elle constitue un secteur d'activité en pleine expansion, en particulier pour les effets spéciaux, vu que les films et les publicités utilisent toujours plus d'effets visuels, et que les spectateurs deviennent de plus en plus exigeants sur la qualité de ces effets. La représentation d'une scène obtenue par scannage 3D se prête bien à la réalisation de « trucages », comme par exemple l'ajout d'éléments à l'existant (personnages ou décors), le rééclairage réaliste ou la pré-visualisation des effets spéciaux lors d'un tournage.

Reconstruction 3D

Le scannage 3D se distingue de la **reconstruction 3D**, qui cherche à retrouver les seules informations géométriques de la scène. Les techniques de reconstruction 3D à partir de photographies sont regroupées sous le terme générique de *shape-from*-X, « X » signifiant que différents indices peuvent être utilisés pour la reconstruction 3D (ombre, contour, texture, etc.). Les principales techniques de *shape-from*-X sont regroupées dans la table 1. En réalité, c'est généralement une combinaison de ces techniques qui est utilisée, ce qui permet de tirer parti des avantages de chacune.

	Techniques géométriques	Techniques photométriques
Techniques mono-vue ($n = 1$ image)	Projection de lumière structurée [53] *Shape-from-shadows* [152] *Shape-from-contour* [25] *Shape-from-texture* [171] *Shape-from-template* [14]	*Shape-from-shading* [74]
Techniques multi-vues ($n \geqslant 2$ images)	*Structure-from-motion* [118] Stéréoscopie [47] *Shape-from-silhouettes* [69] *Shape-from-focus* [123]	Stéréophotométrie [141]

TABLE 1 – Principales techniques de *shape-from*-X.

Les techniques de reconstruction 3D photométriques sont généralement sous-estimées, car elles requièrent des conditions opératoires contraignantes. Elles sont souvent perçues comme de purs exercices de style, alors que les techniques géométriques sont très en vogue. Ces dernières ont d'ailleurs fait le succès de solutions commerciales « clé en main » comme la Kinect de Microsoft, qui utilise la projection de lumière structurée, ou de logiciels libres comme AliceVision [7], qui utilise le *structure-from-motion*.

Le pipeline de reconstruction 3D AliceVision

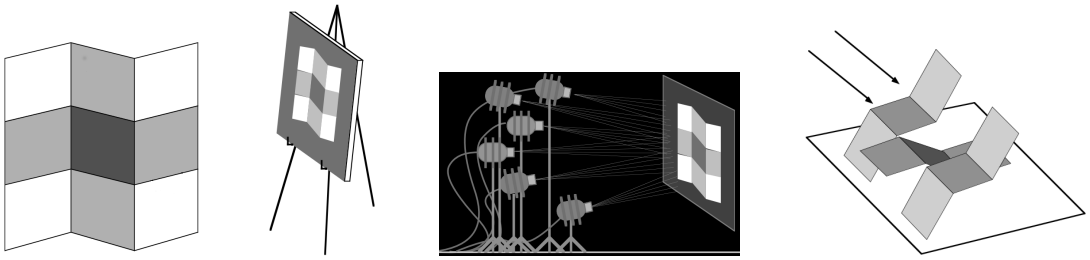
La société Mikros Image, acteur reconnu de l'industrie de la post-production, développe depuis 2010 un pipeline de reconstruction 3D qui permet de créer un maillage 3D (information géométrique) à partir d'un ensemble non ordonné de photographies. Le logiciel libre Meshroom, qui est l'interface de la suite logicielle AliceVision [7], est le fruit d'une longue collaboration entre Mikros Image et le laboratoire IRIT de Toulouse, qui s'est matérialisée par la participation aux projets européens POPART (2015-2016) et LADIO (2017-2018), en partenariat avec d'autres laboratoires européens tels que Simula (Oslo) et CTU (Prague), et par la présente thèse, qui s'est déroulée dans le cadre d'une convention CIFRE.

Cette solution de **photogrammétrie** utilise la disparité entre images acquises sous différents angles, à l'instar de la vision binoculaire humaine. Le pipeline classique de photogrammétrie comporte deux étapes. Tout d'abord, le ***structure-from-motion*** permet d'estimer les différentes poses de l'appareil photographique, ainsi qu'une représentation parcimonieuse de la scène, qui prend la forme d'un nuage de points 3D très peu dense. Est appliqué ensuite un algorithme de **stéréoscopie multi-vues**, plus connu sous sa dénomination anglaise ***multi-view stereo*** (MVS). À partir des poses de la caméra préalablement estimées, le MVS fournit une reconstruction 3D dense par pose. Néanmoins, comme aucune de ces deux étapes n'utilise explicitement les caractéristiques photométriques de la scène, la qualité des résultats dépend de la présence ou non de texture sur la surface des objets à reconstruire. Lorsque la scène est peu texturée, ce pipeline ne parvient pas à fournir des résultats satisfaisants. Or, comme nous l'avons déjà dit, il est souvent opportun de combiner différentes techniques de reconstruction 3D, afin de tirer parti des avantages de chacune. Les lacunes de ce pipeline pourraient donc être comblées en combinant une technique photométrique aux techniques géométriques utilisées par AliceVision.

Reconstruction 3D par l'ombrage

Les techniques photométriques reposent sur l'analyse de la quantité de lumière reçue par chaque site du capteur photosensible de l'appareil photographique. Si celui-ci est étalonné, alors le niveau de gris (ou le niveau de couleur) est proportionnel à la luminance émise par la scène, qui est caractéristique, en partie du moins, du relief.

Dans ce mémoire, nous nous intéressons particulièrement à la technique du ***shape-from-shading*** (SfS), qui désigne la reconstruction 3D à partir de l'ombrage. La luminance émise par une surface peut être modélisée par une fonction faisant intervenir l'éclairage, la réflectance de la surface et sa géométrie. Le SfS, qui vise à inverser cette fonction pour reconstruire le relief de la scène, est un problème fondamentalement mal posé en l'absence de connaissance sur l'éclairage et sur la réflectance, comme cela est illustré par la « métaphore de l'atelier » d'Adelson et Pentland [1] (cf. figure 1).



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

FIGURE 1 – La « métaphore de l'atelier » (extraite d'un article d'Adelson et Pentland [1]). L'image (a) peut être interprétée : (b) en supposant que les variations de niveau de gris sont uniquement dues à la réflectance (principe du « trompe-l'œil ») ; (c) en éclairant une surface blanche plane par un jeu d'éclairages complexe ; (d) en modelant une forme 3D uniformément blanche éclairée par un faisceau parallèle et uniforme. Le *shape-from-shading* vise à estimer le relief à partir d'une image telle que (a), en supposant que l'éclairage incident et la réflectance de la surface sont connus.

Si l'on demande à un peintre, à un « régisseur lumière » et à un sculpteur de réaliser une œuvre expliquant une image, chacun d'eux apportera une réponse différente, mais toutes seront plausibles. Le peintre supposera la surface plane et l'éclairage uniforme, les changements de clarté étant expliqués par des changements de couleur. Le régisseur lumière proposera une configuration d'éclairages complexe, placée face à une surface plane de réflectance uniforme. Enfin, le sculpteur expliquera les changements de clarté uniquement par l'ombrage, qui résulte des variations locales d'orientation de la surface, l'éclairage et la réflectance étant supposés uniformes. Cette métaphore illustre le caractère fondamentalement mal posé du SfS : il existe une infinité de triplets réflectance/éclairage/relief permettant d'expliquer une image. Cela explique pourquoi le SfS suppose généralement que la réflectance et l'éclairage sont connus, et à quel point la fiabilité de ces connaissances est déterminante pour reconstruire le relief avec précision.

L'utilisation du SfS étant limitée par des conditions opératoires très contrôlées, cela explique en partie sa mauvaise réputation. En outre, même en connaissant parfaitement l'éclairage et la réflectance, le problème à résoudre reste mal posé, car il subsiste des ambiguïtés dites de « basses fréquences », comme l'ambiguïté concave/convexe. Malgré tout, nous montrons dans ce mémoire comment les informations apportées par le SfS peuvent aider à surmonter les lacunes d'un pipeline de photogrammétrie, dans la mesure où les méthodes géométriques peuvent « guider » le SfS en lui apportant des informations sur la géométrie de la scène. En outre, si le SfS utilise une seule pose de la caméra, son extension aux techniques multi-vues se révèle prometteuse.

Contributions et organisation du mémoire

Cette thèse porte sur l'amélioration d'un pipeline de photogrammétrie dans le contexte de la post-production, par combinaison de techniques photométriques et géométriques. À terme, l'objectif de ce travail est de transformer un système de reconstruction 3D existant en un outil plus complet permettant simultanément d'affiner le relief et d'estimer les caractéristiques photométriques de la scène que sont la réflectance de la surface et l'éclairage incident. Les différentes contributions de cette thèse sont organisées comme suit.

Le pipeline classique de photogrammétrie est présenté dans le chapitre 1. Nous y esquissons les principes des techniques de *structure-from-motion* et de stéréoscopie multi-vues. Nous illustrons également les limites de ce pipeline, lorsque la scène est peu ou pas texturée. Cette difficulté peut être surmontée grâce à l'apport du SfS, auquel nous nous intéressons dans le chapitre 2. Nous proposons notamment une nouvelle méthode de résolution du SfS, décrite initialement dans [139], qui est adaptée à des conditions d'éclairage naturel. Le chapitre 3, qui reprend les travaux présentés dans [111, 112], décrit une méthode de reconstruction 3D d'objets non texturés combinant MVS et SfS. Les hypothèses nécessaires à l'utilisation du SfS, à savoir la connaissance de la réflectance et de l'éclairage, sont reconsidérées dans le chapitre 4. Cette étude, fondée sur les travaux présentés dans [113, 114, 115], porte sur l'estimation conjointe de ces grandeurs à partir d'une géométrie grossière obtenue par MVS. En guise de conclusion, nous présentons un algorithme qui estime la réflectance et la profondeur de la surface d'une scène 3D, ainsi que l'éclairage, pour chaque pose de l'appareil photographique.

Chapitre 1

Photogrammétrie

Ce chapitre vise à présenter le pipeline « classique » de reconstruction 3D qui, à partir de photographies d'une scène 3D prises sous différents angles, fournit un modèle 3D de la scène. La facilité de prise en main de ce pipeline explique en partie son succès. Le paragraphe 1.1 présente le modèle sténopé, qui décrit la façon dont une scène 3D forme une image 2D par projection sur un plan. Dans le paragraphe 1.2, nous établissons les relations géométriques entre les poses d'un appareil photographique qui, en se déplaçant autour de la scène, permet d'obtenir un nuage de points, c'est-à-dire une reconstruction 3D faiblement dense de la scène. La méthode permettant de densifier la reconstruction 3D ainsi obtenue est présentée dans le paragraphe 1.3. Enfin, nous terminons ce chapitre en illustrant les limites de ce pipeline, notamment lorsque la scène est peu texturée.

1.1 De la scène 3D à la photographie

Trois éléments interviennent lors d'une prise de vue. Deux de ces éléments déterminent la **scène** : l'**objet 3D** photographié, qui possède un certain nombre de caractéristiques géométriques (position, orientation, etc.) et photométriques (couleur, texture, etc.), et **l'éclairage**. Enfin, après d'éventuelles réflexions sur la surface de l'objet, les rayons lumineux atteignent le troisième élément du dispositif : l'**appareil photographique**.

La reconstruction 3D photographique, ou **photogrammétrie**, permet d'obtenir un modèle 3D d'une scène à partir d'un ensemble de photographies. Elle nécessite de comprendre comment la scène 3D se projette sur le plan image de l'appareil photographique (nous utiliserons indifféremment l'anglicisme **caméra** pour désigner ce dernier). Le paragraphe ci-après présente le **modèle sténopé**, dont une description détaillée est fournie dans [65].

1.1.1 Scène 3D

Le modèle sténopé est fréquemment utilisé en vision par ordinateur, car il constitue un bon compromis entre simplicité et réalisme. Avant de nous pencher sur ce modèle, commençons par caractériser brièvement la scène 3D observée.

Nous ne nous intéressons dans ce mémoire qu'aux **objets opaques**, pour lesquels seule la **surface** $\mathcal{S}$ est visible, et laissons donc de côté les phénomènes de transparence et de transluminescence. Dorénavant, nous n'utiliserons que cette notion pour caractériser le relief à reconstruire. La surface est définie comme un ensemble de **points 3D**, chaque point étant repéré par ses coordonnées, relativement à un repère de référence.

En supposant la surface différentiable, nous pouvons définir, en tout point $\mathbf{X} \in \mathcal{S}$, la **normale à la surface** comme le vecteur orthogonal au plan tangent à $\mathcal{S}$, unitaire et

sortant. La géométrie de la surface $\mathcal{S}$ peut donc aussi être caractérisée par ses normales.

Ces notions étant introduites, nous pouvons à présent décrire géométriquement la projection d'un objet 3D en une image 2D.

1.1.2 Modèle sténopé

Un appareil photographique comporte deux éléments essentiels : l'objectif photographique, constitué d'un assemblage de lentilles, permet de projeter la scène 3D observée sur le plan image ; la surface photosensible capte les photons pendant un court intervalle de temps pour former l'image.

Bien que relativement complexe, le fonctionnement d'un tel système optique peut être décrit par le **modèle sténopé** de la *camera obscura* (« chambre noire »), qui permet de modéliser l'objectif photographique par une lentille unique. La création d'une image photographique se ramène alors à une projection centrale, de centre de projection le **centre optique** de l'appareil photographique, dite **projection perspective**.

En réalité, le modèle sténopé fait abstraction de l'aberration de distorsion [27]. Cette aberration peut être corrigée si les paramètres intrinsèques de la caméra sont connus. Nous supposons dorénavant que l'appareil photographique est calibré (cf. paragraphe 1.1.3) et que les images sont corrigées de l'aberration de distorsion.

1.1.2.1 Définition des repères usuels

Afin de modéliser géométriquement cette projection, nous devons définir les différents repères utilisés, qui sont illustrés sur la figure 1.1.

Repère caméra Un repère tridimensionnel direct est associé à l'appareil photographique, qui a pour origine le centre optique $\mathbf{C}$. Ses axes X et Y sont parallèles aux bords de l'image, et l'axe Z est parallèle à l'axe optique. Dans la mesure du possible, nous essaierons de choisir comme repère de référence le repère caméra d'une des poses de la caméra.

Repère image La projection perspective met en bijection chaque point visible $\mathbf{X} \in \mathbb{R}^3$ de la scène avec un point image $\mathbf{x} \in \mathbb{R}^2$. Le repère dans lequel les coordonnées de $\mathbf{x}$ sont définies a pour origine le **point principal** $\mathbf{x}_0$, qui se situe à l'intersection de l'axe optique et du plan image, ses axes x et y étant parallèles aux axes X et Y du repère caméra.

Repère pixels Une image numérique étant composée d'un ensemble régulier de pixels, celui-ci semble constituer un repère naturel pour les points de l'image. En assimilant les pixels à des carrés réguliers, et en choisissant le côté du carré comme unité de longueur, nous pouvons définir un nouveau repère 2D dont les axes u et v sont parallèles aux axes x et y du repère image, mais dont l'origine est généralement positionnée dans le coin supérieur gauche de l'image.

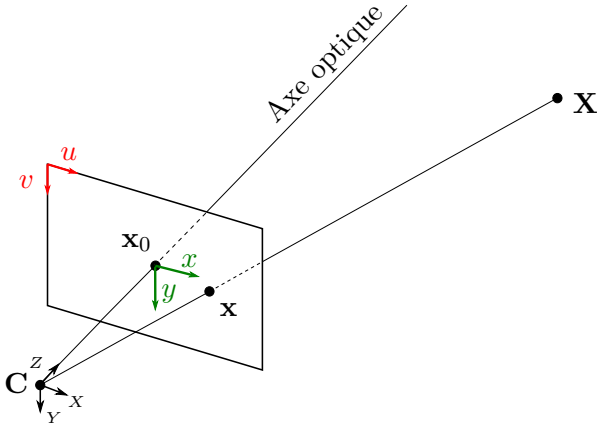


FIGURE 1.1 – Modèle sténopé. L'axe optique coupe le plan image en un point $\mathbf{x}_0$ appelé point principal. Le point 3D $\mathbf{X}$ se projette, par projection centrale de centre $\mathbf{C}$ sur le plan image, en un point qui peut être exprimé par ses coordonnées $\mathbf{x} = [x, y]^\top$ dans le repère image (en vert), ou par ses coordonnées $\mathbf{p} = [u, v]^\top$ dans le repère pixels (en rouge).

Distance focale La distance focale est l'une des principales caractéristiques d'un objectif photographique : f désigne la **distance focale image**, qui est égale à la distance entre le point principal $\mathbf{x}_0$ et le centre optique $\mathbf{C}$, généralement exprimée en millimètres.

1.1.3 Matrice de calibrage

Les deux repères 2D que nous avons définis utilisent deux unités différentes : celle du repère image est la même que celle de la scène 3D (généralement le millimètre), alors que celle du repère pixels est le pixel.

Si nous notons k le nombre de pixels par unité de longueur, les coordonnées d'un point image exprimées dans le repère image $\mathbf{x} = [x, y]^\top$ et dans le repère pixels $\mathbf{p} = [u, v]^\top$ sont liées par la relation affine suivante :

$$\begin{cases} u = k\,x + u_0 \\ v = k\,y + v_0 \end{cases} \tag{1.1}$$

où $[u_0, v_0]^\top$ désignent les coordonnées du point principal $\mathbf{x}_0$, exprimées dans le repère pixels. Afin d'intégrer les constantes u_0 et v_0 à une formulation matricielle, il est d'usage de réécrire ces expressions au moyen des coordonnées homogènes.

Coordonnées homogènes Pour un point $\mathbf{r} = [r_1, r_2, \ldots, r_n]^\top$ de $\mathbb{R}^n$, les coordonnées homogènes prennent la forme d'un vecteur $\widetilde{\mathbf{r}} = [r_1, r_2, \ldots, r_n, 1]^\top$ de $\mathbb{R}^{n+1}$. Le passage des coordonnées homogènes aux coordonnées cartésiennes s'effectue en divisant les n premiers

éléments de $\widetilde{\mathbf{r}}$ par son $n+1^{\text{ème}}$ élément. Par conséquent, deux ensembles de coordonnées homogènes proportionnels correspondent au même point de $\mathbb{R}^n$.

Soit $\widetilde{\mathbf{p}} = [u, v, 1]^\top$ et $\widetilde{\mathbf{x}} = [x, y, 1]^\top$ les vecteurs de coordonnées homogènes associés, respectivement, à $\mathbf{p}$ et à $\mathbf{x}$. La relation affine (1.1) peut être réécrite sous forme matricielle :

$$\widetilde{\mathbf{p}} = \begin{bmatrix} k & 0 & u_0 \\ 0 & k & v_0 \\ 0 & 0 & 1 \end{bmatrix} \widetilde{\mathbf{x}} \tag{1.2}$$

En outre, comme cela est indiqué sur la figure 1.2, le théorème de Thalès permet de relier les coordonnées d'un point 3D $\mathbf{X} = [X, Y, Z]^\top$ à celles de son image $\mathbf{x} = [x, y]^\top$:

$$\begin{cases} \dfrac{x}{\mathsf{f}} = \dfrac{X}{Z} \\ \\ \dfrac{y}{\mathsf{f}} = \dfrac{Y}{Z} \end{cases} \quad \Leftrightarrow \quad \begin{cases} x = \mathsf{f}\,\dfrac{X}{Z} \\ \\ y = \mathsf{f}\,\dfrac{Y}{Z} \end{cases} \tag{1.3}$$

Grâce aux équations (1.3), nous pouvons réécrire (1.2) :

$$\begin{aligned} \widetilde{\mathbf{p}} &= \begin{bmatrix} k & 0 & u_0 \\ 0 & k & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathsf{f}\,X/Z \\ \mathsf{f}\,Y/Z \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} k\,\mathsf{f} & 0 & u_0 \\ 0 & k\,\mathsf{f} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X/Z \\ Y/Z \\ 1 \end{bmatrix} \end{aligned} \tag{1.4}$$

La **matrice de calibrage** est définie comme suit :

$$\mathbf{K} = \begin{bmatrix} k\,\mathsf{f} & 0 & u_0 \\ 0 & k\,\mathsf{f} & v_0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1.5}$$

où $f = k\,\mathsf{f}$ est la distance focale exprimée en pixels. Cette matrice dépend de trois **paramètres intrinsèques** (ou paramètres internes) de l'appareil photographique, à savoir u_0, v_0 et f, qui sont tous trois exprimés en pixels. Nous introduisons le point 3D $\mathbf{y}$ défini de la façon suivante (cf. figure 1.2) :

$$\mathbf{y} = \begin{bmatrix} X/Z \\ Y/Z \\ 1 \end{bmatrix} \tag{1.6}$$

Grâce à (1.5) et (1.6), nous pouvons réécrire (1.4) :

$$\widetilde{\mathbf{p}} = \mathbf{K}\,\mathbf{y} \tag{1.7}$$

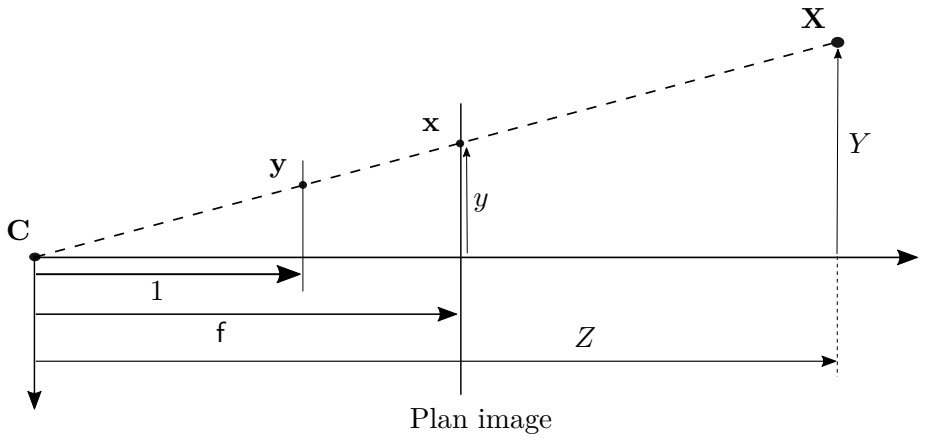


FIGURE 1.2 – Vue de profil de la projection centrale, de centre **C**, d'un point 3D **X** en un point **x** de l'image. Le point **y** est également obtenu par projection centrale, de centre **C**, du point 3D **X**, mais sur un plan parallèle au plan image distant de 1 du point **C**.

1.1.4 Paramétrage de la surface

Par projection, chaque point visible $\mathbf{X} = [X, Y, Z]^\top$ de la surface est mis en bijection avec un point image de coordonnées $\mathbf{p} = [u, v]^\top$ dans le repère pixels. La surface $\mathcal{S}$ peut donc être paramétrée par le **domaine de reconstruction** $\Omega \subset \mathbb{R}^2$, défini comme la projection de $\mathcal{S}$ sur le plan image :

$$\mathcal{S} = \left\{ \mathbf{X}(u, v), [u, v]^\top \in \Omega \right\} \tag{1.8}$$

L'ensemble Ω peut être défini indifféremment dans le repère image ou dans le repère pixels.

1.1.4.1 Fonction de profondeur

Des équations (1.6) et (1.7), nous déduisons :

$$\widetilde{\mathbf{p}} = \frac{\mathbf{K}}{Z} \mathbf{X} \tag{1.9}$$

qui devient, après inversion de la matrice $\mathbf{K}$:

$$\mathbf{X} = Z \begin{bmatrix} \frac{1}{f} & 0 & -\frac{u_0}{f} \\ 0 & \frac{1}{f} & -\frac{v_0}{f} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{1.10}$$

soit encore :

$$\mathbf{X}(u, v) = Z \begin{bmatrix} \frac{u - u_0}{f} \\ \frac{v - v_0}{f} \\ 1 \end{bmatrix} \tag{1.11}$$

Nous introduisons la fonction z, appelée **fonction de profondeur** (*depth map*), qui associe à chaque point image, de coordonnées $\mathbf{p} = [u, v]^\top$ dans le repère pixels, la troisième coordonnée $z(u, v) = Z$ du point 3D conjugué. Ce faisant, nous réécrivons (1.11) :

$$\mathbf{X}(u, v) = z(u, v) \begin{bmatrix} \frac{u - u_0}{f} \\ \frac{v - v_0}{f} \\ 1 \end{bmatrix} \tag{1.12}$$

La fonction de profondeur z constitue donc l'inconnue du problème de reconstruction 3D.

1.1.4.2 Projection orthogonale

Dans ce mémoire, nous traiterons aussi la **projection orthogonale**, que nous désignerons abusivement sous le terme de **projection orthographique**[1], pour laquelle le lien entre un point 3D et son image devient :

$$\mathbf{X}(u, v) = \begin{bmatrix} u \\ v \\ z(u, v) \end{bmatrix} \tag{1.13}$$

Cette projection est moins réaliste que la projection perspective, mais également plus simple. En particulier, il n'est plus nécessaire de positionner l'origine du repère 2D en un point particulier. En effet, la relation (1.13) montre qu'un déplacement du point principal de l'appareil photographique ne fait que translater la surface $\mathcal{S}$ en bloc, sans causer de déformation.

Pour ces deux types de projection, la connaissance de la profondeur $z(u, v)$ permet de déterminer sans ambiguïté le point 3D $\mathbf{X}(u, v)$ recherché, à ceci près que, dans le cas de la projection perspective, les paramètres intrinsèques (u_0, v_0, f) de l'appareil photographique doivent être connus.

1.2 Utilisation de plusieurs vues

Le pipeline usuel de reconstruction 3D est connu sous le nom de **photogrammétrie**. Son principe consiste à imiter la vision humaine, qui dispose de deux images prises sous deux angles différents. En l'étendant au cas de n images, il est possible d'obtenir des modèles 3D à grande échelle [2].

La première étape de ce pipeline, qui vise à estimer le déplacement de la caméra, permet d'obtenir conjointement une description « parcimonieuse » (c'est-à-dire faiblement dense) de la surface observée, ainsi que les différentes **poses** de la caméra. Elle procure donc la structure de la scène à partir du mouvement de l'appareil photographique. C'est pourquoi on l'appelle ***structure-from-motion*** (SfM).

1. Il s'agit effectivement d'un abus de langage, car la projection orthographique désigne un type de projection légèrement différent.

Un vaste pan de la littérature s'est intéressé aux algorithmes de SfM. Son utilisation en milieu industriel [119] fait que cette technique de reconstruction 3D reste d'actualité et que ses performances sont en constante amélioration [118]. S'il n'est pas dans nos objectifs de décrire le SfM en détail, il est quand même utile d'en donner les principales caractéristiques.

1.2.1 Généralisation des notations

La prise en compte de plusieurs poses de la caméra nécessite d'utiliser plusieurs repères. Dans ce contexte, et pour éviter toute ambiguïté, nous devons indexer la pose de la caméra. Nous notons $\mathbf{C}_i$ l'origine du repère caméra $\mathcal{R}_i$ associé à la $i^{\text{ème}}$ pose. Le repère caméra $\mathcal{R}_j$ est lié au repère caméra $\mathcal{R}_i$ par une transformation rigide de $\mathcal{R}_i$ vers $\mathcal{R}_j$ caractérisée par une matrice de rotation $\mathbf{R}_{i\to j}$ et un vecteur de translation $\mathbf{t}_{i\to j}$, tous deux exprimés dans le repère $\mathcal{R}_i$, de sorte que, pour un point 3D repéré par $\mathbf{X}_i = [X_i, Y_i, Z_i]^\top$ dans $\mathcal{R}_i$ et par $\mathbf{X}_j = [X_j, Y_j, Z_j]^\top$ dans $\mathcal{R}_j$, la formule de changement de repère s'écrit :

$$\mathbf{X}_i = \mathbf{R}_{i\to j}\,\mathbf{X}_j + \mathbf{t}_{i\to j} \tag{1.14}$$

Comme la matrice $\mathbf{R}_{i\to j}$ est orthogonale, cette égalité est facile à inverser :

$$\mathbf{X}_j = \mathbf{R}_{i\to j}^\top\,(\mathbf{X}_i - \mathbf{t}_{i\to j}) \tag{1.15}$$

En intervertissant les indices i et j dans (1.14), et en identifiant avec (1.15), il vient :

$$\mathbf{R}_{j\to i} = \mathbf{R}_{i\to j}^\top \qquad \text{et} \qquad \mathbf{t}_{j\to i} = -\mathbf{R}_{j\to i}\,\mathbf{t}_{i\to j} \tag{1.16}$$

En combinant (1.14) et la relation de proportionnalité (1.9), il est possible d'exprimer les coordonnées $\mathbf{p}_i$ de l'image d'un point 3D $\mathbf{X}$, exprimées dans le repère pixels de la $i^{\text{ème}}$ pose, en fonction des coordonnées $\mathbf{X}_j$ de ce point exprimées dans le repère caméra $\mathcal{R}_j$:

$$\widetilde{\mathbf{p}}_i \propto \mathbf{K}\,(\mathbf{R}_{i\to j}\,\mathbf{X}_j + \mathbf{t}_{i\to j}) \tag{1.17}$$

Si nous définissons la **matrice de projection**, de taille 3×4, par :

$$\mathbf{P}_{i\to j} = \mathbf{K}\,[\mathbf{R}_{i\to j}\,|\,\mathbf{t}_{i\to j}] \tag{1.18}$$

nous pouvons réécrire (1.17) sous une forme plus compacte :

$$\widetilde{\mathbf{p}}_i \propto \mathbf{P}_{i\to j}\,\widetilde{\mathbf{X}}_j \tag{1.19}$$

Le SfM se fondant sur les relations géométriques liant un point 3D de la scène à ses projections dans les différentes images, commençons par établir ces relations. Des analyses plus complètes de la géométrie multi-vues sont proposées dans [65] et [73].

1.2.2 Géométrie à deux vues

Comme le montre le schéma de la figure 1.3, l'utilisation d'une deuxième pose de la caméra permet de déterminer la position d'un point 3D par triangulation à partir de ses projections dans les deux images, à condition que l'on dispose de **paires de points homologues**, c'est-à-dire de paires de points $(\mathbf{x}_1, \mathbf{x}_2)$ correspondant à un même point $\mathbf{X}$.

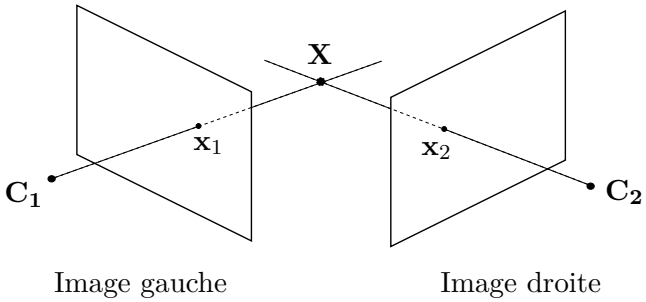


FIGURE 1.3 – L'utilisation de deux poses permet de déterminer la position du point 3D $\mathbf{X}$ par triangulation, à condition que l'on connaisse la paire de points homologues $(\mathbf{x}_1, \mathbf{x}_2)$.

1.2.2.1 Mise en correspondance

La mise en correspondance vise à trouver des éléments correspondants entre deux images. Ce processus peut être scindé en trois étapes : la détection d'éléments remarquables, leur description, puis leur appariement.

Les **points d'intérêt** sont des points de l'image possédant des caractéristiques remarquables, par exemple un fort contraste ou une texture particulière. Ils peuvent prendre la forme de points isolés, de segments, de courbes continues ou de régions. La recherche de points d'intérêt n'intéresse pas seulement la reconstruction 3D, mais également la reconnaissance et le suivi d'objets [20], la réalisation d'images panoramiques [29], la segmentation [144], etc. Nous faisons ci-dessous une présentation succincte des différentes étapes de la mise en correspondance. Il existe bien sûr des états de l'art [67], dont certains se focalisent sur l'étape de détection [162].

La première étape de la mise en correspondance extrait les points d'intérêt d'une image grâce à un **détecteur de points d'intérêt**. Une prise de décision locale détermine si une zone de l'image présente une caractéristique intéressante. Un détecteur de points d'intérêt est censé posséder les deux propriétés suivantes : **précision** et **robustesse**. Cette dernière propriété est primordiale, puisque l'invariance à certaines transformations géométriques (rotation, translation, changement d'échelle) ou photométriques (changement d'éclairage) permet d'accroître les chances de succès de l'étape d'appariement. Depuis le détecteur de coins de Harris et Stephens [63], première approche à avoir été utilisée à grande échelle de par son invariance à l'orientation de la structure, de nombreuses améliorations portant sur la robustesse et la répétabilité ont été proposées, jusqu'à l'apparition des détecteurs multi-échelles tels que le détecteur de points SIFT proposé par Lowe [106], ou ses variantes [176]. Des détecteurs plus récents comme SURF [17], certes moins précis, permettent d'accélérer sensiblement les calculs, voire de s'approcher du temps réel [6].

L'étape de **description** assigne ensuite une signature au point d'intérêt, de manière à pouvoir repérer les projections d'un même point 3D dans une séquence d'images. La robustesse déjà mentionnée aux transformations géométriques et photométriques est la principale propriété recherchée chez un descripteur. Il est notable que les étapes de détection et de description sont généralement effectuées par le même agent. On peut de nouveau citer

le descripteur SIFT [106], qui fait également figure de référence. Plusieurs améliorations de ce descripteur ont été proposées pour éviter une occupation mémoire trop importante, soit en le représentant de manière plus compacte [157], soit en lui préférant des descripteurs plus « légers » [30].

L'étape d'appariement consiste enfin à identifier les points présentant une forte similarité entre les deux images. Trois éléments sont nécessaires à cette dernière étape : la recherche des plus proches voisins, qui fournit les correspondances possibles entre images ; la métrique, qui mesure la similarité entre deux points potentiellement homologues et doit être choisie en fonction du détecteur utilisé ; une politique de rejet, qui valide les correspondances (grâce à un seuil ou à un ratio utilisant le deuxième plus proche voisin [106]).

1.2.2.2 Droites épipolaires

Soit un point $\mathbf{x}_1$ d'une image appelée **image gauche**. Nous cherchons à apparier ce point avec un point $\mathbf{x}_2$ d'une deuxième image de la même scène 3D, appelée **image droite**. Puisque $\mathbf{x}_1$ et $\mathbf{x}_2$ sont censés correspondre à un même point 3D $\mathbf{X}$, ces deux points, considérés comme des points de l'espace 3D, doivent se situer sur le plan défini par les points $\mathbf{C}_1$, $\mathbf{C}_2$ et $\mathbf{X}$. Or, ce plan coupe l'image gauche et l'image droite selon, respectivement, deux droites $\mathbf{D}_1$ et $\mathbf{D}_2$. Ces deux **droites épipolaires** sont donc appariées. Dans chaque image, l'ensemble des droites épipolaires forme un faisceau dont le centre est l'intersection du plan image et de la droite $(\mathbf{C}_1, \mathbf{C}_2)$. Ce point s'appelle l'**épipôle**. Ces différentes notions sont illustrées sur la figure 1.4.

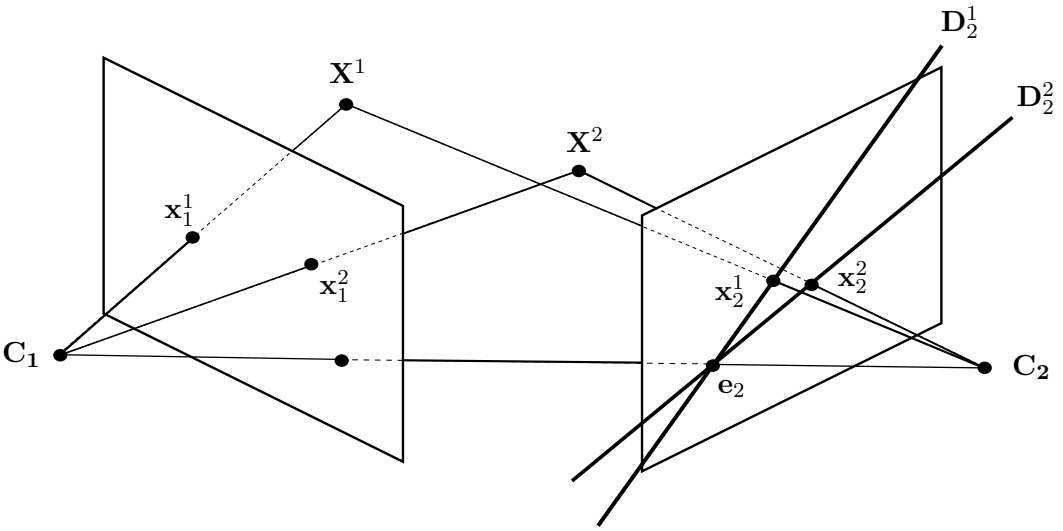


FIGURE 1.4 – Le rayon issu de $\mathbf{C}_1$ et passant par $\mathbf{x}_1^1$ se projette sur l'image droite en une droite $\mathbf{D}_2^1$ passant par le point $\mathbf{x}_2^1$ conjugué de $\mathbf{x}_1^1$. Les droites épipolaires de l'image droite forment un faisceau dont le centre est l'épipôle $\mathbf{e}_2$. Bien entendu, ces définitions peuvent être généralisées à l'image gauche, par un raisonnement en tout point symétrique.

1.2.2.3 Matrice fondamentale

Les relations entre points image et droites épipolaires peuvent être exprimées de manière compacte. Soit $\mathbf{X}$ un point 3D, et $\mathbf{x}_1$ et $\mathbf{x}_2$ ses projections dans les deux images. Essayons de caractériser la droite épipolaire $\mathbf{D}_2$ associée au point $\mathbf{x}_1$, c'est-à-dire la projection sur l'image droite du rayon issu de $\mathbf{C}_1$ et passant par $\mathbf{X}$. Pour cela, nous pouvons projeter dans l'image droite deux points 3D situés sur ce rayon, par exemple $\mathbf{X}^1 = \mathbf{C}_1$ et le point $\mathbf{X}^\infty$ situé à l'infini sur ce rayon. Il nous faut donc exprimer, dans le repère pixels de l'image droite, les coordonnées $\mathbf{p}_2^1$ et $\mathbf{p}_2^\infty$ des projections $\mathbf{x}_2^1$ et $\mathbf{x}_2^\infty$ dans l'image droite de $\mathbf{X}^1$ et $\mathbf{X}^\infty$.

Le point $\mathbf{X}^1 = \mathbf{C}_1$ a pour coordonnées homogènes $\widetilde{\mathbf{X}}_1^1 = [0, 0, 0, 1]^\top$ dans le repère $\mathcal{R}_1$. La formule de projection (1.19) dans l'image droite s'écrit donc :

$$\widetilde{\mathbf{p}}_2^1 \propto \mathbf{P}_{2\to1}\,[0, 0, 0, 1]^\top \tag{1.20}$$

Si nous nous restreignons momentanément au cas à deux poses, nous pouvons noter, pour simplifier, $\mathbf{R}$ pour $\mathbf{R}_{2\to1}$ et $\mathbf{t}$ pour $\mathbf{t}_{2\to1}$ dans la définition (1.18) de $\mathbf{P}_{2\to1}$:

$$\mathbf{P}_{2\to1} = \mathbf{K}\,[\mathbf{R}\,|\,\mathbf{t}] \tag{1.21}$$

ce qui nous permet de réécrire (1.20) :

$$\widetilde{\mathbf{p}}_2^1 \propto \mathbf{K}\,\mathbf{t} \tag{1.22}$$

D'autre part, les coordonnées $\mathbf{X}_1$ du point 3D $\mathbf{X}$, exprimées dans le repère $\mathcal{R}_1$, et les coordonnées $\mathbf{p}_1$ de son image $\mathbf{x}_1$, exprimées dans le repère pixels de l'image gauche, vérifient la relation suivante, d'après (1.9) :

$$\mathbf{X}_1 \propto \mathbf{K}^{-1}\,\widetilde{\mathbf{p}}_1 \tag{1.23}$$

Le point à l'infini $\mathbf{X}^\infty$ situé sur le rayon issu de $\mathbf{C}_1$ et passant par $\mathbf{X}$ a donc pour coordonnées homogènes $\widetilde{\mathbf{X}}_1^\infty = \begin{bmatrix}\mathbf{K}^{-1}\,\widetilde{\mathbf{p}}_1 \\ 0\end{bmatrix}$ dans le repère $\mathcal{R}_1$, et la projection de ce point dans l'image droite a pour coordonnées homogènes, d'après (1.19), (1.21) et (1.23) :

$$\widetilde{\mathbf{p}}_2^\infty \propto \mathbf{P}_{2\to1}\begin{bmatrix}\mathbf{K}^{-1}\,\widetilde{\mathbf{p}}_1 \\ 0\end{bmatrix} \quad\Rightarrow\quad \widetilde{\mathbf{p}}_2^\infty \propto \mathbf{K}\,\mathbf{R}\,\mathbf{K}^{-1}\,\widetilde{\mathbf{p}}_1 \tag{1.24}$$

Dans l'image droite, la droite $\mathbf{D}_2$ passant par les deux points $\mathbf{x}_2^1$ et $\mathbf{x}_2^\infty$ est définie par l'équation cartésienne suivante ($\wedge$ désigne le produit vectoriel) :

$$\widetilde{\mathbf{p}}_2^\top\left(\widetilde{\mathbf{p}}_2^1 \wedge \widetilde{\mathbf{p}}_2^\infty\right) = 0 \tag{1.25}$$

En utilisant les expressions (1.22) et (1.24), l'équation (1.25) se réécrit :

$$\widetilde{\mathbf{p}}_2^\top\left[(\mathbf{K}\,\mathbf{t}) \wedge \left(\mathbf{K}\,\mathbf{R}\,\mathbf{K}^{-1}\,\widetilde{\mathbf{p}}_1\right)\right] = 0 \tag{1.26}$$

Grâce à l'identité $(\mathbf{M}\,\mathbf{x}) \wedge (\mathbf{M}\,\mathbf{y}) = \mathbf{M}^{-\top}(\mathbf{x} \wedge \mathbf{y})$, l'équation (1.26) devient :

$$\widetilde{\mathbf{p}}_2^{\top}\,\mathbf{K}^{-\top}\left[\mathbf{t} \wedge \left(\mathbf{R}\mathbf{K}^{-1}\,\widetilde{\mathbf{p}}_1\right)\right] = 0 \tag{1.27}$$

Or, un produit vectoriel $\mathbf{v} \wedge \mathbf{w}$ peut s'écrire sous la forme d'un produit matriciel $[\mathbf{v}]_{\wedge}\,\mathbf{w}$, pourvu que la matrice $[\mathbf{v}]_{\wedge}$ associée à $\mathbf{v} = [v_1, v_2, v_3]^{\top}$ ait pour expression :

$$[\mathbf{v}]_{\wedge} = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix} \tag{1.28}$$

Cela nous permet de réécrire l'équation (1.27) :

$$\widetilde{\mathbf{p}}_2^{\top}\,\mathbf{K}^{-\top}\,[\mathbf{t}]_{\wedge}\,\mathbf{R}\,\mathbf{K}^{-1}\,\widetilde{\mathbf{p}}_1 = 0 \tag{1.29}$$

qui fait apparaître la **matrice fondamentale**, définie (à un facteur près) par :

$$\mathbf{F} \propto \mathbf{K}^{-\top}\,[\mathbf{t}]_{\wedge}\,\mathbf{R}\,\mathbf{K}^{-1} \tag{1.30}$$

De (1.29) et (1.30), nous tirons finalement la **contrainte épipolaire** :

$$\widetilde{\mathbf{p}}_2^{\top}\,\mathbf{F}\,\widetilde{\mathbf{p}}_1 = 0 \tag{1.31}$$

L'équation (1.31) fournit non seulement l'équation cartésienne de la droite épipolaire $\mathbf{D}_2$, sur laquelle doit se trouver le point homologue de $\mathbf{x}_1$, mais également l'équation cartésienne de la droite épipolaire $\mathbf{D}_1$, sur laquelle doit se trouver le point homologue de $\mathbf{x}_2$. Cette équation caractérise donc entièrement la **géométrie épipolaire**.

La matrice fondamentale $\mathbf{F}$ est définie seulement à un facteur multiplicatif près, puisque le produit de $\mathbf{F}$ par n'importe quel réel non nul ne change rien à la validité de la contrainte épipolaire (1.31). Par conséquent, huit éléments au plus de $\mathbf{F}$ sont indépendants. La matrice antisymétrique $[\mathbf{t}]_{\wedge}$ étant singulière, nous en déduisons que $\mathbf{F}$ est également singulière, et donc que son déterminant est nul. Cette contrainte supplémentaire diminue à sept au plus le nombre de degrés de liberté de la matrice fondamentale.

1.2.2.4 Matrice essentielle

Dans le cas d'un appareil photographique calibré, il est possible d'exprimer les coordonnées des points dans les repères image, auquel cas la contrainte épipolaire (1.31) se simplifie. En effet, l'équation (1.7) implique :

$$\mathbf{K}^{-1}\,\widetilde{\mathbf{p}}_1 = \mathbf{y}_1 \quad \text{et} \quad \mathbf{K}^{-1}\,\widetilde{\mathbf{p}}_2 = \mathbf{y}_2 \tag{1.32}$$

En réinjectant ces deux égalités dans (1.29), il vient :

$$\mathbf{y}_2^{\top}\,[\mathbf{t}]_{\wedge}\,\mathbf{R}\,\mathbf{y}_1 = 0 \tag{1.33}$$

Cette équation fait apparaître la **matrice essentielle**[2], définie (à un facteur près) par :

$$\mathbf{E} \propto [\mathbf{t}]_\wedge \mathbf{R} \tag{1.34}$$

grâce à quoi nous pouvons réécrire la contrainte épipolaire (1.33) :

$$\mathbf{y}_2^\top \mathbf{E}\, \mathbf{y}_1 = 0 \tag{1.35}$$

La matrice essentielle $\mathbf{E}$ représente la géométrie épipolaire calibrée, puisque le calibrage de la caméra est nécessaire à la définition des repères image. Elle possède les mêmes propriétés que $\mathbf{F}$. En particulier, elle est définie seulement à un facteur près.

La matrice essentielle peut être estimée par l'algorithme dit « des 8 points » [105], qui consiste en une minimisation en moindres carrés linéaires. Si, en théorie, cet algorithme permet également d'estimer la matrice fondamentale, on lui préfère généralement l'algorithme normalisé des 8 points [64], connu pour être numériquement plus stable.

1.2.3 Estimation de la pose

La matrice $\mathbf{E}$, une fois estimée, permet d'estimer le déplacement de la caméra entre deux poses. On parle d'estimation du mouvement ou d'**estimation de la pose**. Entre les deux poses associées aux repères $\mathcal{R}_1$ et $\mathcal{R}_2$, ce déplacement est défini par une matrice de rotation $\mathbf{R}$ et un vecteur de translation $\mathbf{t}$ (translation du centre optique), qui sont liés à la matrice essentielle $\mathbf{E}$ par la relation (1.34).

Les méthodes d'estimation de la pose à partir de la matrice essentielle, telles qu'elles sont décrites dans [65], mettent en évidence une ambiguïté sur le vecteur $\mathbf{t}$, calculable seulement à un facteur près. Même en fixant l'échelle de la reconstruction 3D par la contrainte $\|\mathbf{t}\| = 1$, ce qui revient à imposer (arbitrairement) au centre optique de se déplacer, entre deux poses consécutives, d'une distance égale à 1, il reste encore deux possibilités pour le vecteur $\mathbf{t}$. À cette ambiguïté s'ajoute une autre ambiguïté sur la matrice de rotation $\mathbf{R}$, pour laquelle il existe deux solutions.

1.2.3.1 Triangulation

La triangulation vise à évaluer la position d'un point 3D à partir de ses projections dans les différentes images, connaissant les poses de la caméra. Nous traitons pour commencer le cas de figure minimal de deux poses.

En théorie, le point 3D $\mathbf{X}$ doit se situer à l'intersection des rayons $(\mathbf{C}_1\, \mathbf{y}_1)$ et $(\mathbf{C}_2\, \mathbf{y}_2)$. Or, de par la définition (1.6), les coordonnées de $\mathbf{X}$ dans le repère $\mathcal{R}_1$ sont liées à $\mathbf{y}_1$ par :

$$\mathbf{X}_1 = Z_1\, \mathbf{y}_1 \tag{1.36}$$

où $Z_1 \in \mathbb{R}$ désigne la profondeur de $\mathbf{X}$ dans ce repère. De même, les coordonnées de $\mathbf{X}$ dans le repère $\mathcal{R}_2$ sont liées à $\mathbf{y}_2$ par :

$$\mathbf{X}_2 = Z_2\, \mathbf{y}_2 \tag{1.37}$$

2. On notera tout de même que la matrice essentielle a été introduite par Longuet-Higgins [105] avant la matrice fondamentale, qui apparaît donc comme une généralisation de la matrice essentielle.

où $Z_2 \in \mathbb{R}$ désigne la profondeur de $\mathbf{X}$ dans $\mathcal{R}_2$. En utilisant la formule (1.15) de changement de repère, il vient :

$$\mathbf{X}_1 = \mathbf{R}^\top (Z_2\, \mathbf{y}_2 - \mathbf{t}) \tag{1.38}$$

La triangulation consiste donc simplement à trouver le couple de réels (Z_1, Z_2) vérifiant :

$$Z_1\, \mathbf{y}_1 = \mathbf{R}^\top (Z_2\, \mathbf{y}_2 - \mathbf{t}) \tag{1.39}$$

En pratique, comme les données sont bruitées, les rayons $(\mathbf{C}_1\, \mathbf{y}_1)$ et $(\mathbf{C}_2\, \mathbf{y}_2)$ ne s'intersectent pas. Il faut donc se contenter d'une solution approchée. Plusieurs méthodes sont décrites dans [66]. On peut par exemple résoudre le problème suivant :

$$(\widehat{Z}_1, \widehat{Z}_2) = \underset{(Z_1,Z_2)\in\mathbb{R}^2}{\operatorname{argmin}} \| Z_1\, \mathbf{y}_1 - \mathbf{R}^\top (Z_2\, \mathbf{y}_2 - \mathbf{t}) \|^2 \tag{1.40}$$

puis choisir le point milieu entre $\widehat{Z}_1\, \mathbf{y}_1$ et $\mathbf{R}^\top \Big(\widehat{Z}_2\, \mathbf{y}_2 - \mathbf{t} \Big)$.

Validation de la paire $(\mathbf{R}, \mathbf{t})$ Comme nous l'avons déjà dit, quatre couples $(\mathbf{R}, \mathbf{t})$ sont déductibles de la matrice essentielle. Après triangulation, il s'avère qu'une seule de ces solutions situe la quasi-totalité des points 3D à l'avant de la caméra.

1.2.4 SfM à n vues

De nombreuses variantes de la technique du SfM ont été proposées, qui peuvent s'inspirer d'une approche séquentielle [28], hiérarchique [163] ou globale [164]. Nous décrivons ici, dans un but pédagogique, une stratégie naïve qui consiste à prendre en compte une troisième image de la même scène.

À ce niveau du raisonnement, nous disposons d'une première reconstruction 3D parcimonieuse de la scène obtenue à partir de deux images, le repère de référence étant le repère caméra de la première pose. L'image droite devient alors l'image gauche, tandis que la troisième image devient l'image droite. Nous pouvons à nouveau estimer un couple $(\mathbf{R}, \mathbf{t})$ décrivant le déplacement entre les poses 2 et 3, et reconstruire un nuage de points 3D parcimonieux caractéristique de la scène. Un simple changement de repère permet ensuite de retrouver le couple $(\mathbf{R}, \mathbf{t})$ décrivant le déplacement entre les poses 1 et 3. Ainsi, de proche en proche, l'ensemble des poses peuvent être estimées, ce qui permet de densifier la reconstruction 3D de la scène.

Cette méthode de reconstruction 3D, qui semble très naturelle, est particulièrement pertinente si l'on dispose d'une séquence d'images ordonnée, comme cela est le cas dans le cadre de l'odométrie visuelle ou du SLAM (*Simultaneous Localization And Mapping*) [160]. Néanmoins, chaque nuage de points 3D est reconstruit à une échelle différente (rappelons que la matrice $\mathbf{E}$ est estimée à un facteur d'échelle près). Or, si les différentes échelles ne sont pas cohérentes entre elles, il en ira forcément de même pour les nuages de points 3D.

1.2.4.1 Estimation de l'échelle a posteriori

Restreignons-nous pour simplifier au cas de $n = 3$ poses. Supposons qu'un nuage de m points 3D $\mathbf{X}^k$, $k \in \{1, \dots, m\}$, ait pu être estimé par triangulation, à partir des deux premières poses. Les coordonnées homogènes des projections dans la troisième image de ces points 3D, qui ont pour coordonnées $\mathbf{X}_2^k$ dans le repère $\mathcal{R}_2$, s'écrivent, d'après (1.19) :

$$\widetilde{\mathbf{p}}_3^k \propto \mathbf{P}_{3\to 2}\, \widetilde{\mathbf{X}}_2^k, \qquad k \in \{1, \dots, m\} \tag{1.41}$$

Or, grâce à l'étape de mise en correspondance, nous connaissons les positions de ces projections dans la troisième image. La matrice de rotation $\mathbf{R}_{3\to 2}$ de la pose 2, relativement à la pose 3, est invariante par changement d'échelle, mais la translation $\mathbf{t}_{3\to 2}$ du centre optique est connue seulement à un facteur $\alpha \in \mathbb{R}$ près, qui constitue l'inconnue du problème d'estimation de l'échelle a posteriori. Par conséquent, la matrice de projection $\mathbf{P}_{3\to 2}$ dépend de α, dont l'estimation peut se ramener à la résolution du problème suivant, qui est un problème en moindres carrés non linéaires :

$$\widehat{\alpha} = \underset{\alpha\in\mathbb{R}}{\operatorname{argmin}} \; \frac{1}{m}\sum_{k=1}^{m} \left\| \mathbf{p}_3^k - \frac{\left([\mathbf{R}_{3\to 2} \,|\, \alpha\, \mathbf{t}_{3\to 2}]\; \widetilde{\mathbf{X}}_2^k\right)_{1,2}}{\left([\mathbf{R}_{3\to 2} \,|\, \alpha\, \mathbf{t}_{3\to 2}]\; \widetilde{\mathbf{X}}_2^k\right)_{3}} \right\|^2 \tag{1.42}$$

Il est facile de linéariser ce problème, en multipliant chaque résidu par le dénominateur :

$$\widehat{\alpha} = \underset{\alpha\in\mathbb{R}}{\operatorname{argmin}} \; \frac{1}{m}\sum_{k=1}^{m} \left\| \mathbf{p}_3^k \left([\mathbf{R}_{3\to 2} \,|\, \alpha\, \mathbf{t}_{3\to 2}]\, \widetilde{\mathbf{X}}_2^k\right)_3 - \left([\mathbf{R}_{3\to 2} \,|\, \alpha\, \mathbf{t}_{3\to 2}]\, \widetilde{\mathbf{X}}_2^k\right)_{1,2} \right\|^2 \tag{1.43}$$

En pratique, les solutions « industrielles » préfèrent à cette solution l'estimation conjointe de l'échelle et de la pose. Une liste des méthodes de résolution de ce dernier problème, appelé PnP (pour *Perspective-n-Point*), est détaillée dans [100].

1.2.4.2 Ajustement de faisceaux

Dans la version de SfM que nous venons de présenter, les poses de la caméra sont estimées l'une après l'autre. Cette manière de procéder est sujette à des effets de dérive et d'accumulation des erreurs. Pour éviter de tels problèmes, il est d'usage d'effectuer a posteriori des itérations d'**ajustement de faisceaux**. Plus connu sous sa dénomination anglaise de ***bundle adjustment***, ce procédé consiste en une optimisation non linéaire visant à réduire les erreurs résiduelles de reprojection des points 3D dans les différentes images. Une synthèse complète des méthodes d'ajustement de faisceaux peut être trouvée dans [166]. Nous en présentons ici uniquement le principe.

L'ajustement de faisceaux consiste à estimer simultanément les poses des caméras et les positions des points 3D, de manière à minimiser l'ensemble des erreurs de reprojection. Ce problème constitue donc une généralisation de (1.42) :

$$\left(\left\{\widehat{\mathbf{P}}_{i\to\text{réf}}\right\}_i, \left\{\widehat{\mathbf{X}}_{\text{réf}}^k\right\}_k\right) = \underset{\{\mathbf{P}_{i\to\text{réf}}\}_i, \left\{\mathbf{X}_{\text{réf}}^k\right\}_k}{\operatorname{argmin}} \; \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\sum_{k=1}^{m} \left\| \mathbf{p}_i^k - \frac{\left(\mathbf{P}_{i\to\text{réf}}\, \widetilde{\mathbf{X}}_{\text{réf}}^k\right)_{1,2}}{\left(\mathbf{P}_{i\to\text{réf}}\, \widetilde{\mathbf{X}}_{\text{réf}}^k\right)_{3}} \right\|^2 \tag{1.44}$$

où n désigne le nombre de poses préalablement estimées et m le nombre de points 3D, et où l'indice « réf » désigne le repère de référence.

Pour simplifier l'écriture de (1.44), nous avons supposé que les points 3D $\mathbf{X}^k$ étaient visibles dans toutes les images. En pratique, il faudrait tenir compte de la visibilité [50]. Le problème (1.44) est un problème d'optimisation en moindres carrés non linéaires, qui peut être résolu par la méthode de Levenberg-Marquardt [108]. Il s'agit cependant d'un problème non convexe : le choix d'une initialisation appropriée est donc crucial.

1.3 Stéréoscopie multi-vues

Plus connue sous la dénomination anglaise de ***multi-view stereo*** (MVS), cette technique de reconstruction 3D est complémentaire du SfM. À partir des poses de la caméra estimées par SfM, le MVS fournit une fonction de profondeur, c'est-à-dire une reconstruction 3D dense, pour chaque pose. Cette méthode utilise l'information fondamentale contenue dans les images, à savoir les niveaux de gris (ou les niveaux de couleur, le cas échéant). Une vue d'ensemble des pratiques liées au MVS est présentée dans [47].

1.3.1 Surface lambertienne

Une surface est lambertienne si la clarté de l'image d'un point 3D est indépendante de la pose de la caméra. On parle également de surface mate ou « diffusante ». Une proportion non négligeable des surfaces réelles présentent un comportement proche de celui d'une surface lambertienne.

Le niveau de gris de l'image d'un point 3D de la scène est donc indépendant de la pose de l'appareil photographique. Contrairement à l'étape préliminaire de SfM du pipeline de reconstruction 3D, le MVS fait cette hypothèse en chaque point de la surface, grâce à quoi nous pouvons espérer obtenir une reconstruction 3D dense.

1.3.2 Modélisation du MVS

Soit $\mathbf{X}$ un point 3D situé sur une surface opaque $\mathcal{S}$, qui se projette en un point image de coordonnées $\mathbf{p}$ dans le repère pixels. La projection centrale π, définie par $\mathbf{p} = \pi(\mathbf{X})$, est inversible si l'on connaît la fonction de profondeur z. Dans ce cas, il existe une bijection entre les points 3D visibles de la scène et leurs images, qui s'écrit $\mathbf{X} = \pi_z^{-1}(\mathbf{p})$, où l'indice z sert à indiquer que, sans la connaissance de la fonction z, cette écriture serait ambiguë.

Nous disposons de n images de la même scène 3D, et des n poses correspondantes de la caméra. La première pose peut être choisie comme référence. Pour un point 3D $\mathbf{X}$ visible dans toutes les images, notons $\mathbf{p} = \pi(\mathbf{X})$ les coordonnées de la projection de ce point dans l'image de référence, et $\mathbf{p}_i = \pi_i(\mathbf{X})$, $i \in \{2, \dots, n\}$, les coordonnées de ses projections dans les $n-1$ autres images, appelées « images témoins ». La loi de Lambert permet d'écrire :

$$I_i(\mathbf{p}_i) = I(\mathbf{p}), \qquad i \in \{2, \dots, n\} \tag{1.45}$$

où I_i et I désignent, respectivement, les fonctions de niveau de gris de l'image i et de l'image de référence. Pour un point $\mathbf{p}$ de l'ensemble Ω, qui regroupe les points de l'image

de référence visibles dans toutes les images témoins :

$$I_i \circ \pi_i \circ \pi_z^{-1}(\mathbf{p}) = I(\mathbf{p}), \qquad i \in \{2, \dots, n\}, \quad \mathbf{p} \in \Omega \tag{1.46}$$

La stéréoscopie multi-vues recherche la fonction de profondeur z correspondant à l'image de référence qui maximise la **cohérence photométrique** de celle-ci avec les $n-1$ images témoins. Les équations (1.46) ne sont jamais vérifiées exactement, entre autres parce que les surfaces réelles ne sont jamais parfaitement lambertiennes. En considérant z comme une fonction inconnue, nous pouvons reformuler les équations (1.46) en un problème en moindres carrés :

$$\widehat{z} = \underset{z\colon \Omega \to \mathbb{R}}{\operatorname{argmin}} \ \frac{1}{n-1} \sum_{i=2}^{n} \left\| I_i \circ \pi_i \circ \pi_z^{-1} - I \right\|_{\ell^2(\Omega)}^2 \tag{1.47}$$

Sachant que nous ne connaissons pas les fonctions de niveau de gris, mais seulement leurs valeurs en chaque pixel, il nous faut reformuler l'équation (1.47) sous forme discrète, en notant $\mathbf{z}$ le vecteur contenant l'ensemble des valeurs $z(\mathbf{p})$ de la profondeur associées aux pixels $\mathbf{p} \in \Omega$:

$$\widehat{\mathbf{z}} = \underset{\mathbf{z} \in \mathbb{R}^{|\Omega|}}{\operatorname{argmin}} \ \frac{1}{n-1} \sum_{i=2}^{n} \frac{1}{|\Omega|} \sum_{\mathbf{p} \in \Omega} \left| I_i \circ \pi_i \circ \pi_z^{-1}(\mathbf{p}) - I(\mathbf{p}) \right|^2 \tag{1.48}$$

Dans cette formulation discrète, nous ne connaissons pas la valeur $I_i \circ \pi_i \circ \pi_z^{-1}(\mathbf{p})$. En effet, il s'agit d'un point de l'image témoin i dont les coordonnées ne sont pas entières (sauf cas exceptionnel). Cette valeur doit donc être approchée par interpolation (au plus proche voisin, bilinéaire, bicubique, etc.). Le problème (1.48) se simplifie, car il peut être résolu séparément en chaque pixel $\mathbf{p} \in \Omega$, ce qui signifie qu'il est séparable :

$$\widehat{z}(\mathbf{p}) = \underset{z(\mathbf{p}) \in \mathbb{R}}{\operatorname{argmin}} \ \frac{1}{n-1} \sum_{i=2}^{n} \left| I_i \circ \pi_i \circ \pi_z^{-1}(\mathbf{p}) - I(\mathbf{p}) \right|^2 \tag{1.49}$$

Dans les équations (1.47), (1.48) et (1.49), la cohérence photométrique se réduit à la comparaison en moindres carrés de deux niveaux de gris. En pratique, tant dans les éléments qu'elle vise à comparer que dans la forme, la cohérence photométrique est plus complexe (cf. le panorama très complet présenté dans [47, chapitre 2]). Nous reviendrons plus en détail sur ces variantes dans le chapitre 3. Notons pour le moment que la fonction à minimiser peut se révéler non linéaire, non dérivable et/ou non convexe, ce qui rend l'optimisation potentiellement difficile. C'est pourquoi la minimisation est généralement effectuée par une recherche **en force brute** (*brute-force*) sur un ensemble de valeurs prédéfinies de la profondeur. Cette stratégie, introduite initialement dans [68], peut paraître simpliste, mais s'avère très efficace pour l'estimation de fonctions de profondeur de scènes 3D fortement texturées [55]. La figure 1.5 en atteste : l'optimisation de la seule cohérence photométrique permet d'estimer une fonction de profondeur dont la forme globale est satisfaisante du point de vue qualitatif. Des détails géométriques très fins sont même retrouvés, malgré la présence de quelques artéfacts qui peuvent être facilement éliminés par une étape de post-traitement [48, 71].

FIGURE 1.5 – Reconstruction 3D par MVS d'une scène fortement texturée. En haut à gauche : image de référence. En haut à droite : quatre images témoins (parmi 77). En bas à gauche : fonction de profondeur correspondant à l'image de référence, obtenue avec la suite logicielle AliceVision [7] par optimisation de la cohérence photométrique. Bien que cette fonction comporte quelques artéfacts, certains détails géométriques très fins sont effectivement retrouvés, comme le montre l'agrandissement en bas à droite.

Une fois qu'une fonction de profondeur par image a été estimée, le pipeline se poursuit par la fusion des fonctions de profondeur en un maillage 3D [24, 81], puis par le plaquage d'une texture sur le maillage [101]. Ceci permet d'obtenir un modèle 3D complet de la scène, comme le montre l'exemple de la figure 1.6. La qualité d'une telle reconstruction 3D témoigne de la maturité des pipelines de photogrammétrie. Notons néanmoins que, faute de mieux, ce sont les images elles-mêmes qui sont utilisées pour tenir lieu de texture : la texture plaquée sur le maillage mélange donc réflectance et ombrage, ce qui compromet un potentiel rééclairage réaliste de la scène et met en évidence une première limitation inhérente aux pipelines de photogrammétrie, sur laquelle nous reviendrons dans le chapitre 4.

FIGURE 1.6 – À gauche : maillage 3D obtenu en sortie de la suite logicielle AliceVision [7], par filtrage puis fusion de 78 fonctions de profondeur telles que celle de la figure 1.5. À droite : modèle 3D complet, après plaquage d'une texture sur le maillage. Comme cette texture provient directement des images d'entrée, elle mélange réflectance et ombrage, ce qui compromet un potentiel rééclairage réaliste de la scène.

Les pipelines de photogrammétrie présentent une deuxième limitation importante : la qualité des résultats dépend très fortement du caractère texturé de la scène à reconstruire. Cela est illustré sur la figure 1.7, où la fonction de profondeur obtenue par optimisation de la cohérence photométrique est tellement bruitée qu'elle ne comporte aucun détail géométrique. Il n'est possible de réduire le bruit et retrouver les détails du relief qu'en fusionnant de très nombreuses fonctions de profondeur. Mais, outre le problème déjà mentionné de la texture, qui mélange réflectance et ombrage, la reconstruction 3D reste perfectible au niveau des détails, qui sont fortement lissés. Comment faire en sorte que les fonctions de profondeur estimées reproduisent effectivement les détails du relief ? Les chapitres 2 et 3 proposent plusieurs réponses, qui toutes font appel à la technique du *shape-from-shading*.

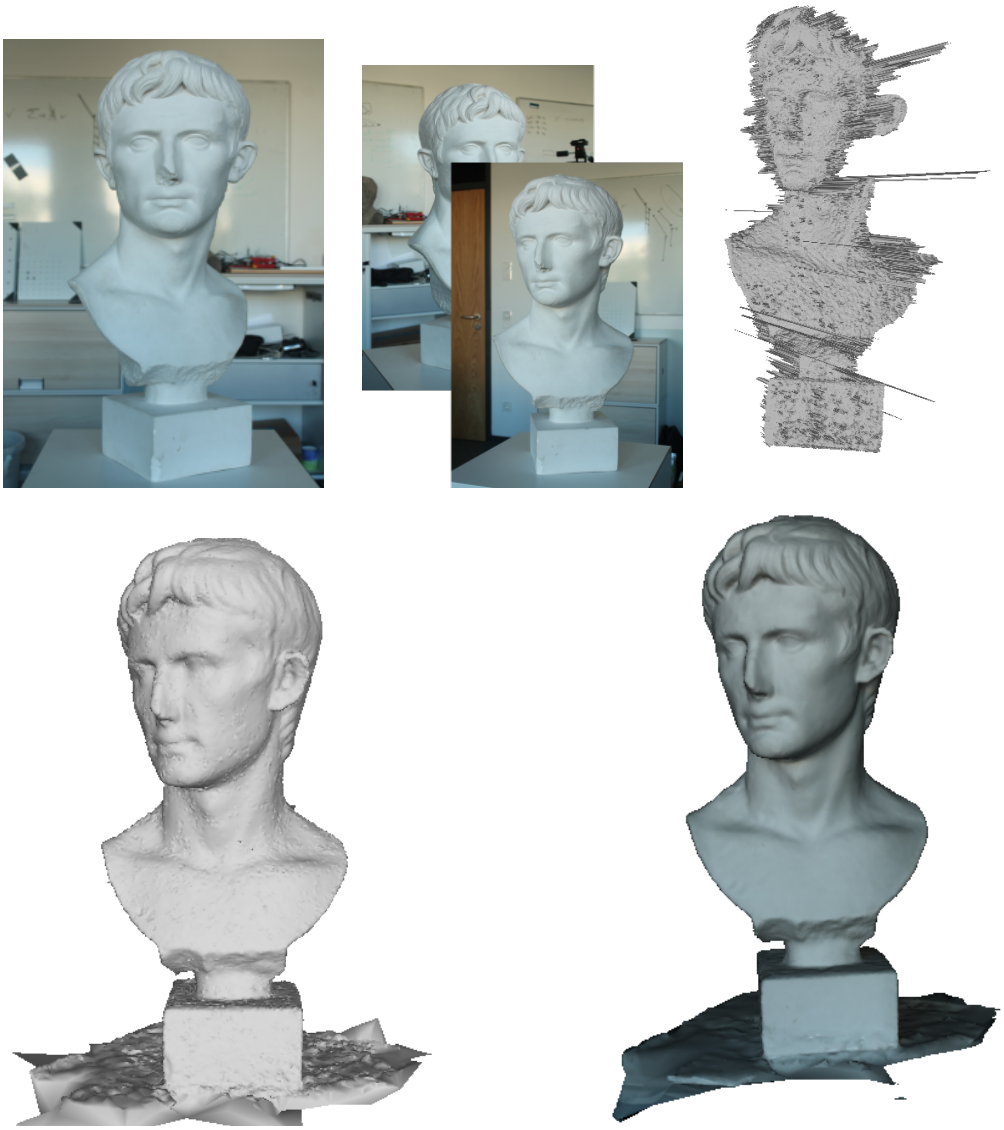


FIGURE 1.7 – Reconstruction 3D par MVS d'une scène peu texturée. En haut, de gauche à droite : image de référence, deux images témoins (parmi 49), et fonction de profondeur correspondant à l'image de référence, obtenue avec la suite logicielle AliceVision [7] par optimisation de la cohérence photométrique. Cette fonction est tellement bruitée qu'elle ne reproduit aucun détail géométrique. En bas à gauche : maillage 3D obtenu par fusion de 50 fonctions de profondeur. Des artéfacts dus au bruit restent présents, et les détails géométriques sont fortement lissés, par exemple au niveau des yeux. En bas à droite : modèle 3D complet. Alors que la texture devrait être uniforme, puisque le matériau est homogène, l'ombrage présent dans les images d'entrée est très nettement visible.

Conclusion

Les pipelines de photogrammétrie existants, fondés sur le SfM et le MVS, constituent des solutions technologiques matures au problème de la reconstruction 3D. Ils restent néanmoins perfectibles, à cause de deux limitations importantes. D'une part, les fonctions de profondeur estimées ne reproduisent pas les détails du relief lorsque la scène est peu texturée. D'autre part, les images d'entrée sont directement plaquées sur le relief, en guise de texture. Or, ces images mélangent réflectance et ombrage, ce qui peut être problématique pour le rééclairage.

La suite de ce mémoire vise à surmonter ces deux limitations à l'aide d'une autre catégorie de techniques de reconstruction 3D, à savoir les **techniques photométriques**. Ces techniques utilisent un modèle physique qui lie la scène 3D, l'éclairage et l'appareil photographique, afin de déterminer le relief de la scène. En particulier, le ***shape-from-shading*** (SfS) permet d'interpréter les variations de niveau de gris d'une image par des caractéristiques géométriques denses de la scène 3D, y compris dans les zones non texturées. Le SfS semble donc capable de combler les lacunes du MVS. De plus, les techniques photométriques permettent d'estimer la réflectance de la surface et l'éclairage de la scène. En plaquant sur le relief estimé non plus les images d'entrée, mais la réflectance estimée, le rééclairage de la scène n'en sera que plus réaliste.

Le principal inconvénient des méthodes photométriques est qu'elles nécessitent des conditions d'éclairage très contrôlées, que seules peuvent garantir des photographies prises en laboratoire. Néanmoins, l'utilisation conjointe de ces méthodes et du pipeline classique de reconstruction 3D semble prometteuse. Comme nous allons le voir dans le chapitre 2, l'utilisation d'un relief initial obtenu par SfM-MVS permet à elle seule d'atteindre des performances très encourageantes, bien que les méthodes photométriques n'utilisent généralement qu'une pose de la caméra. L'extension de ces méthodes au cas multi-vues, qui sera proposée dans le chapitre 3, permet de combiner les avantages très complémentaires de ces deux approches.

Chapitre 2

Shape-from-shading en conditions d'éclairage naturel

Le champ d'application du pipeline de reconstruction 3D présenté dans le chapitre 1 est limité puisque, par principe, il ne peut reconstruire que des scènes texturées. Nous avons vu que, pour une scène peu ou pas texturée, la cohérence photométrique n'était pas suffisamment discriminante pour que la reconstruction 3D soit satisfaisante. Le recours à d'autres techniques de reconstruction 3D semble nécessaire.

Les techniques de reconstruction 3D photométriques interprètent le niveau de gris d'une image comme la combinaison de trois entités : le relief de la scène 3D, la réflectance et l'éclairage. Si le relief et l'éclairage ont déjà été évoqués dans le chapitre 1, il nous faut à présent les caractériser du point de vue photométrique. Enfin, la troisième entité à prendre en compte est la **réflectance** de la surface, qui décrit la façon dont cette dernière réémet la lumière. Après avoir défini plus précisément cette notion, nous décrivons dans le paragraphe 2.1 le modèle photométrique de formation de l'image qui sera utilisé dans ce mémoire. Nous présentons ensuite, dans le paragraphe 2.2, le problème de la reconstruction 3D mono-oculaire, connu sous le nom de ***shape-from-shading***. Enfin, nous proposons dans le paragraphe 2.3 une nouvelle méthode de résolution de ce problème publiée dans [139].

2.1 Réflectance

La réflectance d'une surface permet de décrire la façon dont elle réémet la lumière, c'est-à-dire de faire le lien entre la lumière reçue et la lumière réémise par la surface. C'est cette dernière qui, indirectement, « s'imprime » sur la surface photosensible de l'appareil photographique. Le but de ce paragraphe est de montrer comment l'aspect visuel d'une photographie est lié aux caractéristiques géométriques et photométriques de la scène 3D. Pour définir précisément la réflectance, il nous faut auparavant introduire quelques notions de photométrie. Les explications fournies dans ce paragraphe sont inspirées de [41].

2.1.1 Définitions photométriques

Comme dans le chapitre 1, nous nous intéressons à une scène 3D dont la surface est opaque. La description locale de l'éclairage de cette surface est donnée par la notion d'**éclairement** E, qui est défini comme le flux lumineux incident par unité de surface.

L'énergie lumineuse réémise par la surface est décrite par la notion de **luminance émise** L dans la direction $\mathbf{u}_e$, qui correspond à la « clarté apparente » de cette surface. La luminance émise ne doit pas être confondue avec la **luminance incidente** $\bar{L}$, qui

caractérise l'énergie lumineuse reçue depuis la direction $\mathbf{u}_i$. La réflectance constitue le lien entre éclairement et luminance émise.

Pour les matériaux opaques, la description la plus fine de la réémission de la lumière est donnée par la fonction de distribution de la réflectance bidirectionnelle r, plus connue sous le terme de **réflectance** [125] ou sous l'acronyme anglais **BRDF** (pour *Bidirectional Reflectance Distribution Function*). Pour un élément de surface recevant un flux lumineux incident $\bar{L}(\mathbf{u}_i)$ à l'intérieur d'un angle solide élémentaire $\mathrm{d}\Omega_i$ autour de la direction $\mathbf{u}_i$, la luminance élémentaire $\mathrm{d}L(\mathbf{u}_e)$ émise dans la direction $\mathbf{u}_e$ est proportionnelle à l'éclairement élémentaire $\mathrm{d}E$. Le rapport de proportionnalité entre ces éléments est égal, par définition, à la réflectance r. Si l'on caractérise les directions $\mathbf{u}_i$ et $\mathbf{u}_e$ par les angles (θ_i, ϕ_i) et (θ_e, ϕ_e), définis comme la colatitude et l'azimut dans un repère orthonormé d'axe vertical porté par le vecteur $\mathbf{n}$ normal à la surface (cf. figure 2.1), la réflectance r s'écrit :

$$r(\theta_i, \phi_i, \theta_e, \phi_e) = \frac{\mathrm{d}L(\mathbf{u}_e)}{\mathrm{d}E} \tag{2.1}$$

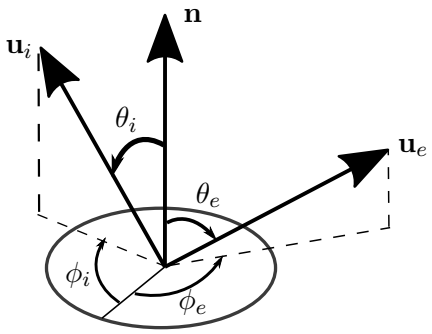


FIGURE 2.1 – Définition des angles θ_i, ϕ_i, θ_e et ϕ_e.

Il est montré dans [41] que l'éclairement élémentaire est lié à la luminance incidente par la relation suivante :

$$\mathrm{d}E = \bar{L}(\mathbf{u}_i)\,\mathrm{d}\Omega_i\,\cos\theta_i \tag{2.2}$$

De plus :

$$\mathrm{d}\Omega_i = \sin\theta_i\,\mathrm{d}\theta_i\,\mathrm{d}\phi_i \tag{2.3}$$

Nous tirons des égalités (2.1) à (2.3) une formulation intégrale de la luminance émise dans la direction $\mathbf{u}_e$:

$$L(\mathbf{u}_e) = \int_{\theta_i=0}^{\pi/2}\int_{\phi_i=0}^{2\pi} r(\theta_i, \phi_i, \theta_e, \phi_e)\bar{L}(\mathbf{u}_i)\,\sin\theta_i\,\cos\theta_i\,\mathrm{d}\theta_i\,\mathrm{d}\phi_i \tag{2.4}$$

Il est notable que l'ensemble de ces grandeurs photométriques sont définies localement, c'est-à-dire qu'elles dépendent du point $\mathbf{X}$ de la surface.

2.1.2 Surface lambertienne

L'expression intégrale (2.4) de la luminance est généralement trop complexe pour être utilisable. Nous nous restreignons dans ce mémoire aux **surfaces lambertiennes**, pour lesquelles la réflectance est indépendante non seulement de la direction $\mathbf{u}_i$ de la lumière incidente, mais également de la direction d'émission $\mathbf{u}_e$: $r(\theta_i, \phi_i, \theta_e, \phi_e) = r_0$. Cette hypothèse est justifiée par le fait que beaucoup de surfaces réelles ont un comportement proche de celui d'une surface lambertienne.

L'**albédo** ρ est une grandeur sans unité, comprise entre 0 et 1, qui mesure la proportion d'énergie lumineuse réémise par la surface. Dans le cas d'une surface lambertienne, la réflectance est entièrement caractérisée par l'albédo. On montre que la luminance s'exprime alors très simplement [41] :

$$L(\mathbf{X}) = \frac{\rho(\mathbf{X})}{\pi} E_s(\mathbf{X}) \tag{2.5}$$

où $E_s(\mathbf{X})$ désigne l'éclairement du point $\mathbf{X}$ de la surface.

En réalité, la BRDF, et donc l'albédo, peuvent dépendre de la longueur d'onde. En toute rigueur, il faudrait manipuler la **BRDF spectrale**. Dans un souci de simplicité, nous supposons dans un premier temps que l'appareil photographique mesure des niveaux de gris et que l'albédo est un scalaire. Nous reviendrons sur ce point dans le chapitre 4, où nous aborderons la question de la couleur.

Pour une source lumineuse ponctuelle, l'éclairage incident en un point $\mathbf{X}$ de la surface peut être décrit par un vecteur $\mathbf{s}(\mathbf{X})$, dont la norme $\psi(\mathbf{X}) = \|\mathbf{s}(\mathbf{X})\|$ est la **densité du flux lumineux** :

$$\begin{array}{rccl} \mathbf{s}: & \mathbb{R}^3 & \to \mathbb{R}^3 & \\ & \mathbf{X} & \mapsto \mathbf{s}(\mathbf{X}) = \psi(\mathbf{X})\, \bar{\mathbf{s}}(\mathbf{X}) & \end{array} \tag{2.6}$$

où $\bar{\mathbf{s}}(\mathbf{X})$ est un vecteur unitaire. Pour simplifier les notations, il est d'usage d'orienter ce vecteur vers la source. En supposant la source lumineuse infiniment distante de la scène, l'éclairage est **parallèle et uniforme**, c'est-à-dire que sa direction et son intensité sont les mêmes en tout point éclairé de la surface : $\mathbf{s}(\mathbf{X}) = \mathbf{s}_0$.

Notons que, tout comme pour l'albédo, le champ vectoriel $\mathbf{s}$ dépend a priori de la longueur d'onde. Dans ce mémoire, nous négligeons cette dépendance spectrale, c'est-à-dire que nous supposons que la source est « blanche ».

Si l'éclairage en un point $\mathbf{X}$ d'une surface est caractérisé par le vecteur $\mathbf{s}(\mathbf{X}) \in \mathbb{R}^3$, l'éclairement en ce point vaut :

$$E_s(\mathbf{X}) = \mathbf{s}(\mathbf{X})^\top \mathbf{n}(\mathbf{X}) \tag{2.7}$$

où $\mathbf{n}(\mathbf{X})$ désigne la normale à la surface en $\mathbf{X}$, sur laquelle nous reviendrons en détail dans le paragraphe 2.2.1. Des égalités (2.5) et (2.7), nous tirons l'expression de la luminance d'une surface lambertienne :

$$L(\mathbf{X}) = \frac{\rho(\mathbf{X})}{\pi} \mathbf{s}(\mathbf{X})^\top \mathbf{n}(\mathbf{X}) \tag{2.8}$$

2.1.3 Réflectance et niveau de gris

Nous supposons à présent que la surface $\mathcal{S}$ est photographiée par une caméra en niveaux de gris, et nous notons $I : \Omega \subset \mathbb{R}^2 \to \mathbb{R}^+$ la fonction de niveau de gris de l'image obtenue.

Comme cela est montré dans [41], un appareil photographique calibré photométriquement peut servir d'instrument de mesure de la luminance. En effet, si les sources d'assombrissement à la périphérie de l'image, à savoir le « vignetage » et le coefficient « $\cos^4 \alpha$ » [3], sont corrigées par un calibrage approprié, l'éclairement $E_i(u, v)$ de l'image au point image $\mathbf{p} = [u, v]^\top$ est proportionnel à la luminance émise par le point 3D conjugué $\mathbf{X}(u, v)$:

$$E_i(u, v) \propto L(\mathbf{X}(u, v)) \tag{2.9}$$

D'autre part, après correction des non linéarités de la fonction de réponse du capteur photosensible [37, 58, 143, 146], le niveau de gris est proportionnel à l'éclairement [1] :

$$I(u, v) \propto E_i(u, v) \tag{2.10}$$

Le coefficient de proportionnalité de l'équation (2.10) dépend de plusieurs facteurs tels que l'ouverture de l'objectif, le temps de pose, le grandissement, etc. Des équations (2.8), (2.9) et (2.10), nous tirons :

$$I(u, v) \propto \rho(u, v)\, \mathbf{s}(u, v)^\top \mathbf{n}(u, v) \tag{2.11}$$

où nous utilisons les notations simplifiées $\rho(u, v) = \rho(\mathbf{X}(u, v))$, $\mathbf{s}(u, v) = \mathbf{s}(\mathbf{X}(u, v))$ et $\mathbf{n}(u, v) = \mathbf{n}(\mathbf{X}(u, v))$.

Le coefficient de proportionnalité de la relation (2.11) peut être intégré à l'albédo ρ. Dans la suite de ce mémoire, $\rho(u, v)$ ne sera plus tenu à être compris entre 0 et 1, mais nous continuerons à l'appeler albédo. Nous écrirons donc dorénavant la relation (2.11) sous la forme d'une égalité :

$$I(u, v) = \rho(u, v)\, \mathbf{s}(u, v)^\top \mathbf{n}(u, v) \tag{2.12}$$

tout en sachant qu'il s'agit en réalité d'une relation de proportionnalité.

2.1.4 Écarts au modèle lambertien

La relation (2.12), qui lie sous une forme particulièrement simple les caractéristiques d'un point d'une surface lambertienne au niveau de gris de son image, présente de nombreux avantages. En particulier, elle modélise bien de nombreux matériaux, bien qu'un certain nombre de phénomènes provoquent des écarts à ce modèle.

Réflexions mutuelles La formulation (2.4) de la luminance n'est pas complète, car chaque point de la surface se comporte lui-même comme une source lumineuse supplémentaire. Dans le cas où la surface présente des concavités, ce flux secondaire éclaire d'autres points de la surface, ces points devenant à leur tour des sources lumineuses supplémentaires.

1. En pratique, le niveau de gris est noyé dans le bruit pour les points très sombres, et est limité par un phénomène de saturation pour les points très brillants.

À cause de ces **réflexions mutuelles**, les niveaux de gris mesurés sont généralement supérieurs aux valeurs prédites par le modèle (2.4), en particulier au voisinage des zones concaves. En toute rigueur, ce modèle de luminance devrait être remplacé par la solution d'une équation, connue en synthèse d'images sous le nom d'**équation du rendu** [86]. Vu que la résolution de cette équation constitue un problème à part entière, nous préférons négliger les réflexions mutuelles.

Reflets brillants La réflectance lambertienne $r(\theta_i, \phi_i, \theta_e, \phi_e) = r_0$ caractérise bien les matériaux mats. Bien entendu, aucun matériau réel n'est parfaitement lambertien, mais la plupart des matériaux mats sont plutôt bien modélisés par la loi de Lambert (2.12), qui présente l'avantage d'être linéaire. De nombreux autres modèles de réflectance ont été proposés pour modéliser soit l'aspect brillant de certains matériaux [116], le plus connu de cette catégorie étant le modèle de Phong [132], soit leur « rugosité » [128], qui découle de la modélisation de la surface par des micro-facettes. Comme pour les réflexions mutuelles, nous négligeons les reflets brillants, que nous traitons comme des données aberrantes.

Ombre propre Pour un éclairage donné, l'**ombre propre** est définie comme l'ensemble des points 3D $\mathbf{X}$ tels que $\mathbf{s}(\mathbf{X})^\top \mathbf{n}(\mathbf{X}) \leq 0$. Nous devrions remplacer l'expression (2.7) de l'éclairement par :

$$E_s(\mathbf{X}) = \max\{\mathbf{s}(\mathbf{X})^\top \mathbf{n}(\mathbf{X}), 0\} \tag{2.13}$$

mais cela ferait perdre à la loi de Lambert (2.12) son caractère linéaire. Dans la présentation générale du *shape-from-shading* du paragraphe 2.2, nous considérons pour simplifier que les points de l'ombre propre sont des données aberrantes. Mais dans le paragraphe 2.3, nous présentons une nouvelle méthode de résolution du SfS qui tient compte de l'ombre propre.

Ombre portée Il n'est pas nécessaire que $\mathbf{s}(\mathbf{X})^\top \mathbf{n}(\mathbf{X}) \leq 0$ pour qu'un point $\mathbf{X}$ soit dans l'ombre. En effet, de nombreux points tels que $\mathbf{s}(\mathbf{X})^\top \mathbf{n}(\mathbf{X}) \geq 0$ ne sont pas éclairés, car la lumière qui devrait les atteindre est arrêtée par un obstacle. Ces points constituent **l'ombre portée** de $\mathbf{s}$. Contrairement à l'ombre propre, qui peut être caractérisée localement, l'ombre portée dépend de la géométrie globale de la scène. Le contexte de la reconstruction 3D accentue la différence entre ces deux types d'ombres : s'il est possible de tenir compte explicitement de l'ombre propre avec le modèle (2.13), cela constitue un problème ouvert pour l'ombre portée. Dans ce mémoire, nous considérons les points de l'ombre portée comme des données aberrantes, qui peuvent provoquer un biais dans l'estimation de l'albédo.

2.2 Reconstruction 3D par *shape-from-shading*

Le ***shape-from-shading*** (SfS) est une technique de reconstruction 3D qui utilise comme indice l'**ombrage** (***shading***). La notion d'ombrage est définie dans [42] comme « l'ensemble des phénomènes à cause desquels un même élément de surface matérielle apparaît plus ou moins clair sur une photographie ». Dans la loi de Lambert (2.12), l'ombrage correspond au produit scalaire $\mathbf{s}(u, v)^\top \mathbf{n}(u, v)$.

La technique du SfS a été développée dans les années 1970 au MIT, sous l'impulsion de Horn [74], dont le but initial était de mettre au point un système de reconnaissance automatique de visages. Dans ce paragraphe, nous revenons sur l'équation (2.12) et son utilisation dans le cadre du SfS.

La prise en compte des réflexions mutuelles et de l'ombre portée rend le problème difficile, puisqu'il ne peut alors être résolu que globalement. Or, comme cela a déjà été dit, nous négligeons ces écarts au modèle lambertien et les considérons dorénavant comme des données aberrantes, tout comme les reflets brillants.

Dans un premier temps, tout en négligeant l'ombre propre, nous cherchons à reconstruire une surface lambertienne éclairée par un flux lumineux parallèle et uniforme caractérisé par le vecteur $\mathbf{s}_0$, c'est-à-dire que nous simplifions encore le modèle (2.12) :

$$I(u,v) = \rho(u,v)\,\mathbf{s}_0^\top \mathbf{n}(u,v) \tag{2.14}$$

2.2.1 Modélisation du SfS

Commençons par modéliser le SfS dans le cas de la projection orthographique, qui est moins réaliste que le modèle perspectif, mais plus simple. Le SfS vise à estimer le relief de la surface observée à partir d'un modèle tel que (2.14). Nous cherchons donc à évaluer la normale $\mathbf{n}(u,v)$ qui, par définition, appartient à la sphère unité $\mathbb{S}^2 \subset \mathbb{R}^3$. La simple observation de l'équation (2.14) suffit à montrer le caractère mal posé du SfS, qui a constitué un sujet d'étude à part entière [91, 126]. En effet, si l'on dispose d'une seule image (en niveaux de gris) de la scène, cette équation comporte trois inconnues scalaires en chaque point (u,v) : une pour l'albédo $\rho(u,v)$ plus les deux degrés de liberté de $\mathbf{n}(u,v)$.

2.2.1.1 Ambiguïtés du SfS

Une première ambiguïté survient si l'albédo est inconnu, puisque l'albédo $\rho(u,v)$ et l'ombrage $\mathbf{s}_0^\top \mathbf{n}(u,v)$ peuvent alors être intervertis dans (2.14). Cette ambiguïté est connue sous le nom de **trompe-l'œil**. Même si l'albédo est connu, une deuxième ambiguïté survient lorsque l'éclairage est inconnu. Cette ambiguïté, connue sous le nom d'**ambiguïté de bas-relief**, a été formalisée dans [18]. Comme nous le verrons plus tard, l'éclairage peut être soit préalablement calibré, soit estimé à partir d'une connaissance, même imprécise, du relief. Un cas particulier de l'ambiguïté de bas-relief est l'**ambiguïté concave/convexe** [130]. Sur l'exemple de la figure 2.2, comme le système visuel humain suppose généralement que la lumière vient d'en haut, l'interprétation du relief dépend de l'orientation de l'image [2].

Afin de nous focaliser sur l'estimation du relief de la surface, nous supposons que $\rho(u,v)$ est connu (en pratique, nous posons $\rho(u,v) = 1$) et que l'éclairage $\mathbf{s}_0$, supposé parallèle et uniforme, est connu. Ces hypothèses peuvent être évitées par une estimation simultanée de la géométrie, de l'éclairage et de la réflectance [13], mais nous nous concentrons dans un premier temps sur l'estimation de la géométrie. Une estimation conjointe de la géométrie, de l'albédo et de l'éclairage sera présentée dans le chapitre 4.

2. Cet effet ne fonctionnerait pas avec une scène usuelle, pour laquelle l'interprétation du relief par le système visuel humain utilise des informations dites « de haut niveau ».

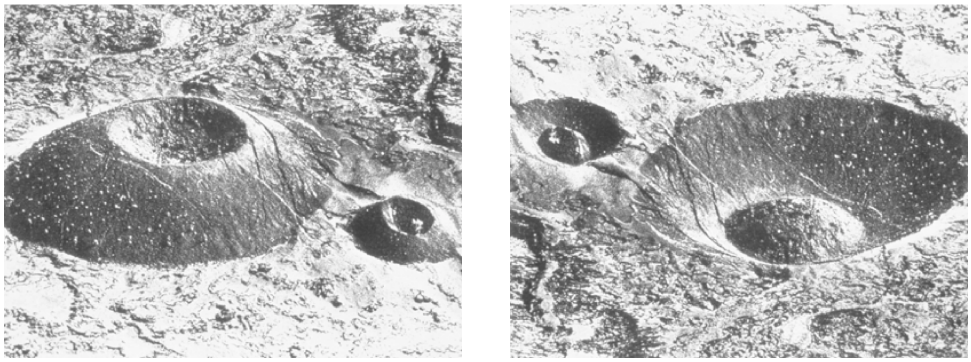


FIGURE 2.2 – À gauche : photographie de deux monticules de cendres des îles Hawaii. À droite : après retournement de la photographie, les monticules sont « devenus » des cratères. Cet exemple illustre l'ambiguïté concave/convexe inhérente au SfS : un même relief peut être interprété de différentes manières. Cette illustration provient de [130] (crédit photographique : Whitman Richards).

Ces hypothèses (plus ou moins justifiées) sur l'albédo et sur l'éclairage ne suffisent quand même pas à faire du SfS un problème bien posé. En effet, en notant $\theta(u,v)$ l'angle formé par les vecteurs $\mathbf{s}_0$ et $\mathbf{n}(u,v)$, et en supposant que $\|\mathbf{s}_0\| = 1$, la loi (2.14) devient :

$$I(u,v) = \cos\theta(u,v) \tag{2.15}$$

Cette équation montre que les normales vérifiant (2.14) forment un cône ayant pour axe le vecteur d'éclairage $\mathbf{s}_0$. La figure 2.3 illustre cette ambiguïté.

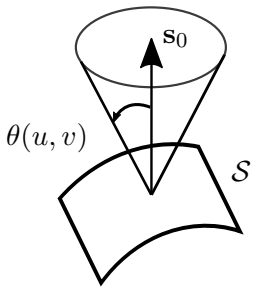


FIGURE 2.3 – Si l'albédo $\rho(u,v)$ et l'éclairage $\mathbf{s}_0$ sont connus, les normales $\mathbf{n}(u,v)$ vérifiant l'équation (2.14) forment un cône ayant pour axe le vecteur d'éclairage $\mathbf{s}_0$. D'après (2.15), le demi-angle au sommet de ce cône vaut $\theta(u,v) = \arccos I(u,v)$.

2.2.1.2 Expression de la normale

Avant de détailler les équations du SfS, il nous faut commencer par expliciter le champ de normales à la surface. En effet, l'équation (2.14) fait apparaître la normale, dont l'existence même suppose que la fonction z est différentiable, ce que nous faisons dorénavant.

En utilisant le paramétrage de la surface par les coordonnées (u, v) dans le repère pixels de l'image, nous définissons le champ de normales sur le domaine de reconstruction Ω :

$$\begin{array}{rccl} \mathbf{n}: & \Omega & \to \mathbb{S}^2 \subset \mathbb{R}^3 \\ & (u,v) & \mapsto \mathbf{n}(u,v) = [n_1(u,v), n_2(u,v), n_3(u,v)]^\top \end{array} \tag{2.16}$$

Pour une surface différentiable, la normale est colinéaire au produit vectoriel des deux dérivées partielles $\partial_u \mathbf{X}(u,v)$ et $\partial_v \mathbf{X}(u,v)$ de $\mathbf{X}(u,v)$:

$$\mathbf{n}(u,v) = \pm \frac{\partial_u \mathbf{X}(u,v) \wedge \partial_v \mathbf{X}(u,v)}{\|\partial_u \mathbf{X}(u,v) \wedge \partial_v \mathbf{X}(u,v)\|} \tag{2.17}$$

Sous l'hypothèse de la projection orthographique, nous déduisons de (1.13) :

$$\partial_u \mathbf{X}(u,v) = [1, 0, z_u(u,v)]^\top \quad \text{et} \quad \partial_v \mathbf{X}(u,v) = [0, 1, z_v(u,v)]^\top \tag{2.18}$$

Comme l'axe z est orienté vers la scène (cf. figure 1.1), la troisième composante de la normale, exprimée dans le repère caméra, doit être négative. Nous déduisons donc facilement des égalités (2.17) et (2.18) :

$$\mathbf{n}(u,v) = \frac{1}{\sqrt{\|\nabla z(u,v)\|^2 + 1}} \begin{bmatrix} \nabla z(u,v) \\ -1 \end{bmatrix} \tag{2.19}$$

Cette expression montre que la linéarité du modèle (2.14), relativement aux composantes de la normale, est perdue vis-à-vis de la profondeur z. Cela est dû au dénominateur du membre droit de (2.19), dont la présence s'explique par la contrainte $\|\mathbf{n}(u,v)\| = 1$.

2.2.1.3 Équation eikonale

Si l'éclairage, supposé parallèle et uniforme, est caractérisé par le vecteur $\mathbf{s}_0 = [s_1, s_2, s_3]^\top$, les équations (2.14) et (2.19) nous permettent d'établir le lien entre la géométrie de la scène 3D et le niveau de gris de l'image associée :

$$\frac{I(u,v)}{\rho(u,v)} \sqrt{\|\nabla z(u,v)\|^2 + 1} - [s_1, s_2]^\top \nabla z(u,v) + s_3 = 0 \tag{2.20}$$

Sous l'hypothèse $\rho(u,v) = 1$, l'équation (2.20) devient :

$$I(u,v) \sqrt{\|\nabla z(u,v)\|^2 + 1} - [s_1, s_2]^\top \nabla z(u,v) + s_3 = 0 \tag{2.21}$$

La formulation la plus simple du SfS suppose de plus que l'éclairage est frontal et de norme égale à 1, c'est-à-dire que $\mathbf{s}_0 = [0, 0, -1]^\top$. L'équation (2.21) devient alors :

$$\|\nabla z(u,v)\| = \sqrt{\frac{1}{I(u,v)^2} - 1} \tag{2.22}$$

qui est connue sous le nom d'**équation eikonale**. Cette équation très simple suffit à mettre en évidence l'ambiguïté concave/convexe. En effet, pour toute solution z_0 de (2.22), il est clair que $-z_0$ est également solution.

L'estimation de la profondeur par SfS consiste, sous ces hypothèses, à résoudre l'équation (2.22), qui est une équation aux dérivées partielles (EDP) non linéaire du premier ordre. Mais une autre modélisation du SfS, non différentielle, est également très utilisée.

2.2.1.4 Modélisation non différentielle du SfS

La modélisation non différentielle du SfS nécessite l'introduction d'inconnues p et q supplémentaires, égales aux dérivées partielles de la fonction de profondeur [41] :

$$\begin{cases} p = z_u \\ q = z_v \end{cases} \tag{2.23}$$

L'équation eikonale (2.22) devient alors :

$$p(u,v)^2 + q(u,v)^2 = \frac{1}{I(u,v)^2} - 1 \tag{2.24}$$

La résolution en (p, q) de cette équation ne suffit pas à résoudre le problème du SfS. En effet, une deuxième étape est nécessaire, qui consiste à calculer la profondeur z par résolution des équations (2.23), c'est-à-dire à résoudre un problème d'**intégration**. Contrairement aux apparences, l'intégration d'un champ de normales constitue un problème délicat, qui a été étudié en détail dans [137].

2.2.2 Résolution du SfS

La résolution du SfS a donné lieu à une littérature abondante. Trois états de l'art sur les méthodes de résolution ont paru depuis les premiers travaux de Horn [43, 76, 178].

2.2.2.1 Résolution des modèles différentiels

La première catégorie de méthodes de résolution vise à résoudre de manière exacte le modèle différentiel du SfS, qui prend par exemple la forme de l'EDP non linéaire (2.21), en supposant la surface différentiable (presque partout). Historiquement, la première approche proposée par Horn [74] tentait de résoudre cette EDP non linéaire par la méthode des caractéristiques. Depuis, des méthodes numériques plus efficaces ont été proposées, par exemple grâce aux « solutions de viscosité » [35, 44, 103, 134, 148] ou aux « courbes de niveau » [89, 90]. Si elles ne nécessitent aucun réglage de paramètre, ces méthodes manquent cependant parfois de robustesse et nécessitent généralement une condition au bord.

2.2.2.2 Résolution des modèles non différentiels

Contrairement aux modèles différentiels, les modèles non différentiels du SfS sont tellement mal contraints que l'utilisation d'une condition au bord ne suffit pas à rendre la solution unique [41]. Une possibilité consiste à utiliser un modèle de surface local [34, 130]. De biens meilleurs résultats peuvent être obtenus grâce à la « théorie de la régularisation », qui consiste à ajouter au modèle des a priori sur la solution, de manière à mieux contraindre le problème et à garantir la régularité de la solution. Les méthodes variationnelles permettent de trouver une solution approchée à cette catégorie de problèmes. Mais si l'ajout d'un a priori est simple, le réglage des paramètres peut se révéler fastidieux. Cette approche a été introduite par Horn et Brooks pour résoudre le SfS [75], afin d'ajouter un a priori d'intégrabilité sur le champ vectoriel $[p, q]^\top$ [45], a priori qui a été réutilisé dans différents travaux [159] et sur lequel nous reviendrons dans le paragraphe 2.3.4.

2.3 Comment rendre le SfS opérationnel

De nombreuses approches ont cherché à supprimer, ou du moins à assouplir, les hypothèses du SfS. En effet, les méthodes de résolution du SfS sont limitées par le nombre d'hypothèses plus ou moins réalistes qu'elles requièrent. Et, même en reformulant le problème de manière à le rendre bien posé, la robustesse de la solution est difficile à garantir. Non seulement la résolution de l'équation liant le niveau de gris au gradient de profondeur peut être difficile en soi, mais la présence de données aberrantes risque fort, en pratique, de dégrader le relief obtenu.

Dans ce paragraphe, nous commençons par présenter le moyen le plus simple pour rendre le SfS bien posé, qui consiste à utiliser plusieurs photographies prises sous différents éclairages. La mise en œuvre de cette technique très performante nécessite néanmoins de bien contrôler l'éclairage, ce qui peut s'avérer contraignant. Nous proposons ensuite une nouvelle modélisation générique du SfS, qui reste valide pour différents jeux d'hypothèses et permet en particulier, conformément à l'objectif visé, d'utiliser le SfS en milieu extérieur. Enfin, nous proposons un nouvelle méthode de résolution numérique de ce modèle, que nous validons par des tests sur données de synthèse et sur données réelles.

2.3.1 Stéréophotométrie

Le moyen le plus connu pour faire de la reconstruction 3D photométrique un problème bien posé est d'utiliser $n > 1$ images prises sous le même point de vue, mais sous différents éclairages. Si la surface est lambertienne et que les éclairages sont parallèles et uniformes, c'est-à-dire caractérisés par des vecteurs $\mathbf{s}_i$, $i \in \{1, \dots, n\}$, l'équation (2.14) peut être remplacée par le système suivant :

$$I_i(u, v) = \rho(u, v)\, \mathbf{s}_i^\top \mathbf{n}(u, v), \quad i \in \{1, \dots, n\} \tag{2.25}$$

Dans cette variante du SfS, appelée **stéréophotométrie** (ou *photometric stereo*) [172], l'albédo $\rho(u, v)$ et la normale $\mathbf{n}(u, v)$ peuvent être estimés sans aucune ambiguïté, dès lors que $n \geq 3$ éclairages $\mathbf{s}_i$ non coplanaires sont utilisés.

De nombreuses variantes de cette technique ont été proposées. Par exemple, elle a été étendue au cas où les éclairages ne sont pas calibrés [16, 140], à la prise en compte des différents canaux colorés des images [138], à des réflectances non lambertiennes [110]. Une étude récente très complète de la stéréophotométrie a été menée dans [141].

Une variante astucieuse de la stéréophotométrie consiste à prendre une seule photographie en couleur d'une scène 3D éclairée par trois sources lumineuses colorées (en rouge, vert et bleu). En utilisant les trois canaux de l'image, il est effectivement possible d'estimer le relief d'une surface blanche par stéréophotométrie [83]. Cette idée a été mise à profit pour effectuer la reconstruction 3D d'une scène déformable en temps réel [70].

Le champ d'application de nos recherches ne permet malheureusement pas d'utiliser la stéréophotométrie. En règle générale, l'éclairage est à la fois « naturel », c'est-à-dire non contrôlé, et identique d'une photographie à l'autre. Il nous faut donc nous cantonner au cadre du SfS classique, tout en relâchant certaines hypothèses pour nous approcher le plus possible des conditions opératoires visées.

2.3.2 Vers des hypothèses plus réalistes

2.3.2.1 Projection perspective

La projection perspective (cf. chapitre 1) est plus réaliste que la projection orthographique, dont nous avons fait l'hypothèse jusqu'à présent. De nombreux travaux se sont intéressés au SfS sous l'hypothèse de la projection perspective [26, 99, 161], d'autant que le SfS devient alors mieux posé et permet de relâcher certaines contraintes [134].

Afin de rendre la méthode de résolution que nous proposons la plus générique possible, il nous faudra néanmoins considérer les deux types de projection (orthographique et perspective). Commençons donc par établir l'expression de la normale $\mathbf{n}(u,v)$ sous l'hypothèse de la projection perspective.

Nous suivons dans ce paragraphe le même cheminement que dans le paragraphe 2.2.1.2. Rappelons tout d'abord que, sous l'hypothèse de la projection perspective, d'après (1.12) :

$$\mathbf{X}(u,v) = z(u,v)\left[\frac{u-u_0}{f}, \frac{v-v_0}{f}, 1\right]^\top \tag{2.26}$$

De cette expression, nous déduisons le produit vectoriel suivant :

$$\partial_u \mathbf{X}(u,v) \wedge \partial_v \mathbf{X}(u,v) \propto \begin{bmatrix} -f\, z_u(u,v) \\ -f\, z_v(u,v) \\ z(u,v) + (u-u_0)\, z_u(u,v) + (v-v_0)\, z_v(u,v) \end{bmatrix} \tag{2.27}$$

Sauf cas exceptionnel, la normale en un point visible de la surface observée doit pointer vers la caméra. Il nous faut donc modifier le signe du produit vectoriel (2.27), ce qui nous permet d'écrire, en recourant au changement de variable $\tilde{z} = \log(z)$:

$$\mathbf{n}(u,v) = \frac{\begin{bmatrix} f\,\nabla\tilde{z}(u,v) \\ -1 - [u-u_0, v-v_0]^\top \nabla\tilde{z}(u,v) \end{bmatrix}}{\sqrt{f^2\,\|\nabla\tilde{z}(u,v)\|^2 + \left\{1 + [u-u_0, v-v_0]^\top \nabla\tilde{z}(u,v)\right\}^2}} \tag{2.28}$$

Afin de simplifier les équations à venir, nous regroupons les expressions (2.19) et (2.28) sous une forme générique :

$$\begin{aligned} \mathbf{n} : \quad \Omega \quad &\to \mathbb{S}^2 \subset \mathbb{R}^3 \\ (u,v) \quad &\mapsto \mathbf{n}(u,v) = \frac{1}{d_{\nabla z}(u,v)} \begin{bmatrix} f\,\nabla z(u,v) \\ -1 - [\tilde{u},\tilde{v}]^\top \nabla z(u,v) \end{bmatrix} \end{aligned} \tag{2.29}$$

où :

$$(z, f, \tilde{u}, \tilde{v}) = \begin{cases} (z, 1, 0, 0) & \text{pour la projection orthographique} \\ (\tilde{z}, f, u-u_0, v-v_0) & \text{pour la projection perspective} \end{cases} \tag{2.30}$$

et où la fonction $d_{\nabla z}$ permet de garantir la contrainte $\|\mathbf{n}(u,v)\| = 1$:

$$\begin{aligned} d_{\nabla z} : \quad \Omega \quad &\to \mathbb{R} \\ (u,v) \quad &\mapsto d_{\nabla z}(u,v) = \sqrt{f^2\,\|\nabla z(u,v)\|^2 + \left\{1 + [\tilde{u},\tilde{v}]^\top \nabla z(u,v)\right\}^2} \end{aligned} \tag{2.31}$$

2.3.2.2 Réflectance non lambertienne

Sous l'hypothèse lambertienne, la réflectance de la surface est entièrement caractérisée par l'albédo. Celui-ci est supposé connu afin de réduire les ambiguïtés du problème [3]. Ces hypothèses sont très fortes. D'une part, une grande majorité de surfaces présentent des écarts au modèle (2.12), de par leur réflectance. D'autre part, même dans le cas d'une surface parfaitement lambertienne, la connaissance de l'albédo constitue en soi une réelle difficulté.

De nombreux travaux sur le SfS se sont intéressés à des modèles de réflectance plus complexes [4, 12, 165], particulièrement au modèle d'Oren-Nayar [84, 85] et au modèle de Phong [168, 169]. Mais dans les deux cas, cela se fait au prix de difficultés numériques non négligeables. Nous préférons profiter de la simplicité du modèle lambertien et considérer que les écarts à ce modèle, tels que les reflets brillants, constituent des données aberrantes. Dans un premier temps, nous continuons de supposer l'albédo connu, une hypothèse sur laquelle nous reviendrons dans le chapitre 4.

2.3.2.3 Éclairage non parallèle

La méthode de résolution du SfS que nous allons décrire a pour vocation d'être appliquée à des scènes 3D réelles, photographiées en « conditions d'éclairage naturel ». Il semble donc impératif de dépasser le modèle d'éclairage parallèle et uniforme, dont la direction et la densité de flux sont identiques en tout point éclairé de la surface.

Des modèles d'éclairage plus réalistes ont été étudiés. L'hypothèse d'une source lumineuse ponctuelle proche a permis une généralisation intéressante des méthodes de résolution du SfS [26, 79, 135], mais cette extension reste tout de même très éloignée d'un éclairage naturel. Certains travaux ont montré l'intérêt qu'il y a à supposer l'éclairage purement diffus [95, 155], auquel cas les photographies en extérieur sont contraintes à être prises par un temps nuageux.

Nous utilisons dans ce mémoire une modélisation de l'éclairage par les **harmoniques sphériques**. Le modèle (2.12) ne tient compte que d'une seule direction d'éclairage $\mathbf{s}(u, v)$ en chaque point de la surface. Un éclairage naturel peut être vu comme un ensemble de sources ponctuelles situées à l'infini, c'est-à-dire comme un ensemble d'éclairages parallèles et uniformes caractérisés par autant de directions $\mathbf{u}_i$. Reprenons le raisonnement ayant abouti au modèle (2.12), en sommant les contributions de ces différentes sources lumineuses. Notons $\mathcal{H}(u, v)$ l'hémisphère « extérieure » reposant sur le plan localement tangent à la surface en $\mathbf{X}(u, v)$. La totalité de l'énergie lumineuse atteignant $\mathbf{X}(u, v)$ traverse cette hémisphère. En utilisant l'expression (2.2) de $\mathrm{d}E$, nous obtenons l'expression suivante pour l'éclairement en $\mathbf{X}(u, v)$:

$$E(u, v) = \int_{\mathcal{H}(u,v)} \bar{L}(\mathbf{u}_i)\, \mathbf{u}_i^\top \mathbf{n}(u, v)\, \mathrm{d}\Omega_i \tag{2.32}$$

3. Généralement, l'albédo est même supposé uniformément égal à 1.

De (2.5) et (2.32), nous déduisons la luminance émise par une surface lambertienne :

$$L(u,v) = \frac{\rho(u,v)}{\pi} \int_{\mathcal{H}(u,v)} \bar{L}(\mathbf{u}_i)\, \mathbf{u}_i^\top \mathbf{n}(u,v)\, \mathrm{d}\Omega_i \tag{2.33}$$

De (2.9), (2.10) et (2.33), nous tirons enfin :

$$I(u,v) \propto \rho(u,v) \int_{\mathcal{H}(u,v)} \bar{L}(\mathbf{u}_i)\, \mathbf{u}_i^\top \mathbf{n}(u,v)\, \mathrm{d}\Omega_i \tag{2.34}$$

Dans (2.34), nous pouvons étendre le domaine d'intégration à la sphère unité $\mathbb{S}^2$, pour peu que seule l'hémisphère $\mathcal{H}(u,v)$ soit prise en compte, c'est-à-dire en tenant compte uniquement des directions $\mathbf{u}_i$ telles que $\mathbf{u}_i^\top \mathbf{n}(u,v) \geq 0$:

$$I(u,v) \propto \rho(u,v) \int_{\mathbb{S}^2} \bar{L}(\mathbf{u}_i)\, \max\{\mathbf{u}_i^\top \mathbf{n}(u,v), 0\}\, \mathrm{d}\Omega_i \tag{2.35}$$

Or, grâce au théorème de Funk-Hecke, l'intégrale dans (2.35) peut être approchée par des harmoniques sphériques [15, 142], et ce de la manière suivante :

$$\int_{\mathbb{S}^2} \bar{L}(\mathbf{u}_i)\, \max\{\mathbf{u}_i^\top \mathbf{n}(u,v), 0\}\, \mathrm{d}\Omega_i \approx \boldsymbol{\sigma}^\top \boldsymbol{\nu}(u,v) \tag{2.36}$$

où $\boldsymbol{\sigma} \in \mathbb{R}^9$ constitue une représentation compacte de l'éclairage, indépendante de (u,v), et où $\boldsymbol{\nu}(u,v) \in \mathbb{R}^9$ caractérise l'information géométrique locale. Ce vecteur, qui constitue une généralisation de la normale $\mathbf{n}(u,v) = [n_1(u,v), n_2(u,v), n_3(u,v)]^\top$, est défini comme suit :

$$\boldsymbol{\nu}(u,v) = \begin{bmatrix} 1 \\ \mathbf{n}(u,v) \\ n_1(u,v)\, n_2(u,v) \\ n_1(u,v)\, n_3(u,v) \\ n_2(u,v)\, n_3(u,v) \\ n_1(u,v)^2 - n_2(u,v)^2 \\ 3\, n_3(u,v)^2 - 1 \end{bmatrix} \tag{2.37}$$

Au même titre que la base de Fourier permet de décomposer fréquentiellement un signal, les harmoniques sphériques offrent une décomposition des fonctions définies sur $\mathbb{S}^2$ adaptée à notre cas, puisque la réflectance lambertienne agit comme un filtre passe-bas. L'éclairage peut donc être approché par quelques coefficients sur la base des harmoniques sphériques.

En réalité, la taille des vecteurs $\boldsymbol{\sigma}$ et $\boldsymbol{\nu}(u,v)$ est modulable. Plus le nombre de coefficients pris en compte est grand, plus la description de l'éclairage est précise. Le premier coefficient (approximation « à l'ordre zéro ») ne permet de modéliser que la composante « diffuse » de l'éclairage. Les trois coefficients suivants correspondent à l'éclairage parallèle et uniforme utilisé jusqu'ici. Les cinq coefficients suivants correspondent à des éclairages plus complexes [142]. Il a été montré que l'approximation « au premier ordre », qui utilise les quatre premiers coefficients, permet de représenter $87,5\%$ de l'énergie lumineuse, tandis que l'approximation « au deuxième ordre », qui utilise neuf coefficients, permet de représenter jusqu'à $99,2\%$ de cette énergie [46].

Les équations (2.35) et (2.36) nous permettent d'établir le modèle suivant, qui constitue une extension de la loi de Lambert (2.12) [4] :

$$I(u,v) = \rho(u,v)\,\boldsymbol{\sigma}^\top \boldsymbol{\nu}(u,v) \tag{2.38}$$

Il est notable que les cinq dernières composantes de l'expression (2.37) de $\boldsymbol{\nu}(u,v)$ sont non linéaires vis-à-vis des composantes de la normale, mais ces non linéarités ne créent pas de difficulté nouvelle. En effet, nous avons déjà vu que la linéarité du modèle (2.12), qui est effective vis-à-vis des composantes de la normale, est perdue vis-à-vis de la fonction de profondeur z, à cause de la contrainte $\|\mathbf{n}\| = 1$. En pratique, le modèle (2.38) ne se révèle donc pas beaucoup plus difficile à manipuler que (2.12). Il permet de modéliser l'éclairage d'une façon bien plus réaliste, au prix d'une non linéarité légèrement accrue.

Des harmoniques sphériques du premier ordre ont été utilisées dans différents travaux [78, 127], ainsi que des harmoniques sphériques du deuxième ordre [83, 145]. Ces dernières ont également servi [13] à résoudre le problème complexe de la décomposition intrinsèque de l'image à partir de l'ombrage (baptisée SIRFS, pour *Shape, Illumination and Reflectance From Shading*) [5].

Nous désignons par σ_i la $i^{\text{ème}}$ composante du vecteur d'éclairage $\boldsymbol{\sigma}$. D'après la définition (2.37) du vecteur $\boldsymbol{\nu}(u,v)$, un éclairage parallèle et uniforme correspond à un vecteur $\boldsymbol{\sigma}$ tel que $\sigma_i = 0$, pour $i \in \{1,5,6,7,8,9\}$. Comme nous l'avons déjà fait pour l'albédo, nous supposons dans ce chapitre l'éclairage connu, son estimation étant reportée au chapitre 4.

2.3.2.4 Utilisation d'un relief préalablement estimé

L'ambiguïté concave/convexe restant un problème fondamental du SfS, il semble pertinent d'utiliser davantage d'information pour guider la résolution, par exemple en forçant la solution à « ressembler » à un relief préalablement estimé, qui peut être fourni par un capteur externe (caméra de profondeur de type Kinect) [59, 61, 62, 127, 174, 177]. Cette procédure permet également de fusionner les informations photométriques contenues dans les images avec les données géométriques obtenues par d'autres techniques de reconstruction 3D [77, 98, 179], comme celles que nous avons présentées dans le chapitre 1.

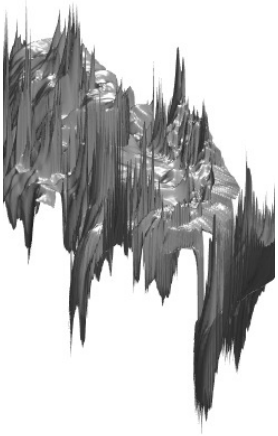
2.3.3 Modèle différentiel générique

Les travaux évoqués dans le paragraphe précédent [13, 78, 127] ne sont malheureusement pas totalement satisfaisants. En effet, les méthodes conçues pour lever les ambiguïtés du problème et tenir compte du bruit se révèlent souvent motivées par des considérations de stabilité numérique. Par exemple, les résultats obtenus par la méthode SIRFS [13] ne sont plus du tout acceptables sans la stratégie multi-échelles. La méthode proposée dans [127] devient, quant à elle, instable sans le terme de régularisation. Ces affirmations sont confirmées par les résultats de la figure 2.4.

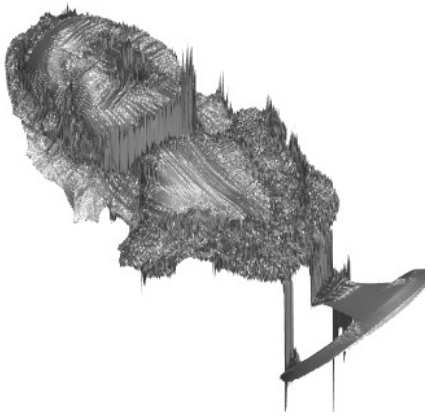
4. Comme pour (2.12), l'égalité (2.38) est en réalité une relation de proportionnalité.

5. La méthode SIRFS peut être considérée comme une méthode de résolution du SfS, si l'albédo et l'éclairage sont connus.

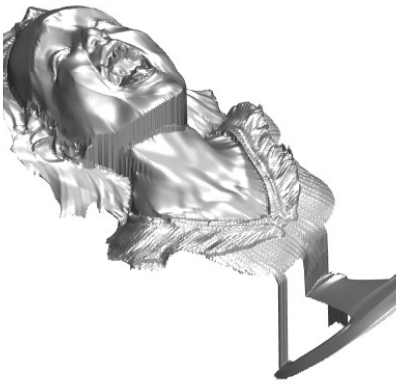
Données : image de synthèse et éclairage.



Algorithme du point fixe (sans régularisation) [127].



SIRFS (sans multi-échelles) [13].



Notre algorithme (sans régularisation).

FIGURE 2.4 – Résultats d'algorithmes de SfS utilisant les harmoniques sphériques du premier ordre. Les stratégies de linéarisation comme le point fixe [127] ne sont pas concluantes sans terme de régularisation. Des problèmes similaires apparaissent avec la méthode SIRFS [13] quand l'approche multi-échelles est désactivée. Notre méthode de résolution du SfS n'utilise de régularisation que pour lever les ambiguïtés du problème et améliorer la robustesse au bruit, et non pour garantir la stabilité de l'algorithme. Nous utilisons la même surface initiale (l'initialisation « réaliste » de la figure 2.7) pour ces trois tests.

Bien que les mécanismes de régularisation contournent quelque peu ces instabilités numériques dans la pratique, une méthode de résolution idéale ne devrait s'appuyer sur la régularisation que pour lever les ambiguïtés du problème et tenir compte du bruit, et non pour garantir une quelconque stabilité numérique.

Chacune des deux catégories de méthodes de résolution du SfS du paragraphe 2.2.2 possède des avantages et des inconvénients. Nous l'avons vu, les méthodes fondées sur la résolution d'EDP ne nécessitent aucun réglage de paramètre, mais elles manquent de robustesse. A contrario, l'approche variationnelle permet de garantir la régularité de la solution grâce à des connaissances a priori, mais elle nécessite de régler des paramètres. Nous allons montrer comment combiner les avantages de ces deux approches, grâce à une nouvelle méthode de résolution variationnelle des EDP du SfS.

L'utilisation des EDP présente l'avantage d'imposer implicitement la dérivabilité (presque partout) sans nécessiter l'ajout d'un terme de régularisation. Nous introduirons plus tard des termes de régularisation, mais uniquement « pour de bonnes raisons », à savoir pour lever les ambiguïtés du SfS et tenir compte du bruit.

En injectant la forme générique (2.29) de la normale dans la loi de Lambert (2.38), le modèle de formation de l'image se réécrit sous la forme de l'EDP en z suivante :

$$\mathbf{a}_{\nabla z}^{\top} \nabla z + b_{\nabla z} = I \tag{2.39}$$

où $\mathbf{a}_{\nabla z}$ et $b_{\nabla z}$ désignent, respectivement, un champ vectoriel et un champ scalaire dépendant non linéairement de ∇z, qui sont définis par :

$$\begin{aligned} \mathbf{a}_{\nabla z} : \quad \Omega \quad &\to \mathbb{R}^2 \\ (u,v) \quad &\mapsto \mathbf{a}_{\nabla z}(u,v) = \frac{\rho(u,v)}{d_{\nabla z}(u,v)} \begin{bmatrix} f\sigma_2 - \widetilde{u}\,\sigma_4 \\ f\sigma_3 - \widetilde{v}\,\sigma_4 \end{bmatrix} \end{aligned} \tag{2.40}$$

$$\begin{aligned} b_{\nabla z} : \quad \Omega \quad &\to \mathbb{R} \\ (u,v) \quad &\mapsto b_{\nabla z}(u,v) = \rho(u,v) \begin{bmatrix} \sigma_1 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \\ \sigma_7 \\ \sigma_8 \\ \sigma_9 \end{bmatrix}^{\top} \begin{bmatrix} 1 \\ -\frac{1}{d_{\nabla z}(u,v)} \\ \frac{f^2\, z_u(u,v)\, z_v(u,v)}{d_{\nabla z}(u,v)^2} \\ \frac{f\, z_u(u,v)\left(-1-[\tilde{u},\tilde{v}]^{\top}\nabla z(u,v)\right)}{d_{\nabla z}(u,v)^2} \\ \frac{f\, z_v(u,v)\left(-1-[\tilde{u},\tilde{v}]^{\top}\nabla z(u,v)\right)}{d_{\nabla z}(u,v)^2} \\ \frac{f^2\left(z_u(u,v)^2 - z_v(u,v)^2\right)}{d_{\nabla z}(u,v)^2} \\ \frac{3\left(-1-[\tilde{u},\tilde{v}]^{\top}\nabla z(u,v)\right)^2}{d_{\nabla z}(u,v)^2} - 1 \end{bmatrix} \end{aligned} \tag{2.41}$$

Rappelons que $d_{\nabla z}(u,v)$ et $(\tilde{u},\tilde{v})$ sont définis en (2.31) et (2.30), respectivement.

Dans le cas d'une caméra orthographique, si l'éclairage est parallèle, uniforme et frontal, donc si $\mathbf{s} = [0,0,s_3]^{\top}$, et si l'albédo vaut $\rho(u,v) = 1$, hypothèses classiques en SfS, alors le modèle (2.39) se ramène à l'équation eikonale (2.22). De plus, ce modèle est générique. En effet, il est valide pour la projection orthographique et la projection perspective, pour un éclairage directionnel ou un éclairage plus complexe caractérisé par des harmoniques sphériques du deuxième ordre.

Nous présentons à présent une solution variationnelle à ce modèle générique. Si nous supposons que le modèle de formation de l'image utilisé est satisfait à un bruit additif blanc gaussien près, alors la solution du maximum de vraisemblance est obtenue en estimant la

fonction de profondeur z qui minimise la fonction de coût en moindres carrés suivante :

$$\mathcal{E}(\nabla z; I) = \left\| \mathbf{a}_{\nabla z}^{\top} \nabla z + b_{\nabla z} - I \right\|_{\ell^2(\Omega)}^2 \tag{2.42}$$

Dans les travaux récents d'affinage de la géométrie à partir de l'ombrage [127], il est suggéré de minimiser une fonction de coût similaire à (2.42) de manière itérative, en « gelant » les champs non linéaires $\mathbf{a}_{\nabla z}$ et $b_{\nabla z}$ à chaque itération. Cette stratégie doit être évitée, pour la simple et bonne raison qu'elle ne parvient pas à résoudre le problème dans le cas élémentaire où la projection est orthographique et où l'éclairage est parallèle, uniforme et frontal : l'expression (2.40) devient alors $\mathbf{a}_{\nabla z} \equiv \mathbf{0}$, puisque $\widetilde{u} = \widetilde{v} = 0$ (projection orthographique) et $\sigma_2 = \sigma_3 = 0$ (éclairage parallèle, uniforme et frontal). Par conséquent, l'expression (2.42) ne dépend même plus de la fonction de profondeur z, une fois $b_{\nabla z}$ gelé. Comme cela est illustré sur la figure 2.4, cette stratégie reste instable dans des cas moins élémentaires, ce qui explique le recours à la régularisation dans [127].

Nous aurons également recours à la régularisation, mais uniquement pour lever les ambiguïtés du SfS et tenir compte du bruit. La méthode de résolution que nous proposons reste stable en l'absence de régularisation (cf. figure 2.4). Décrivons maintenant cette méthode, en discutant sa parenté avec l'approche variationnelle de résolution du SfS proposée par Horn et Brooks [75].

2.3.4 La méthode de Horn et Brooks revisitée

Dans [75], Horn et Brooks introduisent une approche variationnelle pour résoudre l'équation eikonale du SfS qui est, nous l'avons vu, un cas particulier du modèle (2.39). Le modèle de SfS utilisé dans [75] est non différentiel. La méthode de résolution proposée fonctionne donc en deux étapes : estimation du gradient de profondeur ∇z, puis intégration de ce champ vectoriel pour obtenir la fonction de profondeur z. L'équation (2.39) est d'abord minimisée relativement au gradient $\theta := \nabla z$, puis θ est intégré en une fonction de profondeur z. Or, l'estimation locale du gradient est ambiguë. Pour cette raison, la **contrainte d'intégrabilité** du gradient est introduite à la première étape de la résolution.

Le champ vectoriel θ dérive du champ scalaire z. Dans le cas où θ est intégrable, la valeur de z estimée doit être indépendante du chemin d'intégration, c'est-à-dire que θ doit être un champ vectoriel conservatif. Or, la circulation d'un champ vectoriel entre deux points ne dépend pas du chemin d'intégration si le rotationnel de ce champ est nul. En notant $\nabla\cdot$ l'opérateur divergence, nous pouvons donc écrire la contrainte d'intégrabilité :

$$\underbrace{\left\| \nabla \cdot \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \theta \right\|_{\ell^2(\Omega)}^2}_{:=\mathcal{I}(\theta)} = 0 \tag{2.43}$$

En pratique, la contrainte « dure » (2.43) est transformée en un terme de régularisation. Ceci nécessite la mise en place d'une pondération entre les termes d'attache aux données et d'intégrabilité du champ estimé :

$$\widehat{\theta} = \underset{\theta\colon \Omega \to \mathbb{R}^2}{\operatorname{argmin}} \; \lambda\, \mathcal{E}(\theta; I) + \mu\, \mathcal{I}(\theta) \tag{2.44}$$

Une fois menée l'estimation (2.44), le champ $\widehat{\theta}$ est intégré en une fonction de profondeur z, par résolution de l'équation $\nabla z = \widehat{\theta}$. Cependant, la contrainte d'intégrabilité ayant été « réduite » à un terme de régularisation, l'intégrabilité de $\widehat{\theta}$ n'est pas garantie. Cela explique pourquoi Horn et Brooks reformulent le problème d'intégration sous la forme d'un nouveau problème variationnel :

$$\min_{z\colon \Omega \to \mathbb{R}} \left\| \nabla z - \widehat{\theta} \right\|^2_{\ell^2(\Omega)} \tag{2.45}$$

Cette approche en deux étapes, qui consiste à résoudre d'abord (2.44), puis (2.45), est malgré tout sujette à la propagation d'erreurs, c'est-à-dire qu'une petite erreur lors de l'estimation du gradient peut avoir des conséquences très importantes lors de l'étape d'intégration. Nous suggérons donc d'éviter cette approche séquentielle, d'autant plus qu'elle n'est pas nécessaire. En effet, θ étant un champ conservatif par essence, son intégrabilité ne devrait plus être invoquée.

Nous proposons une nouvelle approche permettant d'estimer le relief de la surface à partir de l'image, en utilisant l'estimation locale du gradient, comme dans la méthode de Horn et Brooks, mais qui contraigne explicitement le gradient à être conservatif. Autrement dit, nous estimons simultanément la fonction de profondeur et son gradient, en transformant la minimisation de la fonction de coût (2.42) en un problème contraint :

$$\begin{aligned} \min_{\substack{\theta\colon \Omega \to \mathbb{R}^2 \\ z\colon \Omega \to \mathbb{R}}} \ & \mathcal{E}(\theta; I) \\ \text{s.c. } & \theta = \nabla z \end{aligned} \tag{2.46}$$

Le problème variationnel (2.46) peut être résolu par une méthode de type lagrangien augmenté. En comparaison de la méthode de Horn et Brooks, cela nous permet non seulement de ne pas introduire d'hyper-paramètre comme dans (2.44), mais également d'éviter les erreurs de propagation dues à une résolution en deux étapes. De plus, cette approche peut être facilement étendue avec, au besoin, l'ajout de termes de régularisation, dont nous allons présenter deux exemples : le premier permet de réduire les ambiguïtés du SfS grâce à un a priori sur la profondeur ; le second vise à effectuer un lissage de la surface.

2.3.5 Termes de régularisation

2.3.5.1 A priori sur la profondeur

Il existe plusieurs moyens simples pour obtenir un a priori sur la profondeur, par exemple en utilisant une caméra de profondeur ou le pipeline SfM/MVS décrit dans le chapitre 1. Une fonction de profondeur z^0, même bruitée ou incomplète, peut être obtenue par ce biais et se révéler un bon « guide » pour le SfS. Nous proposons le terme de régularisation suivant :

$$\mathcal{P}(z; z^0) = \left\| z - z^0 \right\|^2_{\ell^2(\Omega^0)} \tag{2.47}$$

où $\Omega^0 \subseteq \Omega \subset \mathbb{R}^2$ est la région de l'image pour laquelle l'a priori est disponible.

2.3.5.2 Lissage de la surface

Afin de ne pas interpréter le bruit de l'image par de petites oscillations du relief, il est souhaitable d'améliorer la robustesse de notre modèle en lui ajoutant explicitement un terme de lissage. La régularisation par **variation totale**, qui est fréquemment utilisée en vision par ordinateur [170], pourrait avoir tendance à favoriser les surfaces fronto-parallèles et provoquer ainsi un « effet d'escalier ». Nous préférons mesurer l'aire totale de la surface. Il a été montré récemment [57] que cette approche était plus adaptée à la régularisation de fonctions de profondeur. En termes de géométrie différentielle, la fonction $d_{\nabla z}$ définie en (2.31) est égale à la racine carrée du déterminant de la première forme fondamentale de la fonction z (tenseur métrique). Son intégrale sur Ω est proportionnelle à l'aire de la surface. Nous proposons donc le terme de lissage de la surface suivant :

$$\mathcal{A}(\nabla z) = \|d_{\nabla z}\|_{\ell^1(\Omega)} \tag{2.48}$$

2.3.5.3 Modèle complet

En combinant (2.42), (2.47) et (2.48), et en utilisant le même changement de variable $\theta := \nabla z$ que dans (2.46), nous obtenons l'approche variationnelle régularisée suivante :

$$\begin{aligned} &\min_{\substack{\theta:\,\Omega\to\mathbb{R}^2 \\ z:\,\Omega\to\mathbb{R}}} \lambda\,\mathcal{E}(\theta; I) + \mu\,\mathcal{P}(z; z^0) + \nu\,\mathcal{A}(\theta) \\ &\text{s.c. } \theta = \nabla z \end{aligned} \tag{2.49}$$

qui est une version régularisée du modèle (2.46), où $(\lambda, \mu, \nu) \geq (0, 0, 0)$ sont des hyper-paramètres qui contrôlent l'influence de chaque terme.

Il est notable que le modèle variationnel (2.49) peut être utilisé dans différents contextes. Il apporte une solution à un problème de pur SfS si $\mu = \nu = 0$. Si $\lambda = 0$ et $\Omega^0 = \Omega$, il permet de débruiter une fonction de profondeur, de manière similaire à la méthode présentée dans [57]. Enfin, il permet d'affiner et de compléter une fonction de profondeur par l'ombrage si $\lambda > 0$, $\mu > 0$ et $\Omega^0 \subsetneq \Omega$.

2.3.6 Optimisation

Le problème d'optimisation (2.49) est à la fois non linéaire, à cause du terme d'attache aux données (2.42) et du terme de lissage de la surface (2.48), et non local, puisqu'il dépend du gradient de profondeur. Le changement de variable $\theta := \nabla z$ dans (2.49) est essentiel pour la résolution numérique, car il sépare les deux difficultés. L'optimisation peut être effectuée en alternant l'estimation du gradient, qui est non linéaire mais locale, et l'estimation de la profondeur, qui est globale mais linéaire. Nous optons pour une méthode de type **lagrangien augmenté**.

En introduisant une variable auxiliaire Ψ, le problème (2.49) peut être résolu itérativement par un **schéma à direction de descente alternée** (noté ADMM, pour *Alternating Direction Method of Multipliers*). Cette méthode classique d'optimisation sous contrainte a été proposée dans les années 70 [54]. Une description plus complète en est donnée dans [23].

Le lagrangien augmenté associé à (2.49) est défini de la manière suivante :

$$\mathcal{L}_\alpha(\theta, z, \Psi) = \lambda\,\mathcal{E}(\theta; I) + \mu\,\mathcal{P}(z; z^0) + \nu\,\mathcal{A}(\theta) + \langle \Psi, \theta - \nabla z \rangle + \frac{\alpha}{2}\,\|\theta - \nabla z\|_2^2 \tag{2.50}$$

où $\Psi : \Omega \to \mathbb{R}^2$ est la fonction vectorielle contenant les multiplicateurs de Lagrange, $\langle \cdot \rangle$ désigne le produit scalaire induit par $\|\cdot\|_2$ sur Ω, et $\alpha > 0$.

Les itérations du schéma ADMM s'écrivent :

$$\theta^{(k+1)} = \underset{\theta\colon \Omega \to \mathbb{R}^2}{\operatorname{argmin}}\ \mathcal{L}_{\alpha^{(k)}}(\theta, z^{(k)}, \Psi^{(k)}) \tag{2.51}$$

$$z^{(k+1)} = \underset{z\colon \Omega \to \mathbb{R}}{\operatorname{argmin}}\ \mathcal{L}_{\alpha^{(k)}}(\theta^{(k+1)}, z, \Psi^{(k)}) \tag{2.52}$$

$$\Psi^{(k+1)} = \Psi^{(k)} + \alpha^{(k)}\left(\theta^{(k+1)} - \nabla z^{(k+1)}\right) \tag{2.53}$$

où $\alpha^{(k)}$ est déterminé de manière automatique [23].

Alors que le problème (2.49) est, rappelons-le, non linéaire et non local, chacun des sous-problèmes issus du schéma ADMM est plus simple à résoudre.

En effet, le problème (2.51) est non linéaire, mais il s'agit d'un problème local qui peut être résolu indépendamment en chaque pixel. Nous pouvons, par exemple, utiliser une méthode de type « quasi-Newton » telle que BFGS [104, 149].

Le problème (2.52), quant à lui, est discrétisé par des différences finies avant d'ordre un, auxquelles nous ajoutons une condition au bord de type Neumann. Le problème ainsi discrétisé prend la forme d'un problème en moindres carrés linéaires, dont les équations normales forment un système linéaire symétrique, défini positif (semi-défini positif, dans le cas où $\mu = 0$). Bien que ce système soit creux, il reste trop grand pour pouvoir être résolu directement. Une résolution par gradient conjugué est préférable.

Ces différents sous-problèmes offrent également une interprétation de plus « haut niveau » de notre algorithme. À l'étape (2.51) de mise à jour de θ, le gradient de z (c'est-à-dire le champ de normales) est estimé localement par l'ombrage, tout en garantissant que ce gradient soit lisse et proche de celui de la fonction de profondeur courante. À l'inverse de l'approche par point fixe [127], l'orientation locale de la surface est estimée à partir du modèle complet (2.39), et non pas uniquement à partir de sa partie linéaire. Nous observons, en pratique, que notre algorithme devient ainsi beaucoup plus stable (cf. figure 2.4). Les mises à jour (2.52) de z reviennent à intégrer les nouvelles normales en une nouvelle fonction de profondeur qui reste proche de l'a priori.

La non convexité des termes d'ombrage $\mathcal{E}(\theta; I)$ et de lissage $\mathcal{A}(\theta)$ ne nous permettent pas de garantir la convergence de cet algorithme. Nous n'avons cependant observé aucun cas de divergence. Nous conjecturons qu'une preuve de convergence pourrait être apportée grâce aux dernières études sur le schéma ADMM [72, 102].

Le prochain paragraphe montre l'efficacité de l'algorithme proposé pour la résolution du SfS en conditions d'éclairage naturel.

2.4 Résultats expérimentaux

Nous commençons par valider la capacité de notre méthode à résoudre le SfS en conditions d'éclairage naturel, c'est-à-dire à résoudre l'équation (2.38). Notre approche est comparée à la méthode SIRFS [13], seule méthode de résolution du SfS en conditions d'éclairage naturel dont le code soit disponible. Afin de comparer des choses comparables, les estimations de l'albédo et de l'éclairage par la méthode SIRFS sont désactivées, et la stratégie multi-échelles est conservée afin d'éviter les artéfacts mis en évidence sur la figure 2.4.

2.4.1 Évaluation quantitative sur des données de synthèse

Puisque, dans un premier temps, nous voulons seulement comparer la capacité de ces deux méthodes à expliquer l'image, les termes de régularisation de notre méthode sont désactivés ($\mu = \nu = 0$), tout comme ceux de SIRFS. Afin de quantifier les résultats, nous mesurons la racine carré de l'erreur quadratique moyenne (RMSE, pour *Root Mean Square Error*) entre les images d'entrée et les images reprojetées, comme cela est conseillé dans [43].

Les jeux de données ont été créés à partir du modèle 3D « Joyful Yell » par projection orthographique, afin d'effectuer une comparaison équitable, SIRFS n'étant pas conçu pour tenir compte de la projection perspective. Des images en niveaux de gris non bruitées sont simulées avec deux scénarios d'éclairage, décrits par un vecteur d'éclairage $\boldsymbol{\sigma}_1$ du premier ordre et par un vecteur d'éclairage $\boldsymbol{\sigma}_2$ du deuxième ordre définis comme suit :

$$\boldsymbol{\sigma}_1 = [\ 0,1 \quad -0,25 \quad -0,7 \quad 0,2 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0\ ]^\top \tag{2.54}$$

$$\boldsymbol{\sigma}_2 = [\ 0,2 \quad 0,3 \quad -0,7 \quad 0,5 \quad -0,2 \quad -0,2 \quad 0,3 \quad 0,3 \quad 0,2\ ]^\top \tag{2.55}$$

Deux surfaces sont utilisées comme relief initial : l'une est très éloignée de la vérité terrain (fonction « peaks » de MATLAB), tandis que l'autre en est relativement proche (obtenue par un simple filtrage gaussien appliqué à la vérité terrain). La vérité terrain, ainsi que ces deux surfaces initiales, sont représentées sur la figure 2.5.

FIGURE 2.5 – De gauche à droite : vérité terrain ; surface initiale « non réaliste » (fonction « peaks » de MATLAB) ; surface initiale « réaliste » obtenue en dégradant la vérité terrain.

Bien que μ soit nul pour les tests des figures 2.6 et 2.7, la méthode que nous proposons ne s'éloigne pas trop de la vérité terrain : la surface obtenue est qualitativement satisfaisante, dès lors qu'une initialisation adéquate est utilisée. Il n'est pas étonnant de constater que l'initialisation joue un rôle prépondérant, à cause des ambiguïtés inhérentes au SfS.

Dans tous les tests, les images sont mieux expliquées par notre méthode, ce qui montre que la stratégie numérique proposée résout le problème fortement non linéaire du SfS en conditions d'éclairage naturel de manière plus précise que la méthode SIRFS. De plus, les temps de calcul des deux méthodes sont comparables : quelques minutes pour chaque test, avec une implémentation en MATLAB sur un ordinateur portable standard, pour des ensembles Ω comportant environ 150000 pixels.

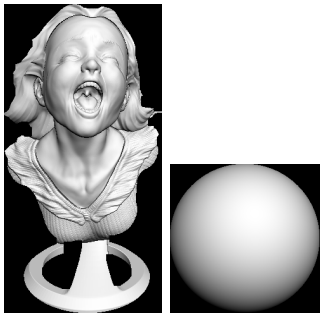
2.4.2 Évaluation qualitative sur des données réelles

L'importance de l'initialisation se confirme sur la figure 2.8. Dans ces tests, notre méthode de SfS ($\mu = \nu = 0$) est évaluée sur des données réelles obtenues par un capteur RGB-D [61], en utilisant un plan fronto-parallèle comme surface initiale. Les images sont préalablement transformées en niveaux de gris. Bien que les détails des surfaces soient retrouvés, les résultats présentent un biais évident dans les basses fréquences, c'est-à-dire que la forme générale de l'objet n'est pas satisfaisante. De plus, de nombreux artéfacts dus au bruit apparaissent. Ceci illustre les ambiguïtés inhérentes au SfS et la nécessité de recourir à un terme de régularisation de la fonction de profondeur.

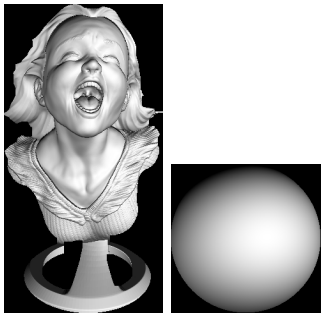
2.4.3 Influence des hyper-paramètres

Afin d'illustrer l'influence des hyper-paramètres μ et ν qui contrôlent, respectivement, le poids de l'a priori sur la profondeur et le poids du lissage de la surface, nous utilisons le jeu de données de la figure 2.7, mais nous ajoutons un bruit blanc à la fois à l'image et à la fonction de profondeur initiale (nous utilisons, pour ces tests, l'initialisation « réaliste » de la figure 2.5). Ces données sont représentées sur la figure 2.9.

Les résultats obtenus avec divers jeux de paramètres sont regroupés sur la figure 2.10. Choisir $\lambda = 1$ et $(\mu, \nu) = (0, 0)$ revient à résoudre un problème de SfS : les détails du relief sont parfaitement retrouvés, mais la surface peut s'éloigner de l'initialisation, car le bruit est attribué à tort au relief, ce qui provoque l'apparition d'artéfacts. Si $\mu \to +\infty$, l'estimation initiale (dont la forme globale est proche de la vérité terrain, sans les détails) n'est pas modifiée. Si $\nu \to +\infty$, seul le terme de lissage de la surface est pris en compte. Le résultat est alors exagérément lissé. Dans ces tests, nous évaluons également l'erreur angulaire moyenne sur les normales (MAE, pour *Mean Angular Error*), exprimée en degrés, qui est minimale lorsque les paramètres sont réglés de manière appropriée, et non lorsque la RMSE est minimale. Ceci s'explique par le fait que la minimisation de la RMSE revient à expliquer le bruit de l'image par des oscillations rapides du relief.

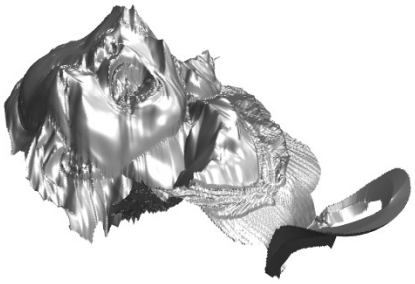


Éclairage $\boldsymbol{\sigma}_1$ d'ordre 1 (2.54).

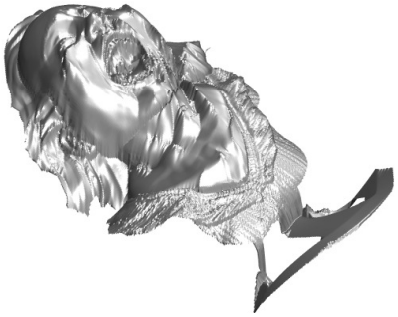


Éclairage $\boldsymbol{\sigma}_2$ d'ordre 2 (2.55).

Nos résultats

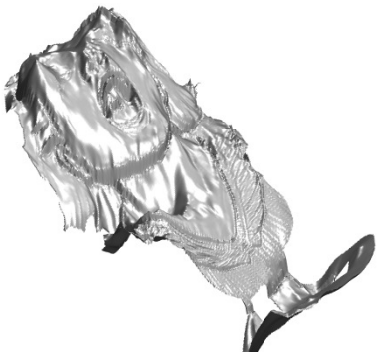


RMSE = **0,02**

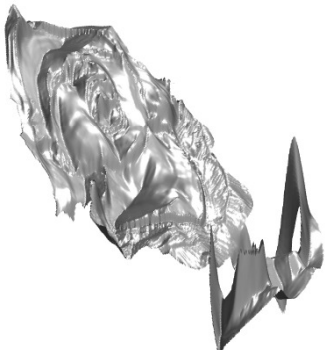


RMSE = **0,03**

SIRFS

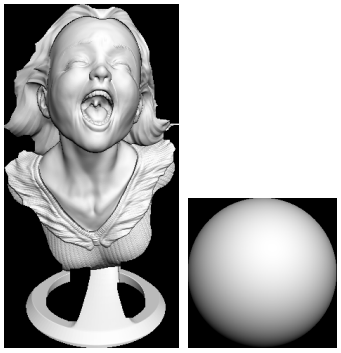


RMSE = $0,08$



RMSE = $0,10$

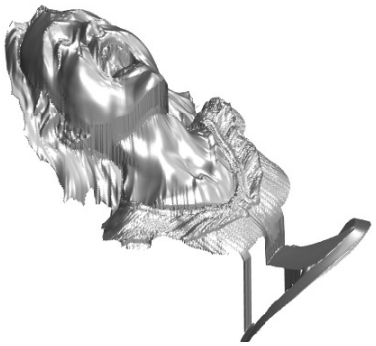
FIGURE 2.6 – Comparaison de notre méthode de résolution du SfS et de l'approche multi-échelles SIRFS [13], sous deux éclairages différents, avec la surface initiale « non réaliste » de la figure 2.5. Pour chaque test, nous affichons la fonction de profondeur estimée et la RMSE entre l'image initiale et l'image reprojetée.



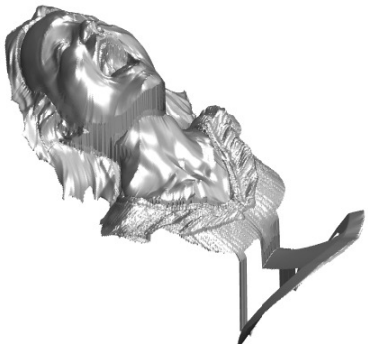
Éclairage $\boldsymbol{\sigma}_1$ d'ordre 1 (2.54).

Éclairage $\boldsymbol{\sigma}_2$ d'ordre 2 (2.55).

Nos résultats



RMSE = **0,03**



RMSE = **0,03**

SIRFS

RMSE = 0,06

RMSE = 0,07

FIGURE 2.7 – Comparaison de notre méthode de résolution du SfS et de l'approche multi-échelles SIRFS [13], sous deux éclairages différents, avec la surface initiale « réaliste » de la figure 2.5. Pour chaque test, nous affichons la fonction de profondeur estimée et la RMSE entre l'image initiale et l'image reprojetée.

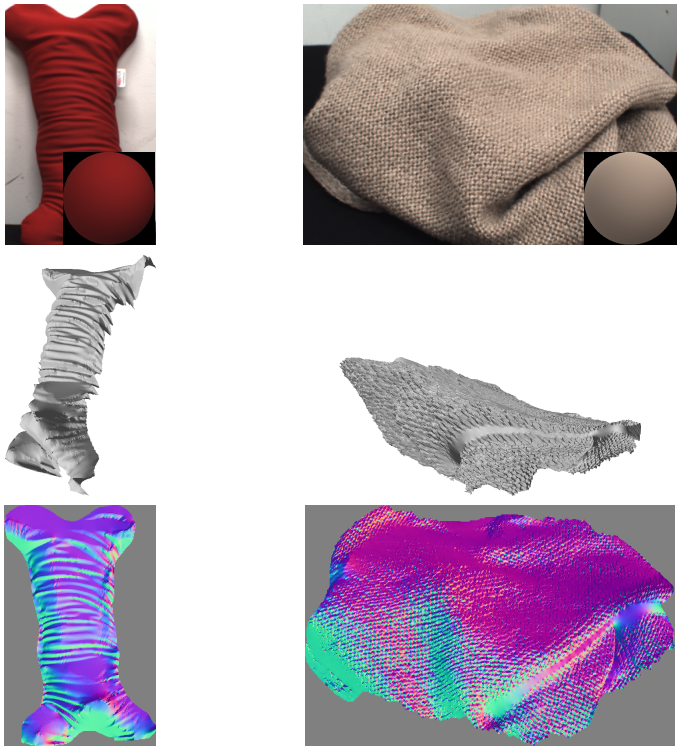


FIGURE 2.8 – Résultats de notre approche non régularisée sur deux jeux de données réelles [61]. Ligne du haut : image d'entrée et éclairage associé. Ligne du milieu : résultat du SfS (sans régularisation). La surface estimée est déformée à cause des ambiguïtés du SfS, et des artéfacts apparaissent à cause du bruit, mais les détails du relief sont retrouvés. Ligne du bas : champ de normales estimé, affiché en fausses couleurs.

FIGURE 2.9 – À gauche : image d'entrée bruitée (écart-type égal à 2 % du niveau de gris maximal). À droite : surface initiale bruitée (écart-type égal à $0,2\,\%$ de la profondeur maximale), représentée sous la forme d'un champ de normales en fausses couleurs afin de mettre en évidence les détails.

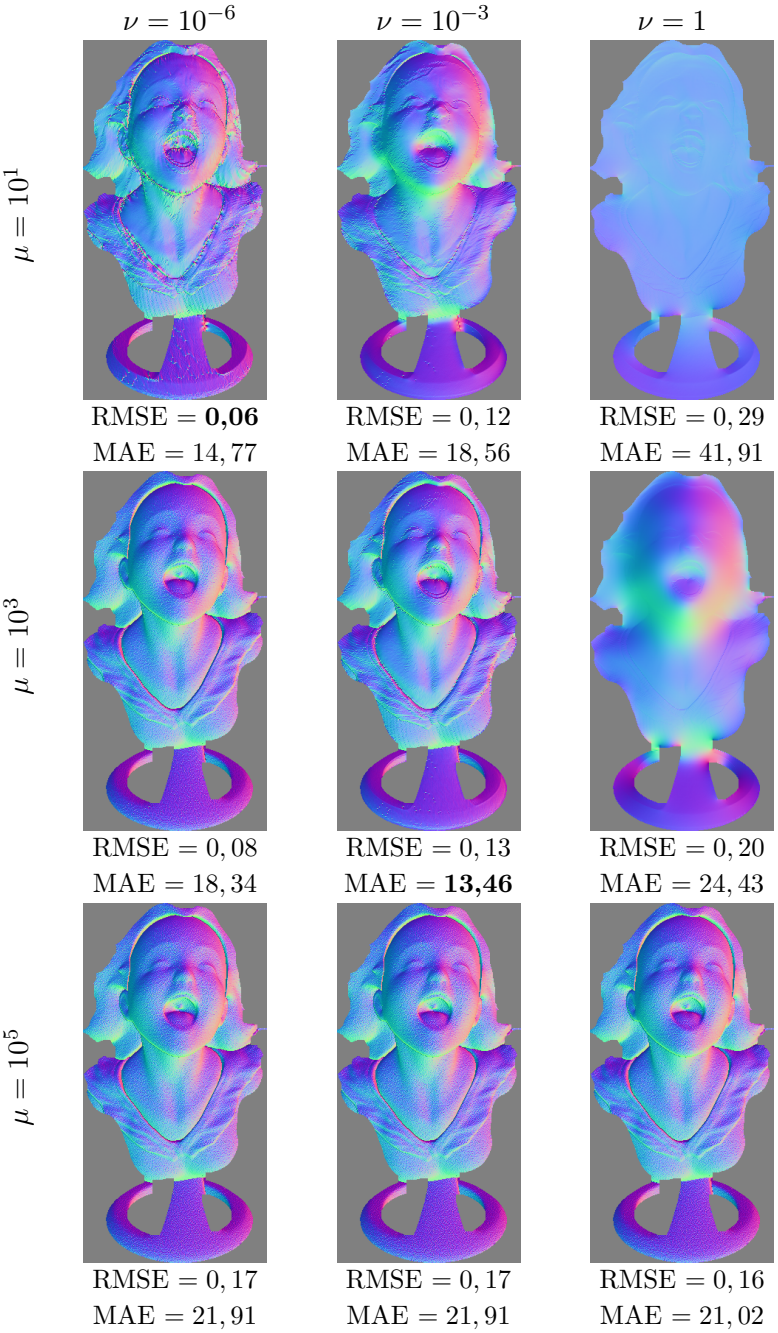


FIGURE 2.10 – Surfaces estimées pour $\lambda = 1$ et différentes valeurs de μ et ν. La RMSE entre l'image reprojetée et l'image initiale est minimale lorsque μ et ν sont minimaux, mais cela n'est pas le cas pour l'erreur angulaire moyenne (MAE, en degrés) entre les normales de la surface estimée et celles de la vérité terrain.

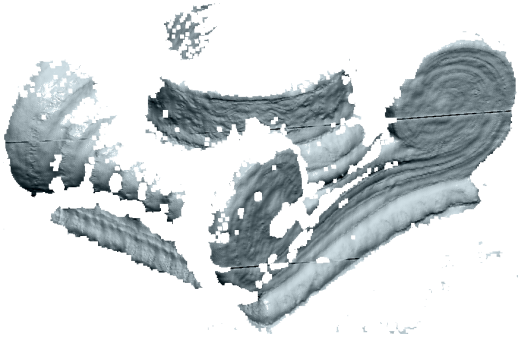
2.4.4 Évaluation de l'approche régularisée sur des données réelles

Le réglage approprié des hyper-paramètres μ et ν dépend de la confiance accordée à l'image et à la surface initiale. La fonction de profondeur obtenue par un capteur RGB-D se révélant généralement plus bruitée que dans nos tests sur données de synthèse, il nous faut diminuer la valeur de μ. Le triplet $(\lambda, \mu, \nu) = (1, 1, 5.10^{-5})$, déterminé de manière empirique, nous permet d'obtenir de bons résultats dans tous les tests sur données réelles.

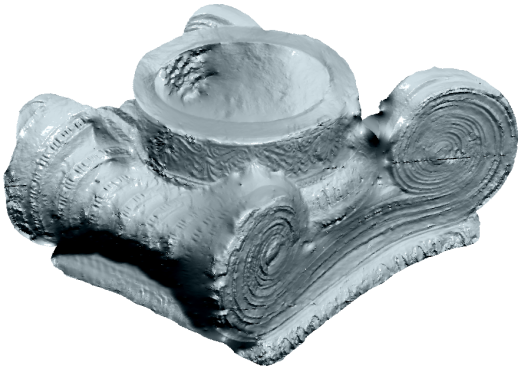
Afin de montrer qu'un a priori sur la profondeur permet de lever les ambiguïtés du SfS, nous utilisons à présent l'a priori z^0 fourni par le capteur de profondeur, qui sert également d'initialisation $z^{(0)}$. La première ligne de la figure 2.12, qui représente le relief fourni par un tel capteur pour les données de la figure 2.8, illustre l'aspect incomplet et bruité de cet a priori. Cependant, notre méthode permet d'effectuer à la fois la complétion, le débruitage et l'affinage de cet a priori, en adéquation avec l'ombrage de l'image. Les autres lignes de la figure 2.12 illustrent l'utilité du SfS pour l'affinage de la fonction de profondeur, par rapport à un débruitage à l'aveugle uniquement fondé sur la régularisation de la profondeur [57].

Image I_1.

Image I_2.

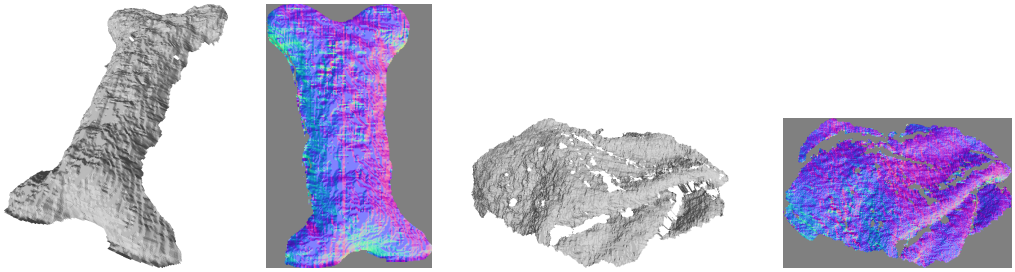


Fonction de profondeur initiale z_2^0 [80].

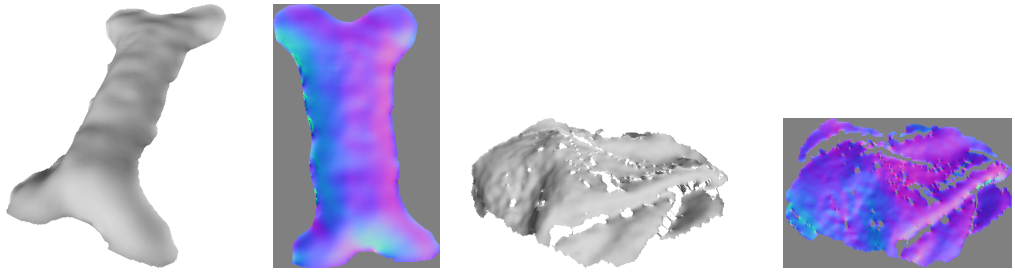


Fonction de profondeur z_2 affinée et complétée.

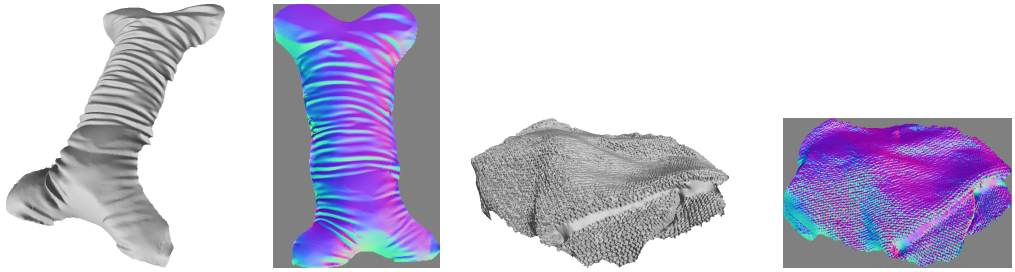
FIGURE 2.11 – En haut : deux images (parmi 30) du jeu de données « Figure » [179]. En bas à gauche : fonction de profondeur z_2^0 obtenue par MVS [80] (avant l'étape de maillage). En bas à droite : fonction de profondeur z_2 affinée et complétée. Ces deux fonctions de profondeur sont affichées sous l'angle correspondant à l'image I_2.



Fonctions de profondeur z^0 incomplètes et bruitées, et champs de normales associés [61].



Débruitage par surface minimale avec $(\lambda, \mu, \nu) = (0, 1, 5.10^{-5})$.



Affinage de la fonction de profondeur par l'ombrage avec $(\lambda, \mu, \nu) = (1, 1, 5.10^{-5})$.

FIGURE 2.12 – Ligne du haut : fonctions de profondeur et champs de normales associés, obtenus avec un capteur RGB-D, pour les deux scènes de la figure 2.8. Ligne du milieu : débruitage de cette fonction de profondeur. La profondeur est effectivement débruitée, mais les détails du relief sont perdus. Ligne du bas : lorsque l'information d'ombrage est utilisée, l'approche proposée permet de supprimer le bruit et de retrouver les détails du relief.

Enfin, la figure 2.11 nous ramène au contexte de la reconstruction 3D multi-vues, avec un jeu de données réelles issu de [179]. Dans ce test, la fonction de profondeur initiale est obtenue par un algorithme de stéréoscopie multi-vues (MVS). L'éclairage est estimé à partir de cette fonction de profondeur initiale, en supposant l'albédo uniforme. Nous laissons ensuite à notre algorithme le soin de retrouver les détails du relief que le MVS n'a pu restituer. Des ambiguïtés surviennent dans les nombreuses parties manquantes du relief initial, sachant que notre modèle favorise les surfaces lisses. Nous observons, en particulier, que les concavités ne sont pas bien retrouvées et que les discontinuités de profondeur sont partiellement lissées. Cependant, de nombreux détails du relief ont pu être retrouvés et la surface est globalement satisfaisante, au moins d'un point de vue qualitatif.

Conclusion

Dans ce chapitre, nous avons montré comment effectuer la reconstruction 3D par *shape-from-shading* dans des conditions opératoires peu contrôlées. De par les hypothèses, nombreuses et peu réalistes, qui sont nécessaires à sa mise en œuvre, il est d'usage de considérer que cette technique n'est pas utilisable en dehors du laboratoire. Dans la lignée de plusieurs travaux récents, nous avons proposé une méthode de résolution du SfS en conditions d'éclairage naturel. Cette méthode utilise une formulation différentielle et variationnelle du problème. L'utilisation d'un schéma ADMM permet d'effectuer la résolution en séparant les difficultés du problème, qui est à la fois non linéaire et non local.

Notre méthode de résolution peut être utilisée dans de nombreux scénarios, puisqu'elle reste valide pour différents types de projection et différents modèles d'éclairage. Nous avons en outre proposé deux termes de régularisation qui permettent de lever les ambiguïtés du SfS. Le premier permet de prendre en compte un a priori sur la profondeur qui peut être obtenu, par exemple, *via* l'approche multi-vues présentée dans le chapitre 1. Le deuxième terme de régularisation proposé vise à lisser la surface. Ces deux termes permettent en outre d'étendre le champ d'utilisation de notre algorithme. À partir d'une fonction de profondeur bruitée et incomplète, il est possible de procéder à un débruitage et, si l'on dispose également d'une image, de procéder à un affinage de la profondeur guidé par l'ombrage.

Si les résultats de ce dernier cas d'usage, à partir d'une fonction de profondeur obtenue par SfM/MVS, sont très satisfaisants, il semble évident qu'une telle approche, que nous pouvons qualifier de « séquentielle », n'est pas optimale. En effet, l'aspect multi-vues, qui n'est utilisé ici que pour obtenir une fonction de profondeur grossière, affinée a posteriori, ne tire aucun profit des informations apportées par l'ombrage. Dans le prochain chapitre, nous allons présenter une nouvelle approche de reconstruction 3D multi-vues spécifiquement conçue pour la reconstruction 3D d'objets peu texturés, qui s'appuie conjointement sur les méthodes SfM/MVS et SfS.

Chapitre 3

Stéréoscopie multi-vues et *shape-from-shading*

Les différentes techniques de reconstruction 3D présentées jusqu'ici, qui comportent des forces et des faiblesses, sont complémentaires : le pipeline SfM/MVS permet d'obtenir une surface globalement satisfaisante, mais les détails du relief sont perdus, tandis que le SfS explique bien les détails d'une image mais fournit des résultats dont la forme globale est biaisée. Chaque technique pouvant pallier les faiblesses de l'autre, il semble naturel de chercher à les faire coopérer. La méthode de résolution du SfS présentée dans le chapitre 2 nous a permis non seulement de nous affranchir de la nécessité d'opérer en conditions de laboratoire, mais également d'affiner une fonction de profondeur grossière obtenue par SfM/MVS. Il semble possible de pousser encore plus loin le bénéfice que chacune de ces techniques pourrait tirer de l'autre, dans la mesure où le pipeline SfM/MVS ne fait aucun usage du lien entre ombrage et relief, lien qui est le fondement même du SfS. Nous présentons dans ce chapitre une stratégie simple, mais efficace, qui permet de véritablement faire coopérer ces différentes techniques de reconstruction 3D.

Nous montrons dans le paragraphe 3.1 comment combiner MVS et SfS dans un modèle variationnel générique. Nous présentons ensuite, dans le paragraphe 3.2, un schéma numérique permettant de décomposer le problème d'optimisation qui résulte de ce nouveau modèle en une séquence de sous-problèmes plus simples à résoudre. Le paragraphe 3.3 étudie différents termes de régularisation adaptés au couplage MVS/SfS. Enfin, le paragraphe 3.4 illustre cette nouvelle approche par des résultats expérimentaux. Les travaux présentés dans ce chapitre ont été publiés dans [111] et [112].

3.1 Coopération entre MVS et SfS

Précisons dans un premier temps le cadre des travaux présentés dans ce chapitre. La méthode de reconstruction 3D qui a été décrite dans le chapitre 1 comporte deux étapes : le SfM estime les différentes poses de la caméra et un nuage de points 3D faiblement dense, puis le MVS, connaissant les poses, estime une fonction de profondeur dense par pose. Les lacunes mises en évidence dans le chapitre 1 concernent le MVS. Nous cherchons à améliorer cette deuxième étape du pipeline dans les mêmes conditions opératoires, c'est-à-dire en supposant que les poses de la caméra sont connues grâce au SfM. Rappelons à présent les forces et les faiblesses des deux techniques de reconstruction 3D dense que nous cherchons à combiner.

3.1.1 Des forces complémentaires

La technique du MVS, qui a été présentée dans le chapitre 1, permet d'obtenir une surface globalement satisfaisante, grâce au critère de la cohérence photométrique. Néanmoins, elle n'est pas adaptée aux parties non texturées de la surface, où la cohérence photométrique est sujette à des ambiguïtés. L'exemple de la figure 3.1 montre la dépendance entre la cohérence photométrique en un pixel situé dans une partie non texturée de la surface, et la profondeur du point 3D conjugué : il peut être délicat, voire impossible, de départager les différents minima globaux d'une telle courbe. Dans une situation de ce type, qui est illustrée sur la deuxième ligne de la figure 3.2, l'utilisation d'une deuxième vue (obtenue ici par translation de la caméra perspective) et de la seule cohérence photométrique procurent un résultat très bruité. Le recours à la régularisation semble naturel pour supprimer le bruit, mais les termes de régularisation utilisés sont généralement « purement géométriques » et consistent à lisser la surface [51], ce qui favorise les surfaces fronto-parallèles et gomme les détails du relief [175]. Les résultats accessibles sur le site [150], qui complète l'état de l'art [151], confirment bien que les méthodes actuelles les plus performantes peinent à retrouver convenablement les « hautes fréquences » de la surface.

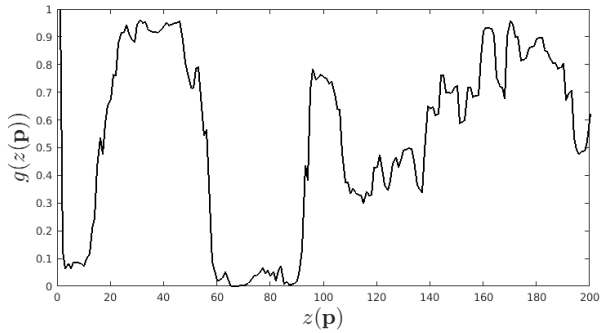


FIGURE 3.1 – Cohérence photométrique $g(z(\mathbf{p}))$, telle que définie en (3.2), en fonction de la profondeur $z(\mathbf{p})$, en un pixel $\mathbf{p}$ situé dans une partie non texturée de la surface. La présence de plusieurs minima globaux compromet l'optimisation.

Par ailleurs, la nouvelle méthode de résolution du SfS décrite dans le chapitre 2 cherche l'origine des variations d'ombrage dans les oscillations du relief. Néanmoins, les hypothèses nécessaires à sa mise en œuvre sont très fortes, puisque l'albédo de la surface à reconstruire et l'éclairage sont supposés connus. Comme nous l'avons déjà dit, nous reportons l'estimation de ces deux entités au chapitre 4. Mais, même sous ces hypothèses pourtant très favorables, le SfS reste sujet à l'ambiguïté concave/convexe, qui provoque un biais sur la forme globale de la surface, comme cela est illustré sur la troisième ligne de la figure 3.2.

En d'autres termes, le MVS et le SfS permettent de retrouver, respectivement, les « basses fréquences » et les « hautes fréquences » de la fonction de profondeur associée à chaque pose [124]. La combinaison de ces techniques doit donc nous permettre d'obtenir une fonction de profondeur restituant à la fois la forme globale et les détails du relief.

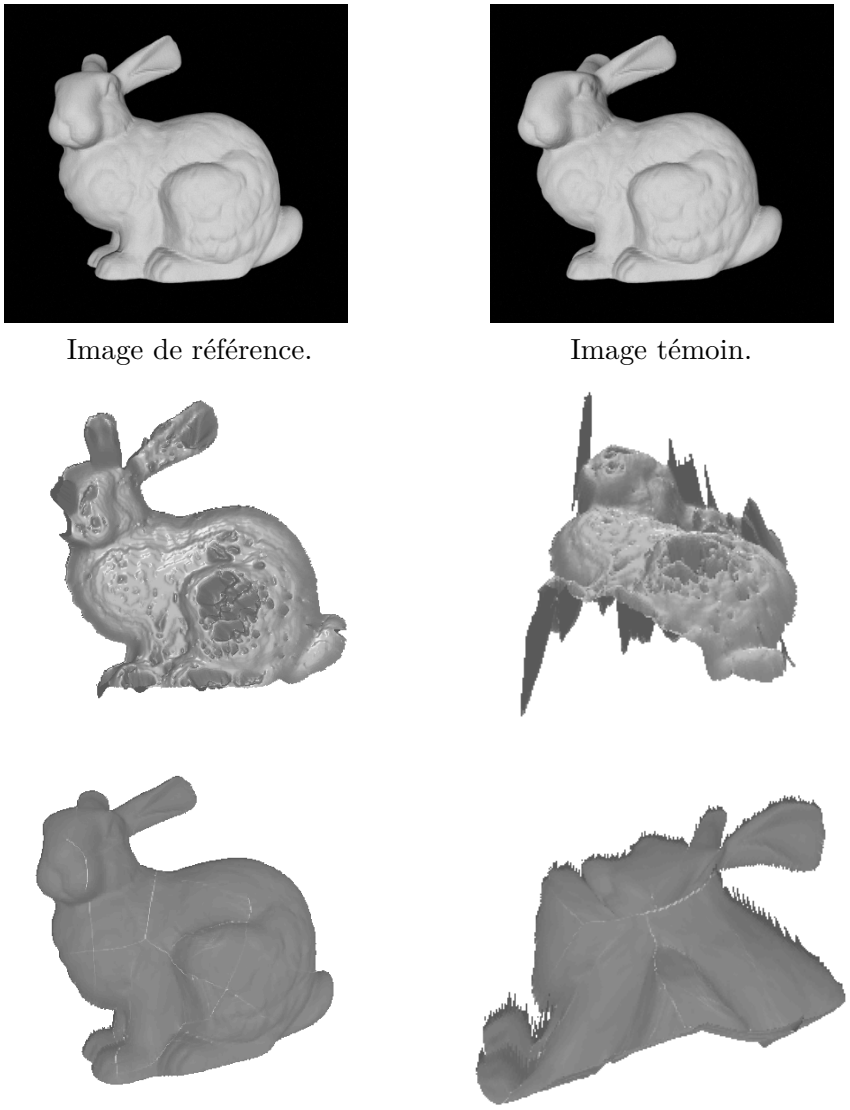


Image de référence. Image témoin.

FIGURE 3.2 – Première ligne : images de synthèse utilisées lors de nos premiers tests. L'image témoin est obtenue par translation de la caméra perspective. Deuxième ligne : MVS non régularisé. Si la forme globale de l'objet est retrouvée, les détails du relief ont disparu et, comme la mesure de la cohérence photométrique n'est pas suffisamment discriminante dans les zones non texturées, des artéfacts apparaissent (à droite, la surface est vue sous un autre angle). Troisième ligne : le SfS permet de retrouver les détails du relief, mais la forme globale de l'objet est biaisée à cause de l'ambiguïté concave/convexe (les angles de vue sont les mêmes que ceux de la deuxième ligne).

3.1.2 Différentes approches

La fusion d'informations provenant de la cohérence photométrique entre vues, d'une part, et d'informations fournies par l'ombrage, d'autre part, a été identifiée depuis longtemps comme une piste prometteuse [21]. Différentes approches ont été suivies. Il a notamment été proposé d'exécuter séparément MVS et SfS, puis de fusionner les deux reliefs obtenus, la difficulté consistant à effectuer cette fusion de manière pertinente [36, 60].

De nombreuses méthodes ont proposé une approche séquentielle : l'ombrage est alors considéré comme un moyen d'affiner la reconstruction 3D fondée principalement sur le MVS [96, 107, 173, 175, 179], sans que le MVS puisse profiter des informations apportées par l'ombrage. Réciproquement, l'extension du SfS à un cadre multi-vues présentée dans [82] se prive des informations précieuses apportées par le MVS. Afin que les deux techniques puissent « partager » les informations qu'elles sont censées fournir, nous optons pour une approche jointe de bout en bout, comme cela a été fait dans un travail très récent [109].

D'un point de vue plus théorique, une preuve d'unicité de la solution a été apportée dans le cas où deux poses de la caméra se déduisent l'une de l'autre par translation [31], mais cette preuve n'est valide que sous les hypothèses suivantes : la surface doit être lisse, blanche, lambertienne et éclairée par une seule source lumineuse parallèle et uniforme ; il ne doit y avoir ni ombre ni occultation. Dans ce chapitre, nous mettons de côté l'étude théorique de l'existence et de l'unicité de la solution, pour nous intéresser à la résolution numérique du problème dans des conditions opératoires réalistes.

3.1.3 Une nouvelle approche générique

Nous notons encore $z : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ la fonction de profondeur à estimer. La stéréoscopie multi-vues régularisée consiste généralement à minimiser la somme d'un terme d'attache aux données $g(z)$ traduisant la cohérence photométrique, et d'un terme de régularisation $h(z)$. Nous nous intéressons donc dorénavant au problème variationnel suivant :

$$\min_{z:\,\Omega\to\mathbb{R}} g(z) + h(z) \tag{3.1}$$

Les deux termes du problème variationnel (3.1) sont construits à partir de notions déjà abordées dans ce mémoire, sur lesquelles il nous faut maintenant revenir en détail.

3.1.3.1 Cohérence photométrique

Le terme d'attache aux données $g(z)$ du problème variationnel (3.1) traduit la cohérence photométrique. Il a déjà été établi dans le paragraphe 1.3 que la connaissance a priori des paramètres intrinsèques et des poses de la caméra nous permettait de définir la projection π d'un point 3D $\mathbf{X}$ visible dans l'image de référence en $\mathbf{p} = \pi(\mathbf{X}) \in \Omega$. Nous définissons $n-1$ autres projections π_i, $i \in \{2, \dots, n\}$, correspondant aux $n-1$ caméras témoins.

Si la fonction de profondeur z est connue, π devient une bijection dont la réciproque est notée π_z^{-1}, où l'indice souligne la nécessité de connaître z. La fonction de profondeur est alors recherchée comme solution du problème discret (1.48), qui compare les niveaux de gris de pixels hypothétiquement correspondants.

En réalité, la cohérence photométrique peut viser à comparer divers éléments : de manière générale, nous désignons par $\mathbf{v}(\mathbf{p}) \in \mathbb{R}^l$ un vecteur de taille l, caractérisant le pixel $\mathbf{p}$ dans l'image de référence. Un tel vecteur peut être simplement défini comme le niveau de gris de ce pixel ($l = 1$), ses coefficients RVB ($l = 3$), la concaténation des niveaux de gris sur un voisinage de taille 3×3 centré en $\mathbf{p}$ ($l = 9$), etc. Pour une caméra témoin $i \in \{2, \ldots, n\}$ donnée, la cohérence photométrique mesure alors l'écart entre le vecteur $\mathbf{v}(\mathbf{p})$ et le vecteur $\mathbf{v}_i \circ \pi_i \circ \pi_z^{-1}(\mathbf{p}) \in \mathbb{R}^l$ calculé en $\pi_i \circ \pi_z^{-1}(\mathbf{p})$. Le terme d'attache aux données $g(z)$ est alors naturellement défini comme la cohérence photométrique moyenne, calculée sur tous les pixels et toutes les caméras témoins :

$$g(z) = \frac{1}{n-1} \sum_{i=2}^{n} \frac{1}{|\Omega|} \sum_{p \in \Omega} \left\| \mathbf{v}_i \circ \pi_i \circ \pi_z^{-1}(\mathbf{p}) - \mathbf{v}(\mathbf{p}) \right\|^2 \tag{3.2}$$

De façon plus générale, cet écart peut être évalué au sens d'une fonction de coût κ :

$$g(z) = \frac{1}{n-1} \sum_{i=2}^{n} \frac{1}{|\Omega|} \sum_{p \in \Omega} \kappa \left(\mathbf{v}_i \circ \pi_i \circ \pi_z^{-1}(\mathbf{p}), \mathbf{v}(\mathbf{p}) \right) \tag{3.3}$$

Plusieurs fonctions de coût κ peuvent être choisies, par exemple la moyenne des différences au carré (SSD, pour *Sum of Squared Deviations*) :

$$\kappa_{\text{SSD}}(\mathbf{x}, \mathbf{y}) = \frac{1}{l} \sum_{j=1}^{l} (x_j - y_j)^2 , \qquad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^l \tag{3.4}$$

L'intérêt d'une fonction quadratique telle que (3.4) est qu'elle se prête bien à une linéarisation par un développement de Taylor, comme cela a été proposé dans [57, 109], mais cette linéarisation fait l'hypothèse de petits incréments de profondeur. De plus, les moindres carrés restent sensibles aux données aberrantes. Nous préférons utiliser une fonction κ robuste (cf. [47, chapitre 2]) calculant la moyenne des différences absolues (SAD, pour *Sum of Absolute Deviations*) :

$$\kappa_{\text{SAD}}(\mathbf{x}, \mathbf{y}) = \frac{1}{l} \sum_{j=1}^{l} |x_j - y_j|, \qquad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^l \tag{3.5}$$

ou mesurant la corrélation croisée (ZNCC, pour *Zero-mean Normalized Cross Correlation*) :

$$\kappa_{\text{ZNCC}}(\mathbf{x}, \mathbf{y}) = \frac{1}{2} \left[1 - \frac{(\mathbf{x} - \overline{\mathbf{x}})^\top (\mathbf{y} - \overline{\mathbf{y}})}{\|\mathbf{x} - \overline{\mathbf{x}}\| \, \|\mathbf{y} - \overline{\mathbf{y}}\|} \right] , \qquad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^l \tag{3.6}$$

Il est d'usage de normaliser les valeurs de la cohérence photométrique par un opérateur non linéaire, par exemple une transformation exponentielle dépendant d'un paramètre σ :

$$\kappa(\mathbf{x}, \mathbf{y}) := 1 - \exp\left\{ -\frac{\kappa(\mathbf{x}, \mathbf{y})^2}{\sigma^2} \right\} , \qquad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^l \tag{3.7}$$

qui permet d'étaler les valeurs de ces différentes fonctions de coût dans l'intervalle $[0; 1]$.

Entre toutes ces possibilités, le terme d'attache aux données peut se révéler non linéaire, non dérivable ou non convexe, et donc compliquer l'optimisation.

3.1.3.2 Régularisation

Nous avons vu que la cohérence photométrique n'était pas un critère pertinent pour les parties non texturées de la surface : en chaque point d'une telle partie, il existe souvent plusieurs minima globaux (cf. figure 3.1). Vu que l'augmentation du nombre d'images témoins ne résout en rien ce problème, l'ajout de termes de régularisation s'impose.

Généralement, les termes de régularisation utilisés pour le MVS visent à lisser la surface, par exemple par variation totale. Le problème régularisé consiste alors à minimiser la fonction $g(z) + \lambda \|\nabla z\|$, où λ est un hyper-paramètre [170]. Cependant, sous l'hypothèse de la projection perspective, la variation totale n'est pas une contrainte physiquement réaliste : il est préférable d'utiliser la régularisation par minimisation de la surface totale [57] déjà introduite dans le paragraphe 2.3.5.2, qui est une forme bilinéaire impliquant à la fois la profondeur z et son gradient ∇z. Grâce au changement de variable $\widetilde{z} = \log z$, ce terme de régularisation dépend du gradient $\nabla\widetilde{z}$, mais plus de la variable $\widetilde{z}$ elle-même. Nous verrons que ce changement de variable se prête bien aussi à l'ajout d'un terme de régularisation provenant du SfS [1]. Nous pourrons ainsi procéder à la comparaison entre deux types de régularisation : un lissage « à l'aveugle » et un lissage fondé sur une loi physique liant l'ombrage au relief. Afin de pouvoir appliquer ce changement de variable le moment venu, nous réécrivons donc le problème (3.1) sous la forme :

$$\min_{z\colon \Omega \to \mathbb{R}} g(z) + h(\nabla \log z) \tag{3.8}$$

La résolution du problème (3.8) peut se révéler complexe. En effet, aux difficultés induites par la fonction d'attache aux données g, la dépendance en $\nabla \log z$ du terme de lissage fait perdre au problème (3.8) son caractère local, puisque la résolution ne peut être faite en un pixel indépendamment de ses voisins. Nous proposons dans le prochain paragraphe une approche découplée, qui permet de décomposer ce problème en une séquence de sous-problèmes plus simples à résoudre.

3.2 Double découplage

Dans ce paragraphe, nous montrons comment transformer le problème variationnel (3.8) en un nouveau problème (3.12) plus simple à résoudre, et nous décrivons un algorithme permettant de résoudre ce dernier. Commençons par établir une formulation discrète du problème (3.8). Pour cela, nous introduisons le vecteur $\mathbf{z} \in \mathbb{R}^m$ ayant pour composantes les valeurs de la profondeur en chacun des m pixels de Ω, et nous notons $\mathbf{D} \in \mathbb{R}^{2m\times m}$ la matrice des différences finies du premier ordre, de sorte que le produit $\mathbf{D}\,\mathbf{z} \in \mathbb{R}^{2m}$ contient les approximations du gradient de la fonction de profondeur.

1. Il a été montré dans [57] que le changement de variable $\widetilde{z} = \frac{z^2}{2}$ permettait également d'exprimer la surface totale comme une fonction de $\nabla\widetilde{z}$ uniquement. Cependant, ce changement de variable se prête moins bien à l'ajout d'un terme de régularisation provenant du SfS.

3.2.1 Modèle variationnel

Comme cela a été dit dans le paragraphe précédent, la fonction d'attache aux données g du problème (3.8) est généralement une fonction de coût robuste mais non différentiable. Le couplage induit par l'opérateur gradient dans le terme de régularisation $h(\nabla \log z)$ rend les choses encore plus compliquées. Nous séparons ces difficultés en découplant les optimisations de ces deux termes, grâce à l'introduction de la variable auxiliaire $\mathbf{u} = \mathbf{z} \in \mathbb{R}^m$, ce qui nous amène à la version discrète suivante du problème (3.8) :

$$\begin{aligned} \min_{\mathbf{u},\mathbf{z}} \quad & g(\mathbf{u}) + h(\mathbf{D} \log \mathbf{z}) \\ \text{s.c.} \quad & \mathbf{u} = \mathbf{z} \end{aligned} \tag{3.9}$$

Dans (3.9), le sous-problème en $\mathbf{u}$ demeure non dérivable et potentiellement non convexe, mais devient désormais local (et donc parallélisable). En effet, g est séparable : dans (3.3), chaque terme de la somme sur i n'implique qu'une valeur de la profondeur $z(\mathbf{p})$. Il est donc possible de rechercher le minimum global de g de façon exhaustive (« en force brute ») sur un ensemble fini de valeurs de la profondeur.

De plus, si la fonction h est lisse, une simple méthode d'optimisation à direction de descente doit suffire pour minimiser le terme de régularisation. Cependant, la contrainte $\mathbf{u} = \mathbf{z}$ ne permet pas à $\mathbf{z}$ de traduire les oscillations rapides du relief, qui risquent d'être perdues à cause de la quantification de la profondeur. Pour cette raison, nous préférons suivre une approche mixte, où le sous-problème en $\mathbf{u}$ est résolu sur un ensemble discret, tandis que le sous-problème en $\mathbf{z}$ est résolu sur $\mathbb{R}$, de manière à ce que le résultat final $\mathbf{z}$ puisse restituer au mieux les détails du relief.

Par conséquent, les deux variables $\mathbf{u}$ et $\mathbf{z}$ diffèrent légèrement par essence, ce qui nous motive d'autant plus à relâcher la contrainte « dure » $\mathbf{u} = \mathbf{z}$ du problème (3.9), pour la transformer en un terme de pénalisation quadratique :

$$\min_{\mathbf{u},\mathbf{z}} g(\mathbf{u}) + h(\mathbf{D} \log \mathbf{z}) + \beta \left\| \log \mathbf{u} - \log \mathbf{z} \right\|^2 \tag{3.10}$$

où $\beta > 0$ est un hyper-paramètre. Nous choisissons d'appliquer cette pénalisation aux logarithmes afin que, dans (3.10), $\mathbf{z}$ apparaisse uniquement par le biais de son logarithme. Le changement de variable déjà utilisé $\widetilde{\mathbf{z}} = \log \mathbf{z}$ nous permet d'optimiser le problème par rapport à la nouvelle variable $\widetilde{\mathbf{z}}$, puis de calculer $\mathbf{z} = \exp \widetilde{\mathbf{z}}$ à la fin du processus. Le nouveau problème d'optimisation s'écrit donc :

$$\min_{\mathbf{u},\widetilde{\mathbf{z}}} g(\mathbf{u}) + h(\mathbf{D}\, \widetilde{\mathbf{z}}) + \beta \left\| \log \mathbf{u} - \widetilde{\mathbf{z}} \right\|^2 \tag{3.11}$$

Comme nous l'avons déjà signalé dans le paragraphe précédent, plusieurs études récentes ont plaidé pour l'utilisation de fonctions de régularisation h non linéaires, ce qui rend le sous-problème en $\widetilde{\mathbf{z}}$ complexe. Nous pouvons toutefois effectuer un nouveau découplage, grâce à la variable auxiliaire $\boldsymbol{\theta} = \mathbf{D}\, \widetilde{\mathbf{z}} \in \mathbb{R}^{2n}$. Le problème (3.11) devient alors :

$$\begin{aligned} \min_{\mathbf{u},\boldsymbol{\theta},\widetilde{\mathbf{z}}} \quad & g(\mathbf{u}) + h(\boldsymbol{\theta}) + \beta \left\| \log \mathbf{u} - \widetilde{\mathbf{z}} \right\|^2 \\ \text{s.c.} \quad & \boldsymbol{\theta} = \mathbf{D}\, \widetilde{\mathbf{z}} \end{aligned} \tag{3.12}$$

3.2.2 Résolution numérique

La contrainte linéaire du problème (3.12) pourrait être prise en compte en ayant recours, par exemple, à une approche de type Lagrangien augmenté. Nous préférons suivre une stratégie plus simple, qui consiste à approcher la solution en résolvant, à chaque itération, un problème de la forme suivante, qui comporte une pénalisation quadratique supplémentaire :

$$\min_{\mathbf{u},\boldsymbol{\theta},\widetilde{\mathbf{z}}} g(\mathbf{u}) + h(\boldsymbol{\theta}) + \alpha^{(k)} \left\| \theta - \mathbf{D}\,\widetilde{\mathbf{z}} \right\|^2 + \beta \left\| \log \mathbf{u} - \widetilde{\mathbf{z}} \right\|^2 \tag{3.13}$$

Dans ce modèle, la valeur du poids $\alpha^{(k)} > 0$ augmente au fil des itérations k. Nous souhaitons en effet que la contrainte dure du problème (3.12) soit satisfaite à la convergence, c'est-à-dire lorsque $k \to +\infty$, contrairement à la contrainte dure $\mathbf{u} = \mathbf{z}$ du problème (3.9), que nous avons volontairement relâchée et dont le poids β dans (3.13) est fixe.

Pour chaque valeur du poids α, la solution du problème (3.13) est approchée par une étape d'optimisation alternée des différents sous-problèmes. Comme nous l'avons déjà signalé, le sous-problème en $\mathbf{u}$ est résolu par une recherche exhaustive du minimum (force brute). Nous nous intéressons ici à des termes de régularisation $h(\boldsymbol{\theta})$ lisses et séparables, de sorte que le sous-problème en $\boldsymbol{\theta}$ peut être non seulement résolu par une méthode d'optimisation itérative utilisant le gradient, mais également parallélisé. Enfin, le sous-problème en $\widetilde{\mathbf{z}}$ est un problème en moindres carrés linéaires pouvant être résolu, par exemple, par la méthode des gradients conjugués. Le processus est répété jusqu'à ce que l'écart relatif e entre deux estimations successives de la solution $\mathbf{z} = \exp \widetilde{\mathbf{z}}$ passe en dessous de 10^{-4}.

Notre méthode de résolution est résumée dans l'algorithme 1. Intuitivement, cet algorithme commence par estimer une fonction de profondeur grossière en optimisant la cohérence photométrique (équation (3.14)). Il régularise ensuite les oscillations du relief (équation (3.15)), avant de calculer la fonction de log-profondeur par intégration du gradient ainsi obtenu (équation (3.16)). Les valeurs de $\alpha^{(0)}$ et β sont déterminées empiriquement.

3.3 Reconstruction 3D d'objets non texturés

Nous ne devons pas perdre de vue que le problème (3.11) est censé nous permettre de combiner les techniques de MVS et de SfS. Si n'importe quel terme de régularisation séparable et lisse est acceptable numériquement, certains sont tout de même préconisés. Nous décrivons dans ce paragraphe les deux termes de régularisation permettant de fournir une solution à notre problème de reconstruction 3D combinant MVS et SfS.

3.3.1 Régularisation par l'ombrage

S'il a déjà été établi dans le paragraphe 3.1.3.1 que l'attache aux données consistait en un terme de cohérence photométrique propre au MVS, la contribution du SfS a été, pour le moment, seulement évoquée.

Rappelons que le modèle générique de formation de l'image d'une surface lambertienne, sous un éclairage $\boldsymbol{\sigma} \in \mathbb{R}^9$, s'écrit $I(\mathbf{p}) = \rho(\mathbf{p})\, \boldsymbol{\sigma}^\top \boldsymbol{\nu}(\mathbf{p})$ (cf. paragraphe 2.3.2.3). Sous l'hypothèse de la projection perspective, il découle des expressions (2.37) et (2.29) que le vecteur

Algorithme 1 : Algorithme générique de reconstruction 3D multi-vues.

Entrées : $\mathbf{z}^{(0)}$, $\alpha^{(0)} > 0$, $\beta > 0$

Sorties : fonction de profondeur affinée $\mathbf{z}$

$\widetilde{\mathbf{z}}^{(0)} = \log \mathbf{z}^{(0)}$, $k = 0$, $e^{(0)} = +\infty$;

tant que $e^{(k)} > 10^{-4}$ **faire**

`// Optimisation de la cohérence photométrique`

$$\mathbf{u}^{(k+1)} = \underset{\mathbf{u}}{\operatorname{argmin}}\ g(\mathbf{u}) + \beta \left\| \log \mathbf{u} - \widetilde{\mathbf{z}}^{(k)} \right\|^2 ; \tag{3.14}$$

`// Régularisation des oscillations du relief`

$$\boldsymbol{\theta}^{(k+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\ h(\boldsymbol{\theta}) + \alpha^{(k)} \left\| \boldsymbol{\theta} - \mathbf{D}\, \widetilde{\mathbf{z}}^{(k)} \right\|^2 ; \tag{3.15}$$

`// Intégration`

$$\widetilde{\mathbf{z}}^{(k+1)} = \underset{\widetilde{\mathbf{z}}}{\operatorname{argmin}}\ \alpha^{(k)} \left\| \mathbf{D}\, \widetilde{\mathbf{z}} - \boldsymbol{\theta}^{(k+1)} \right\|^2 + \beta \left\| \widetilde{\mathbf{z}} - \log \mathbf{u}^{(k+1)} \right\|^2 ; \tag{3.16}$$

`// Mise à jour des variables auxiliaires`

$\alpha^{(k+1)} = 1,5\, \alpha^{(k)}$;

$\mathbf{z}^{(k+1)} = \exp \widetilde{\mathbf{z}}^{(k+1)}$;

$e^{(k+1)} = \frac{\left\| \mathbf{z}^{(k+1)} - \mathbf{z}^{(k)} \right\|}{\left\| \mathbf{z}^{(k)} \right\|}$;

$k = k + 1$;

fin

$\boldsymbol{\nu}(\mathbf{p})$ dépend du gradient de la log-profondeur $\widetilde{\mathbf{z}} = \log \mathbf{z}$, c'est-à-dire de $\boldsymbol{\theta}$. Afin de souligner cette dépendance, nous notons dorénavant $\boldsymbol{\nu}_{\boldsymbol{\theta}}(\mathbf{p}) := \boldsymbol{\nu}(\mathbf{p})$. Nous continuons de supposer l'albédo connu et, sans perte de généralité, nous posons $\rho(\mathbf{p}) = 1$. L'équation du niveau de gris en un pixel $\mathbf{p}$ s'écrit alors :

$$I(\mathbf{p}) = \boldsymbol{\sigma}^\top \boldsymbol{\nu}_{\boldsymbol{\theta}}(\mathbf{p}) \tag{3.17}$$

Cette équation, qui traduit le modèle lambertien, doit permettre de lever les ambiguïtés de mise en correspondance dans les parties non texturées, qui constituent le point faible du MVS. Si nous supposons que le modèle lambertien est satisfait à un bruit gaussien près, centré en zéro et homoscédastique, nous pouvons effectuer la résolution en moindres carrés de l'équation (3.17), dans l'esprit de l'approche variationnelle du SfS en conditions d'éclairage naturel présentée dans le chapitre 2, ce qui nous suggère la fonction de régularisation :

$$h_{\mathrm{SfS}}(\boldsymbol{\theta}) = \sum_{\mathbf{p} \in \Omega} \left| \boldsymbol{\sigma}^\top \boldsymbol{\nu}_{\boldsymbol{\theta}}(\mathbf{p}) - I(\mathbf{p}) \right|^2 \tag{3.18}$$

L'utilisation de la régularisation (3.18) dans l'algorithme 1 constitue effectivement une variante du MVS guidée par l'ombrage. Dans la mesure où $h_{\text{SfS}}(\boldsymbol{\theta})$ est non seulement lisse, mais également séparable, puisque chaque terme de la somme implique les composantes de $\boldsymbol{\theta}$ relatives à un seul pixel $\mathbf{p}$, le problème (3.15) peut être reformulé comme un ensemble de $m = |\Omega|$ problèmes d'optimisation non linéaires en dimension 2, qui peuvent être résolus en parallèle, par exemple par la méthode BFGS [104].

3.3.2 Régularisation par minimisation de la surface totale

La régularisation par l'ombrage requiert la connaissance du vecteur d'éclairage $\boldsymbol{\sigma}$. Dans certains cas, il peut s'avérer compliqué, voire impossible, d'étalonner l'éclairage, auquel cas il peut être préférable de ne pas utiliser de modèle explicite de formation de l'image. Nous revenons alors à une formulation classique du MVS régularisé. Nous préférons utiliser la minimisation de la surface totale, comme cela a déjà été expliqué dans le paragraphe 2.3.5.2. Nous proposons donc un deuxième terme de régularisation, défini comme suit :

$$h_{\text{MS}}(\boldsymbol{\theta}) = \sum_{\mathbf{p}\in\Omega} d_{\boldsymbol{\theta}}(\mathbf{p}) \tag{3.19}$$

où $d_{\boldsymbol{\theta}}(\mathbf{p})$ a déjà été défini en (2.31). La fonction h_{MS} est, tout comme h_{SfS}, lisse et séparable, de sorte que le problème (3.15) peut encore être résolu en parallèle pour les m pixels, par exemple par des itérations de la méthode BFGS.

3.3.3 Régularisation mixte

Bien entendu, le terme (3.19) de régularisation par minimisation de la surface totale, qui tend à favoriser les surfaces lisses, risque de gommer les détails du relief. Inversement, le terme (3.18) de régularisation par l'ombrage cherche à expliquer la moindre variation de niveau de gris par des oscillations de la surface, ce qui peut être une source de mauvaise interprétation du bruit. C'est pourquoi il semble intéressant de combiner ces deux types de régularisation. Nous utilisons donc, dans les tests, le terme de régularisation suivant :

$$h(\boldsymbol{\theta}) = \lambda\, h_{\text{SfS}}(\boldsymbol{\theta}) + \nu\, h_{\text{MS}}(\boldsymbol{\theta}) \tag{3.20}$$

où λ et ν sont des hyper-paramètres. Notons que h est, elle aussi, lisse et séparable, et que le terme combiné (3.20) se ramène à la régularisation par l'ombrage si $\lambda > 0$ et $\nu = 0$, et à la régularisation par minimisation de la surface totale si $\lambda = 0$ et $\nu > 0$.

3.4 Résultats expérimentaux

Dans l'ensemble des tests, les vecteurs caractéristiques $\mathbf{v}(\mathbf{p})$ sont construits par concaténation des niveaux de gris dans un voisinage du pixel $\mathbf{p}$ de taille 3×3. Sauf indication contraire, la fonction de coût κ utilisée dans (3.3) est la transformée exponentielle de la SAD (3.5), avec $\sigma = 0,2$. Les valeurs $\alpha^{(0)} = 1$ et $\beta = 0,1$, déterminées empiriquement, permettent d'obtenir des résultats satisfaisants. Elles sont donc systématiquement utilisées. Nous choisissons enfin comme fonction de profondeur initiale $\mathbf{z}^{(0)}$ un plan fronto-parallèle de profondeur égale à la moyenne des profondeurs de la vérité terrain.

3.4.1 Utilisation de deux images

Dans un premier temps, nous testons la modélisation (3.13), qui nous a permis de combiner MVS et SfS, sur des données de synthèse. La vérité terrain du modèle 3D « Stanford's Bunny » est représentée sur la figure 3.3. Les images, obtenues à partir d'un logiciel de rendu, sont toutes de taille 540×540. L'albédo y est uniforme, et l'éclairage $\boldsymbol{\sigma}$, comme les paramètres des caméras, sont connus. Un bruit gaussien d'écart-type égal à 1% du niveau de gris maximal a été ajouté, afin de rendre ces images plus réalistes.

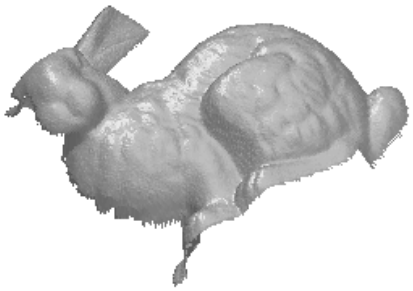
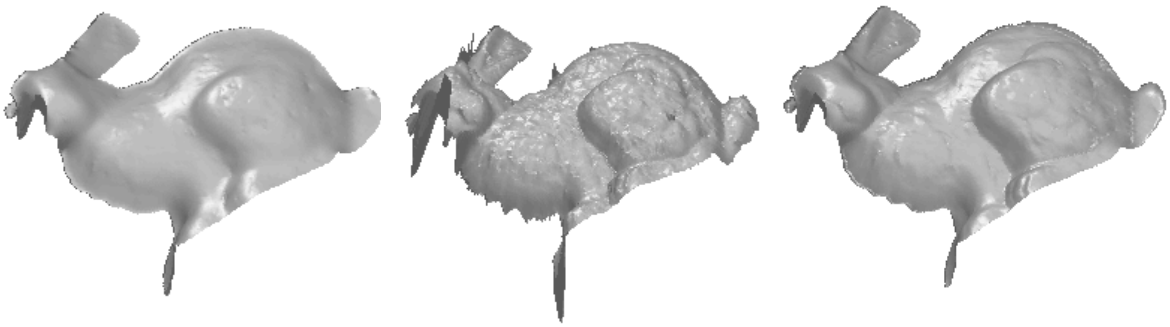


FIGURE 3.3 – Vérité terrain du modèle 3D « Stanford's Bunny » utilisé dans les tests.

Afin de valider la pertinence de notre approche, nous utilisons d'abord les deux images de ce modèle 3D qui sont représentées sur la première ligne de la figure 3.2. La comparaison entre les résultats de la deuxième ligne de la figure 3.2 et ceux de la figure 3.4 montre que l'ajout des termes de régularisation décrits dans le paragraphe précédent améliore sensiblement les résultats du MVS seul. Quand ils ne sont pas nuls, les hyper-paramètres prennent les valeurs suivantes, déterminées de manière empirique : $\lambda = 5.10^{-4}$ et $\nu = 5.10^{-5}$.

Comme nous pouvions nous y attendre, la régularisation par minimisation de la surface totale permet d'obtenir une fonction de profondeur sans bruit, globalement satisfaisante, mais dont les détails ont été gommés. Au contraire, la régularisation par l'ombrage permet de retrouver ces détails, mais l'utilisation d'une seule image témoin ($n = 2$) ne permet pas d'éviter la totalité des erreurs de reconstruction dues à l'ambiguïté concave/convexe. Enfin, l'utilisation simultanée des deux types de régularisation, qui combine leurs avantages, fournit les meilleurs résultats.

Nous présentons sur la figure 3.5, à des fins pédagogiques, un résultat où la contrainte dure $\mathbf{u} = \mathbf{z}$ n'est pas relâchée (en donnant à β une valeur très élevée). Les composantes du vecteur $\mathbf{z}$ ne peuvent donc prendre que les valeurs discrètes sur lesquelles la recherche en force brute est appliquée. La quantification de la profondeur apparaît clairement sur la surface reconstruite, au travers d'un « effet d'escalier ».



RMSE = 27,0 RMSE = 28,8 RMSE = **24,4**

FIGURE 3.4 – Stéréoscopie multi-vues régularisée utilisant une seule image témoin (les deux images sont celles de la figure 3.2). De gauche à droite : régularisation par minimisation de la surface totale ($\lambda = 0$), régularisation par l'ombrage ($\nu = 0$) et combinaison de ces deux types de régularisation ($\lambda > 0$ et $\nu > 0$).

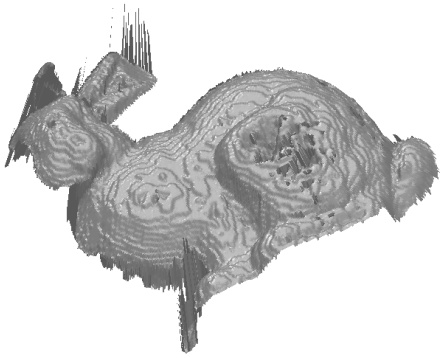


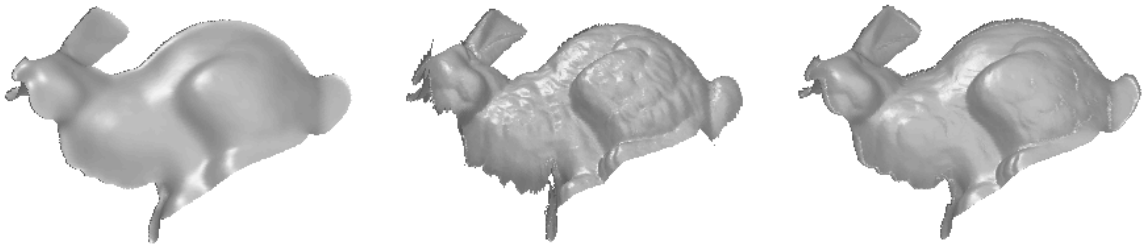
FIGURE 3.5 – Stéréoscopie multi-vues régularisée lorsque la contrainte dure $\mathbf{u} = \mathbf{z}$ n'est pas relâchée. La profondeur ne peut prendre qu'un nombre fini de valeurs, sur lesquelles la recherche en force brute est appliquée, ce qui provoque l'apparition d'un « effet d'escalier » sur la surface reconstruite.

3.4.2 Utilisation d'images supplémentaires

La question de la **visibilité** n'ayant pas été prise en compte explicitement, il est clair que les fonctions de profondeur représentées sur la figure 3.4 ne sont pas fiables pour les points de l'image de référence qui ne sont pas visibles dans l'image témoin. Cela est le cas, en particulier, de la zone située à proximité de l'oreille droite du lapin.

Pour éviter ce problème, il suffit d'augmenter le nombre d'images témoins, de telle sorte que chaque point de l'image de référence soit visible dans au moins une image témoin. La figure 3.6 montre trois des cinq images témoins supplémentaires utilisées. Effectivement, les résultats de la figure 3.7, qui utilisent $n - 1 = 6$ images témoins, ne comportent plus le défaut de la stéréoscopie binoculaire ($n = 2$). Les occultations résiduelles sont alors traitées comme des données aberrantes par la fonction d'attache aux données g, conçue pour être robuste. L'utilisation d'images témoins supplémentaires permet donc vraiment d'améliorer les résultats, comme le confirme la RMSE (en millimètres) mesurée par rapport à la vérité terrain. La régularisation par l'ombrage semble suffisante, alors que la régularisation par minimisation de la surface totale tend à lisser excessivement la surface.

FIGURE 3.6 – Images témoins supplémentaires permettant de garantir que chaque point de l'image de référence soit visible dans au moins une des $n - 1 = 6$ images témoins.



RMSE = 25,3 RMSE = **19,5** RMSE = 23,1

FIGURE 3.7 – Stéréoscopie multi-vues régularisée utilisant $n - 1 = 6$ images témoins. De gauche à droite : régularisation par minimisation de la surface totale ($\lambda = 0$), régularisation par l'ombrage ($\nu = 0$) et combinaison de ces deux types de régularisation ($\lambda > 0$ et $\nu > 0$). Les erreurs de reconstruction dues aux occultations (cf. figure 3.4) ont disparu.

3.4.3 Évaluation sur des données réelles

Pour finir, nous replaçons ces travaux dans le contexte de la prise de vues réelles, en utilisant les images du buste de l'empereur Auguste [179] représentées sur la figure 3.8.

FIGURE 3.8 – Images utilisées dans les tests sur données réelles. À gauche : image de référence. À droite : quatre des $n - 1 = 6$ images témoins.

FIGURE 3.9 – Comparaison des fonctions SAD (première ligne) et ZNCC (deuxième ligne) sur les images réelles de la figure 3.8. De gauche à droite : régularisation par minimisation de la surface totale ($\lambda = 0$, $\nu = 10^{-5}$), régularisation par l'ombrage ($\lambda = 5.10^{-3}$, $\nu = 0$) et combinaison de ces deux types de régularisation ($\lambda = 5.10^{-3}$, $\nu = 10^{-5}$).

Nous utilisons le pipeline de photogrammétrie AliceVision [7] pour estimer les paramètres de la caméra et une fonction de profondeur grossière relative à l'image de référence, à partir de laquelle sont évalués l'éclairage et la profondeur du plan fronto-parallèle initial. Cette fonction de profondeur grossière n'est plus utilisée par la suite (en particulier, elle n'est pas utilisée comme initialisation). Pour montrer que notre approche permet d'utiliser différentes mesures de cohérence photométrique, nous présentons les résultats obtenus avec les transformées exponentielles de SAD (3.5) et de ZNCC (3.6). Parmi les reconstructions 3D de la figure 3.9, la plus satisfaisante, du moins d'un point de vue qualitatif, est celle d'en bas à droite, qui est obtenue avec la transformée exponentielle ZNCC et une combinaison des deux types de régularisation.

Conclusion

Dans ce chapitre, nous avons présenté une méthode de résolution générique de la stéréoscopie multi-vues régularisée, qui permet d'utiliser diverses mesures de cohérence photométrique. Grâce à un double découplage, cette méthode permet de minimiser des fonctions de coût non linéaires, non dérivables et/ou non convexes.

L'algorithme proposé peut être appliqué à tout terme de régularisation lisse et séparable. L'ajout de termes de régularisation adéquats nous a permis de proposer une solution au problème de la reconstruction 3D d'objets peu ou pas texturés. Le terme de régularisation qui découle du SfS permet de recueillir des informations de hautes fréquences sur la profondeur. La régularisation par minimisation de la surface totale, qui est plus usuelle dans le cadre de la reconstruction 3D, permet d'éviter les erreurs de reconstruction résiduelles dues à l'ambiguïté concave/convexe et de rendre la méthode robuste au bruit. Notre approche présente également l'avantage de ne comporter que peu de paramètres à régler. En guise d'extension, elle pourrait être transformée en approche volumétrique, comme cela a été fait récemment dans [107], afin de retrouver un modèle 3D complet, et non pas seulement une fonction de profondeur relative à la vue de référence.

Les résultats sur données réelles valident notre approche. Cependant, comme dans le chapitre 2, les hypothèses de sa mise en œuvre restent très fortes. En particulier, le terme de régularisation par l'ombrage nécessite de connaître l'albédo et l'éclairage de la scène. Aussi, nous allons maintenant nous attacher à estimer ces deux entités, afin de pouvoir employer l'algorithme proposé dans un cadre plus général.

Chapitre 4

Estimation de l'albédo et de l'éclairage

Dans les chapitres précédents, nous avons montré comment estimer de manière précise le relief de la surface observée, grâce à une combinaison des techniques MVS et SfS, mais l'utilisation du SfS suppose l'albédo de la surface et l'éclairage connus. Nous montrons dans ce chapitre comment estimer ces entités à partir d'une connaissance grossière de la géométrie. Cette approche a déjà fait l'objet de plusieurs publications [113, 114, 115].

Dans le paragraphe 4.1, nous étudions le caractère mal posé de l'estimation de l'albédo et de l'éclairage en étendant le modèle lambertien (2.38) au contexte multi-vues. Le paragraphe 4.2 présente un panorama des solutions déjà proposées pour résoudre ce problème. Nous montrons dans le paragraphe 4.3 pourquoi il faut choisir comme variables les points de l'image. Dans le paragraphe 4.4, ce choix nous permet de reformuler le problème sous la forme d'un problème d'inférence bayésienne avec des a priori relativement simples. Nous détaillons ensuite, dans le paragraphe 4.5, la résolution du problème variationnel associé, au moyen d'un schéma numérique de type majoration-minimisation alternée. Enfin, nous présentons dans le paragraphe 4.6 quelques résultats obtenus avec cette approche.

4.1 Un problème d'estimation mal posé

Le processus de formation d'une image fait intervenir à la fois la géométrie de la surface, son albédo et l'éclairage incident. Si l'éclairage est connu, l'estimation de l'albédo est un problème mal posé en l'absence d'a priori sur la géométrie : le principe du trompe-l'œil repose sur cette ambiguïté. Cependant, le contexte multi-vues permet de réduire les ambiguïtés, puisqu'il nous permet d'appliquer la technique du MVS décrite dans le chapitre 1, et par conséquent de disposer d'un a priori sur la géométrie. Nous supposons donc, dans ce chapitre, que la géométrie de la surface est connue, au moins de manière grossière.

En revanche, nous supposons que nous ne connaissons ni l'albédo, ni l'éclairage incident de chaque prise de vue, ce qui complique notablement le problème. En effet, la connaissance du relief ne suffit pas à estimer conjointement l'albédo et l'éclairage, comme cela est illustré par la métaphore de l'atelier déjà mentionnée dans l'introduction : sur la figure 1, les deux explications (b) et (c) de l'image (a) correspondent au même relief plan. Pour lever cette ambiguïté sur le relief, il semble donc nécessaire de contraindre les variations spatiales de l'éclairage. Comme dans les chapitres précédents, nous continuons de modéliser l'éclairage par les harmoniques sphériques. Il est temps maintenant de revenir sur le modèle lambertien (2.38), afin de l'étendre au contexte multi-vues.

Nous autorisons dorénavant le vecteur d'éclairage à varier entre les prises de vue, ce qui nous amène à réécrire le modèle de formation de la $i^{\text{ème}}$ image de la façon suivante :

$$I_i(\pi_i(\mathbf{X})) = \rho(\mathbf{X})\, \boldsymbol{\sigma}_i^\top \boldsymbol{\nu}(\mathbf{X}), \quad i \in \{1, \dots, n\} \tag{4.1}$$

où $\rho(\mathbf{X}) > 0$ est l'albédo de la surface au point 3D $\mathbf{X} \in \mathcal{S}$, $\pi_i(\mathbf{X}) \in \mathbb{R}^2$ est la projection de ce point dans la $i^{\text{ème}}$ image, $\boldsymbol{\sigma}_i \in \mathbb{R}^9$ est le vecteur d'éclairage associé à la $i^{\text{ème}}$ prise de vue, et $\nu(\mathbf{X}) \in \mathbb{R}^9$ est le vecteur qui caractérise la géométrie de la surface au point $\mathbf{X}$. En vertu de nos hypothèses, ce dernier vecteur est supposé connu : il se déduit de la géométrie grossière par la définition (2.37), la normale à la surface étant paramétrée dans un repère de référence, par exemple le repère caméra de la première pose.

L'estimation de l'albédo et de l'éclairage revient donc à estimer la fonction $\rho(\mathbf{X})$, $\mathbf{X} \in \mathcal{S}$, et les vecteurs d'éclairage $\boldsymbol{\sigma}_i$, $i \in \{1, \dots, n\}$. Or, la difficulté de ce problème n'est pas la même selon que l'éclairage dépend ou non de l'indice i de la prise de vue.

4.1.1 Premier cas : éclairage fixe

S'il ne varie pas entre les différentes prises de vue, l'éclairage peut être caractérisé par un unique vecteur $\boldsymbol{\sigma} := \boldsymbol{\sigma}_i$, $i \in \{1, \dots, n\}$. L'équation (4.1) devient alors :

$$I_i(\pi_i(\mathbf{X})) = \rho(\mathbf{X})\, \boldsymbol{\sigma}^\top \boldsymbol{\nu}(\mathbf{X}), \quad i \in \{1, \dots, n\} \tag{4.2}$$

où le vecteur $\boldsymbol{\sigma} \in \mathbb{R}^9$ est inconnu, tout comme l'albédo $\rho(\mathbf{X})$, $\mathbf{X} \in \mathcal{S}$. Si m désigne le nombre de points 3D pour lesquels l'albédo doit être estimé, le problème comporte $m+9$ inconnues. Nous disposons pour les estimer de $m\,n$ équations telles que (4.2). Cependant, pour chaque point 3D, les n équations sont identiques, au bruit de mesure sur $I_i(\pi_i(\mathbf{X}))$ près. Seules m équations sont donc linéairement indépendantes. L'estimation des m paramètres de l'albédo et des 9 paramètres de l'éclairage constitue alors un problème mal contraint.

4.1.2 Deuxième cas : éclairage variable

Si l'éclairage change d'une image à l'autre, le nombre d'inconnues augmente : aux m valeurs de l'albédo s'ajoutent $9\,n$ paramètres d'éclairage. Cependant, les $m\,n$ équations deviennent linéairement indépendantes, pourvu que les vecteurs $\boldsymbol{\sigma}_i$ ne soient pas proportionnels. L'ordre de grandeur du nombre m de points 3D se situe entre 10^3 et 10^6, tandis que le nombre n d'images est plutôt de l'ordre de la dizaine. Il suffit donc que deux des n vecteurs d'éclairage ne soient pas colinéaires pour qu'il y ait plus d'équations linéairement indépendantes que d'inconnues, puisque $2\,m > m + 9\,n$ pour des valeurs typiques de m et n. Cette situation est similaire à celle de la stéréophotométrie non calibrée en éclairage naturel [16], mais en plus favorable : la géométrie étant connue, les ambiguïtés de la stéréophotométrie non calibrée n'ont pas lieu de survenir.

Néanmoins, dans notre cas d'usage, l'éclairage n'est généralement pas contrôlé. Nous ne pouvons donc pas garantir avec certitude qu'il existe au moins deux vecteurs d'éclairage non colinéaires. Si les images sont acquises en un court laps de temps, il est même probable que les éventuelles variations d'éclairage soient limitées. Les n vecteurs d'éclairage $\boldsymbol{\sigma}_i$, $i \in \{1, \dots, n\}$, sont alors presque égaux, ce qui signifie que le problème est mal conditionné.

Malgré ces différences, nous nous contentons dorénavant de supposer l'éclairage variable, vu que cela permet aussi bien de décrire le cas où les variations d'éclairage sont nulles.

4.1.3 Extension aux images RVB

Jusqu'à présent, nous avons toujours fait l'hypothèse d'images en niveaux de gris. Pour étendre notre étude au cas des images RVB, nous supposons (sans doute abusivement) que les canaux colorés ne sont pas corrélés, et nous appliquons les algorithmes proposés dans les chapitres précédents à chaque canal $\star \in \{R, V, B\}$. L'équation (4.1) devient alors :

$$I_i^\star(\pi_i(\mathbf{X})) = \rho^\star(\mathbf{X})\, \boldsymbol{\sigma}_i^{\star\top} \boldsymbol{\nu}(\mathbf{X}), \quad i \in \{1, \dots, n\}, \quad \star \in \{R, V, B\} \tag{4.3}$$

où $I_i^\star$ désigne la fonction de niveau de couleur de la $i^{\text{ème}}$ image, $\rho^\star(\mathbf{X})$ l'albédo au point $\mathbf{X}$, et $\boldsymbol{\sigma}_i^\star$ le $i^{\text{ème}}$ vecteur d'éclairage, toutes ces définitions étant relatives au canal $\star$. Une étude plus complète du modèle (4.3), notamment sur l'interprétation de $\rho^\star(\mathbf{X})$ et de $\boldsymbol{\sigma}_i^\star$ en termes de réponse de l'appareil photographique, est présentée dans [136].

Le modèle simplifié (4.3) rend indépendants les problèmes d'estimation de l'albédo et de l'éclairage dans les différents canaux. Nous pouvons donc nous concentrer sur les images en niveaux de gris, c'est-à-dire sur le modèle (4.2) de formation de l'image. Comme nous l'avons vu, l'estimation de l'albédo et de l'éclairage à partir de ce modèle constitue un problème mal posé, qui peut être seulement mal conditionné si l'éclairage varie faiblement entre les différentes prises de vue. La grande variété de connaissances a priori permettant de lever les ambiguïtés du problème explique probablement pourquoi autant de méthodes ont été proposées pour estimer l'albédo.

4.2 Bref état de l'art

Étant donné la nature mal posée de l'estimation de l'albédo, le recours à des connaissances a priori semble nécessaire. La plupart des travaux existants se fondent sur la **théorie Retinex** [94]. Cette théorie postule que l'albédo d'une image est constant par morceaux, comme dans un tableau de Mondrian, et que les variations progressives de l'intensité dans une image sont dues à l'éclairage. En d'autres termes, les variations importantes de l'albédo sont très localisées, tandis que l'éclairage induit des variations plus faibles du niveau de gris. Cet a priori suggère de séparer l'image en deux parties additives : une partie constante par morceaux, et une autre plus oscillante.

La communauté des mathématiques pour l'image désigne cette séparation sous le terme de **décomposition *cartoon* + texture**. De nombreux modèles variationnels ont été proposés pour résoudre ce problème [8], en utilisant par exemple un terme de régularisation non lisse et l'analyse fréquentielle de Fourier [9], ou un modèle l^1-TV [97]. Cependant, de telles méthodes n'utilisent pas explicitement de modèle photométrique, ce qui les rend relativement empiriques.

La communauté de la vision par ordinateur utilise plutôt la dénomination de **décomposition intrinsèque**. Des méthodes fondées sur la classification [49, 153] ou sur l'hypothèse que la scène comporte un petit nombre de valeurs de l'albédo [52, 121, 153, 154] ont été

développées à partir de la théorie Retinex. Parmi toutes ces méthodes, celle de Baron et Malik est probablement la plus connue [13]. Elle introduit de nombreux a priori sur le relief, l'albédo et l'ombrage. D'autres méthodes nécessitent une intervention de l'utilisateur, qui est invité à annoter les zones de l'image censées partager la même réflectance [22, 121], ou à utiliser une base de données collaborative [19].

Au lieu d'introduire des a priori difficilement vérifiables, ou de s'appuyer sur l'intervention d'un utilisateur, les ambiguïtés inhérentes à l'estimation de l'albédo peuvent être réduites en s'appuyant à la fois sur la connaissance partielle de la géométrie et sur des données multi-vues. Contrairement à certaines méthodes récentes de reconstruction 3D multi-vues visant à estimer conjointement la géométrie, l'albédo et l'éclairage [82, 88, 96, 107, 109], nous nous concentrons sur l'estimation de l'albédo et de l'éclairage, la géométrie étant supposée connue (estimée par MVS, dans notre cas). De ce point de vue, la méthode décrite dans ce chapitre est proche de la décomposition intrinsèque de données RGB-D [32], mais l'utilisation de données multi-vues la rapproche également de [92, 93].

4.3 Estimation de l'albédo et de l'éclairage dans un contexte multi-vues

La première question qui se pose est celle du paramétrage. La géométrie étant supposée connue, l'albédo pourrait être paramétré directement sur la surface (par exemple, en associant une valeur de l'albédo à chaque sommet du maillage). Une autre possibilité consiste à associer à chaque vue une fonction d'albédo. Nous évaluons ci-après les avantages et les inconvénients de ces deux paramétrages.

4.3.1 Estimation de l'albédo sur la surface

Comme nous l'avons fait jusqu'à présent, nous nous situons à la sortie d'un pipeline de reconstruction 3D multi-vues. Par conséquent, les projections π_i sont connues, tout comme la géométrie, sous la forme d'un ensemble de m points 3D $\mathbf{X}^j \in \mathbb{R}^3$, $j \in \{1, \dots, m\}$, et d'un ensemble de normales $\mathbf{n}(\mathbf{X}^j)$, à partir desquelles m vecteurs géométriques $\boldsymbol{\nu}^j := \boldsymbol{\nu}(\mathbf{X}^j)$ peuvent être déduits de la définition (2.37).

Les inconnues sont donc les m valeurs de l'albédo $\rho^j := \rho(\mathbf{X}^j) \in \mathbb{R}$ et les n vecteurs d'éclairage $\boldsymbol{\sigma}_i \in \mathbb{R}^9$. À première vue, nous pourrions penser que leur estimation puisse être menée par une approche de type maximum de vraisemblance, en effectuant la résolution approchée du système d'équations (4.1) à l'aide d'une fonction de coût κ :

$$\min_{\substack{\{\rho^j \in \mathbb{R}\}_j \\ \{\boldsymbol{\sigma}_i \in \mathbb{R}^9\}_i}} \sum_{i=1}^{n} \sum_{j=1}^{m} v_i^j \, \kappa\left(\rho^j \, \boldsymbol{\sigma}_i^\top \boldsymbol{\nu}^j, I_i^j\right) \tag{4.4}$$

où I_i^j est une notation abrégée pour $I_i(\pi_i(\mathbf{X}^j))$, et où v_i^j est un booléen qui indique si le point 3D $\mathbf{X}^j$ est visible ou non dans la $i^{\text{ème}}$ image.

Dans un but purement pédagogique, considérons le cas le plus simple des moindres carrés ($\kappa(x, y) = (x - y)^2$) et supposons la visibilité parfaite ($v_i^j \equiv 1$). Le problème (4.4)

peut alors être réécrit sous forme matricielle :

$$\min_{\substack{\boldsymbol{\rho}\in\mathbb{R}^m \\ \mathbf{S}\in\mathbb{R}^{9\times n}}} \|\mathbf{N}\,(\mathbf{S}\otimes\boldsymbol{\rho}) - \mathbf{I}\|_F^2 \tag{4.5}$$

Dans cette expression, $\|\cdot\|_F$ désigne la norme de Frobenius, le produit de Kronecker $\mathbf{S}\otimes\boldsymbol{\rho}$ est une matrice de $\mathbb{R}^{9m\times n}$, et les notations suivantes sont utilisées :

- le vecteur $\boldsymbol{\rho}\in\mathbb{R}^m$ contient les m valeurs de l'albédo à estimer ;
- la matrice $\mathbf{S}\in\mathbb{R}^{9\times n}$ est construite par concaténation des n vecteurs d'éclairage inconnus ;
- la matrice $\mathbf{N}\in\mathbb{R}^{m\times 9m}$ est diagonale par blocs, le $j^{\text{ème}}$ bloc étant le vecteur ligne ${\boldsymbol{\nu}^j}^\top$;
- la matrice $\mathbf{I}\in\mathbb{R}^{m\times n}$ contient les $m\,n$ niveaux de gris des données.

En utilisant la pseudo-inverse $\mathbf{N}^\dagger$ de $\mathbf{N}$, le problème (4.5) se réécrit :

$$\min_{\substack{\boldsymbol{\rho}\in\mathbb{R}^m \\ \mathbf{S}\in\mathbb{R}^{9\times n}}} \left\|\mathbf{S}\otimes\boldsymbol{\rho} - \mathbf{N}^\dagger\,\mathbf{I}\right\|_F^2 \tag{4.6}$$

qui est un problème de *recherche du produit de Kronecker le plus proche*, dont la résolution peut être effectuée par décomposition en valeurs singulières (SVD) [56, théorème 12.3.1].

Cette approche par factorisation matricielle présente néanmoins trois défauts. Primo, elle n'est valide que si tous les points 3D sont visibles dans toutes les images, ce qui est peu réaliste. En pratique, le problème (4.5) doit être remplacé par :

$$\min_{\substack{\boldsymbol{\rho}\in\mathbb{R}^m \\ \mathbf{S}\in\mathbb{R}^{9\times n}}} \|\mathbf{V}\circ[\mathbf{N}\,(\mathbf{S}\otimes\boldsymbol{\rho}) - \mathbf{I}]\|_F^2 \tag{4.7}$$

où $\mathbf{V}\in\{0,1\}^{m\times n}$ est la matrice de visibilité, qui contient les valeurs v_i^j, et $\circ$ désigne le produit de Hadamard. Or, ce problème de recherche du produit de Kronecker le plus proche avec des données manquantes est encore plus complexe à résoudre.

Secundo, elle n'est adaptée qu'à l'estimation en moindres carrés. Avec une fonction d'attache aux données plus robuste, la résolution directe par SVD n'est plus envisageable.

Tertio, si l'éclairage est constant, c'est-à-dire si $\boldsymbol{\sigma}_i = \boldsymbol{\sigma}, \forall i\in\{1,\dots,n\}$, alors il peut être montré que le problème (4.5) est mal posé, comme nous l'avons suggéré dans le paragraphe 4.1.1. En effet, une des nombreuses solutions est la solution triviale suivante :

$$\mathbf{S}_{\text{triviale}} = \boldsymbol{\sigma}_{\text{diffuse}}\,\mathbb{1}_{1\times n} \tag{4.8}$$

$$\boldsymbol{\rho}_{\text{triviale}} = \left[\overline{I}_i^1,\dots,\overline{I}_i^m\right]^\top \tag{4.9}$$

où :

$$\boldsymbol{\sigma}_{\text{diffuse}} = [1,0,0,0,0,0,0,0,0]^\top \tag{4.10}$$

et $\overline{I}_i^j$ désigne la moyenne des niveaux de gris des observations du point 3D $\mathbf{X}^j$ (moyenne sur l'indice i). L'ombrage n'est donc pas séparé de l'albédo : cette solution naïve de plaquage de texture est pourtant implémentée dans la plupart des pipelines de reconstruction 3D. Cela est le cas, par exemple, des figures 1.6 et 1.7.

Or, nous avons vu que de nombreux scénarios réels se rapprochaient du cas dégénéré où l'éclairage est constant (cf. paragraphe 4.1). Ceci nous incite à ajouter un terme de régularisation au problème (4.4) :

$$\min_{\substack{\{\rho^j \in \mathbb{R}\}_j \\ \{\boldsymbol{\sigma}_i \in \mathbb{R}^9\}_i}} \sum_{i=1}^{n} \sum_{j=1}^{m} v_i^j \, \kappa\Big(\rho^j \, \boldsymbol{\sigma}_i^\top \boldsymbol{\nu}^j, I_i^j\Big) + \sum_{j=1}^{m} \sum_{k \in \mathcal{I}^j} R(\rho^j, \rho^k) \tag{4.11}$$

où $\mathcal{I}^j$ désigne l'ensemble des indices des points 3D voisins de $\mathbf{X}^j$ sur la surface $\mathcal{S}$, et où la fonction de régularisation R doit être choisie de manière appropriée, en forçant par exemple l'albédo à être constant par morceaux.

Cependant, le relief de la surface $\mathcal{S}$ pouvant être complexe, l'échantillonnage des points $\mathbf{X}^j$ sur $\mathcal{S}$ est potentiellement non uniforme. Il semble alors compliqué de définir des termes d'attache aux données et de régularisation adéquats, et d'élaborer un schéma numérique de résolution approprié. De plus, les faibles variations d'intensité dans les images peuvent ne pas être perçues si l'échantillonnage n'est pas suffisamment dense.

L'estimation directe de l'albédo sur la surface, qui semblait prometteuse, se révèle donc délicate à réaliser en pratique. Nous préférons suivre une stratégie plus simple, qui consiste à paramétrer l'albédo dans l'image.

4.3.2 Estimation de l'albédo dans l'image

Nous proposons de décrire l'albédo de la surface par un ensemble de fonctions :

$$\rho_i : \Omega_i \subset \mathbb{R}^2 \to \mathbb{R}, \quad i \in \{1, \dots, n\} \tag{4.12}$$

qui associent à chaque point $\mathbf{p}$ de $\Omega_i := \pi_i(\mathcal{S})$ la valeur de l'albédo au point ${\pi_i}^{-1}(\mathbf{p})$, c'est-à-dire au point 3D conjugué de $\mathbf{p}$. Le modèle (4.1) peut alors être réécrit de la façon suivante :

$$I_i(\mathbf{p}) = \rho_i(\mathbf{p}) \, \boldsymbol{\sigma}_i^\top \boldsymbol{\nu}_i(\mathbf{p}), \quad \mathbf{p} \in \Omega_i, \quad i \in \{1, \dots, n\} \tag{4.13}$$

où $\boldsymbol{\nu}_i(\mathbf{p}) := \boldsymbol{\nu}({\pi_i}^{-1}(\mathbf{p}))$. Au lieu d'estimer une valeur $\rho(\mathbf{X})$ de l'albédo par point 3D, le problème consiste maintenant à estimer n fonctions d'albédo.

Le paramétrage (4.12) présente quand même quelques inconvénients. Tout d'abord, il ne garantit pas la cohérence entre fonctions d'albédo. Celle-ci devra donc être explicitée en introduisant une contrainte inter-vues. De plus, la surface ne sera pas directement coloriée : les fonctions d'albédo estimées devront être fusionnées a posteriori, avant d'êtres plaquées sur le maillage 3D. Fort heureusement, la fusion de fonctions avant plaquage sur un maillage est d'ores et déjà implémentée dans les pipelines de MVS : il nous suffira de remplacer les images brutes utilisées par cette fusion par les fonctions d'albédo estimées. En outre, chaque ensemble Ω_i est un sous-ensemble d'une grille 2D régulière. Les problèmes des données manquantes et de l'échantillonnage étant contournés, cela facilitera d'autant la recherche de termes d'attache aux données et de régularisation adéquats.

Avec un tel paramétrage, le problème régularisé (4.11) devient :

$$\min_{\substack{\{\rho_i:\Omega_i\to\mathbb{R}\}_i \\ \{\boldsymbol{\sigma}_i\in\mathbb{R}^9\}_i}} \sum_{i=1}^{n}\sum_{\mathbf{p}\in\Omega_i} \kappa\Big(\rho_i(\mathbf{p})\,\boldsymbol{\sigma}_i^\top \boldsymbol{\nu}_i(\mathbf{p}), I_i(\mathbf{p})\Big) + \sum_{i=1}^{n}\sum_{\mathbf{p}\in\Omega_i}\sum_{\mathbf{q}\in\mathcal{V}_i(\mathbf{p})} R(\rho_i(\mathbf{p}),\rho_i(\mathbf{q})) \qquad (4.14)$$
$$\text{s.c. } C(\{\rho_i\}_i) = 0$$

où la fonction C doit être définie de façon à garantir la cohérence inter-vues, et où $\mathcal{V}_i(\mathbf{p})$ désigne l'ensemble des pixels de Ω_i voisins de $\mathbf{p}$. Notons au passage que la notion de voisinage est beaucoup plus facile à appréhender que pour le problème (4.11), dans la mesure où Ω_i est un sous-ensemble d'une grille 2D régulière.

Dans le prochain paragraphe, nous discutons du choix de fonctions κ, R et C adaptées au problème variationnel (4.14), en nous fondant sur une approche bayésienne.

4.4 Modèle variationnel fondé sur une approche bayésienne

Suivant l'exemple de Mumford [120], nous présentons dans ce paragraphe une approche bayésienne visant à construire un modèle variationnel pour l'estimation de l'albédo et de l'éclairage. Ce problème consiste à estimer n fonctions d'albédo $\rho_i : \Omega_i \to \mathbb{R}$ et n vecteurs d'éclairage $\boldsymbol{\sigma}_i \in \mathbb{R}^9$, à partir de n images $I_i : \Omega_i \to \mathbb{R}$, $i \in \{1,\dots,n\}$. Nous avons déjà établi qu'une approche par le maximum de vraisemblance était sans espoir, puisqu'il existe une solution triviale (cf. équations (4.8) à (4.10)). Nous optons donc pour une approche bayésienne, et cherchons à estimer $(\{\rho_i\}_i, \{\boldsymbol{\sigma}_i\}_i)$ par le maximum a posteriori de la distribution :

$$\mathcal{P}(\{\rho_i\}_i, \{\boldsymbol{\sigma}_i\}_i | \{I_i\}_i) = \frac{\mathcal{P}(\{I_i\}_i | \{\rho_i\}_i, \{\boldsymbol{\sigma}_i\}_i)\,\mathcal{P}(\{\rho_i\}_i, \{\boldsymbol{\sigma}_i\}_i)}{\mathcal{P}(\{I_i\}_i)} \qquad (4.15)$$

où les facteurs du numérateur correspondent, respectivement, à la vraisemblance et à l'a priori. Notons que le dénominateur de cette expression peut être éliminé du problème de maximisation, puisqu'il ne dépend ni de l'albédo, ni de l'éclairage.

4.4.1 Vraisemblance

En pratique, le modèle de formation de l'image (2.38) n'est jamais strictement satisfait, à cause du bruit, de l'ombre portée et des reflets brillants. Nous faisons l'hypothèse que ces écarts au modèle sont suffisamment peu denses pour être décrits par une loi de Laplace centrée en zéro, d'échelle contrôlée par un paramètre ζ, indépendamment de l'indice i :

$$\begin{aligned}\mathcal{P}(\{I_i\}_i | \{\rho_i\}_i, \{\boldsymbol{\sigma}_i\}_i) &= \prod_{i=1}^{n} \left(\frac{1}{2\zeta}\right)^{|\Omega_i|} \exp\left\{-\frac{1}{\zeta}\left\|\rho_i\,\boldsymbol{\sigma}_i^\top\boldsymbol{\nu}_i - I_i\right\|_{i,1}\right\} \\ &= \left(\frac{1}{2\zeta}\right)^{\sum_{i=1}^{n}|\Omega_i|} \exp\left\{-\frac{1}{\zeta}\sum_{i=1}^{n}\left\|\rho_i\,\boldsymbol{\sigma}_i^\top\boldsymbol{\nu}_i - I_i\right\|_{i,1}\right\} \qquad (4.16)\end{aligned}$$

Dans cette expression, $|\Omega_i|$ désigne le cardinal de Ω_i et $\|\cdot\|_{i,1}$ la norme ℓ^1 sur Ω_i.

4.4.2 A priori

L'a priori du modèle (4.15) peut être factorisé car, non seulement les fonctions d'albédo $\{\rho_i\}_i$ sont indépendantes des vecteurs d'éclairage $\{\boldsymbol{\sigma}_i\}_i$, mais les vecteurs d'éclairage sont indépendants entre eux :

$$\mathcal{P}(\{\rho_i\}_i, \{\boldsymbol{\sigma}_i\}_i) = \mathcal{P}(\{\rho_i\}_i)\, \mathcal{P}(\{\boldsymbol{\sigma}_i\}_i) \tag{4.17}$$

$$= \mathcal{P}(\{\rho_i\}_i) \prod_{i=1}^{n} \mathcal{P}(\boldsymbol{\sigma}_i) \tag{4.18}$$

Comme les vecteurs d'éclairage ne sont pas contraints, nous pouvons faire l'hypothèse d'une distribution uniforme indépendante de l'indice i, soit $\mathcal{P}(\boldsymbol{\sigma}_i) =$ constante. L'a priori sur l'éclairage étant indépendant des inconnues, nous pouvons l'écarter du processus d'inférence. Quant aux fonctions d'albédo, nous nous conformons à la théorie Retinex [94] en supposant chacune d'elles constante par morceaux. L'a priori le plus naturel pour une telle fonction est donné par le modèle de Potts :

$$\mathcal{P}(\rho_i) = \xi_i \exp\left\{-\frac{1}{\iota_i} \|\nabla \rho_i\|_{i,0}\right\} \tag{4.19}$$

Dans cette expression, ξ_i désigne un coefficient de normalisation, ι_i un paramètre d'échelle, et $\nabla \rho_i = [\partial_x \rho_i, \partial_y \rho_i]^\top$ le gradient de la fonction ρ_i (que l'on peut approcher, en pratique par des différences finies avant du premier ordre et une condition au bord de Neumann). Nous utilisons abusivement la notation $\|\nabla \rho_i\|_{i,0}$ de la norme ℓ^0 pour désigner :

$$\|\nabla \rho_i\|_{i,0} = \sum_{\mathbf{p} \in \Omega_i} \sum_{\mathbf{q} \in \mathcal{V}_i(\mathbf{p})} \delta\left(\rho_i(\mathbf{p}) - \rho_i(\mathbf{q})\right) \tag{4.20}$$

où $\delta(x) = 1$ si $x \neq 0$, et $\delta(x) = 0$ sinon.

L'albédo, qui par définition est caractéristique de la surface, ne dépend pas de la vue. Il en découle que les paramètres (ξ_i, ι_i) sont les mêmes pour tous les modèles de Potts (4.19), et que l'a priori $\mathcal{P}(\{\rho_i\}_i)$ sur l'albédo se ramène au produit de n distributions indépendantes de mêmes paramètres (ξ, ι) :

$$\mathcal{P}(\{\rho_i\}_i) = \xi^n \exp\left\{-\frac{1}{\iota} \sum_{i=1}^{n} \|\nabla \rho_i\|_{i,0}\right\} \tag{4.21}$$

pourvu que la cohérence entre les différentes fonctions d'albédo soit garantie par la contrainte linéaire suivante :

$$C_{i,j}(\mathbf{p}_i, \mathbf{p}_j)\left(\rho_i(\mathbf{p}_i) - \rho_j(\mathbf{p}_i)\right) = 0, \quad \forall (i,j) \in \{1, \ldots, n\}^2,\ \forall (\mathbf{p}_i, \mathbf{p}_j) \in \Omega_i \times \Omega_j \tag{4.22}$$

sachant que la « fonction de cohérence » $C_{i,j} : \Omega_i \times \Omega_j \to \{0, 1\}$ se déduit facilement des projections $\{\pi_i\}_i$:

$$C_{i,j}(\mathbf{p}_i, \mathbf{p}_j) = \begin{cases} 1 & \text{si } \pi_i^{-1}(\mathbf{p}_i) = \pi_j^{-1}(\mathbf{p}_j) \\ 0 & \text{sinon} \end{cases} \tag{4.23}$$

4.4.3 Modèle complet

La maximisation de la probabilité a posteriori (4.15) équivaut à la minimisation de l'opposé de son logarithme. En injectant dans (4.15) les équations (4.16) et (4.21), nous obtenons finalement le problème variationnel contraint suivant, qui explicite les fonctions κ, R et C du problème (4.14) :

$$\begin{aligned} &\min_{\substack{\{\rho_i:\,\Omega_i\to\mathbb{R}\}_i \\ \{\boldsymbol{\sigma}_i\in\mathbb{R}^9\}_i}} \sum_{i=1}^{n} \left\| \rho_i\, \boldsymbol{\sigma}_i^\top \boldsymbol{\nu}_i - I_i \right\|_{i,1} + \gamma \sum_{i=1}^{n} \|\nabla \rho_i\|_{i,0} \\ &\text{s.c.} \quad C_{i,j}\,(\rho_i - \rho_j) = 0,\ \forall (i,j) \in \{1,\dots,n\}^2 \end{aligned} \tag{4.24}$$

avec $\gamma = \zeta/\iota$.

4.4.4 Lien avec la décomposition *cartoon* + texture

En calculant le logarithme des deux membres de l'équation (4.13), il vient :

$$\widetilde{I}_i(\mathbf{p}) = \widetilde{\rho}_i(\mathbf{p}) + \log\left(\boldsymbol{\sigma}_i^\top \boldsymbol{\nu}_i(\mathbf{p})\right) \tag{4.25}$$

où le « tilde » désigne le logarithme. Le problème (4.24) peut donc être reformulé :

$$\begin{aligned} &\min_{\substack{\{\widetilde{\rho}_i:\,\Omega_i\to\mathbb{R}\}_i \\ \{\boldsymbol{\sigma}_i\in\mathbb{R}^9\}_i}} \sum_{i=1}^{n} \left\| \widetilde{\rho}_i + \log\left(\boldsymbol{\sigma}_i^\top \boldsymbol{\nu}_i\right) - \widetilde{I}_i \right\|_{i,1} + \gamma \sum_{i=1}^{n} \|\nabla \widetilde{\rho}_i\|_{i,0} \\ &\text{s.c.} \quad C_{i,j}(\widetilde{\rho}_i - \widetilde{\rho}_j) = 0,\ \forall (i,j) \in \{1,\dots,n\}^2 \end{aligned} \tag{4.26}$$

qui peut être interprété comme un problème de **décomposition *cartoon* + texture** multi-vues, où chaque log-image $\widetilde{I}$ est décomposée en une composante $\widetilde{\rho}_i$ lisse par morceaux, appelée « *cartoon* » (ici, le log-albédo), et une composante $\log\left(\boldsymbol{\sigma}_i^\top \boldsymbol{\nu}_i\right)$ qui contient les détails de hautes fréquences, ou « texture » (ici, le log-ombrage). Contrairement aux méthodes classiques visant à effectuer une telle décomposition, celle-ci utilise de manière explicite un modèle d'ombrage qui caractérise la composante de texture. Notons cependant qu'une telle décomposition n'est justifiée que si les log-images $\widetilde{I}_i$ sont utilisées. Dans le cadre de l'utilisation des images originales I_i, notre approche devrait plutôt être considérée comme une **décomposition *cartoon* × texture** multi-vues.

4.4.5 Relaxation bi-convexe du modèle

Le problème (4.24) est non convexe à cause du terme de régularisation en norme ℓ^0, et non différentiable à cause des termes d'attache aux données et de régularisation. Bien que des recherches aient été récemment menées sur la résolution de problèmes comportant des termes de régularisation en norme ℓ^0 [156], nous préférons nous contenter d'approcher ces termes de régularisation par des termes de variation totale anisotrope, qui sont non différentiables mais convexes :

$$\sum_{i=1}^{n} \|\nabla \rho_i\|_{i,0} \approx \sum_{i=1}^{n} \|\nabla \rho_i\|_{i,1} \tag{4.27}$$

En outre, la fonction de cohérence peut se révéler légèrement inexacte en pratique, à cause par exemple d'erreurs dans l'estimation préalable de la géométrie obtenue par MVS. Pour cette raison, nous transformons la contrainte linéaire du problème (4.24) en un terme de pénalisation supplémentaire. Finalement, nous approchons la valeur absolue découlant de la norme ℓ^1, qui est non différentiable, par son enveloppe de Moreau (différentiable), c'est-à-dire par la fonction de Huber [1] :

$$|x| \approx \mathcal{H}_\epsilon(x) := \begin{cases} \dfrac{x^2}{2\,\epsilon} & \text{si } |x| \leq \epsilon \\ |x| - \dfrac{\epsilon}{2} & \text{si } |x| > \epsilon \end{cases} \tag{4.28}$$

Ces différentes approximations nous permettent de réécrire (4.24) sous la forme d'un problème variationnel différentiable et bi-convexe :

$$\begin{aligned} \min_{\substack{\boldsymbol{\rho}:=\{\rho_i:\Omega_i\to\mathbb{R}\}_i \\ \boldsymbol{\sigma}:=\{\boldsymbol{\sigma}_i\in\mathbb{R}^9\}_i}} \varepsilon(\boldsymbol{\rho},\boldsymbol{\sigma}) := & \sum_{i=1}^{n} \sum_{\mathbf{p}\in\Omega_i} \mathcal{H}_\epsilon\left(\rho_i(\mathbf{p})\,\boldsymbol{\sigma}_i^\top \boldsymbol{\nu}_i(\mathbf{p}) - I_i(\mathbf{p})\right) \\ & + \gamma \sum_{i=1}^{n} \sum_{\mathbf{p}\in\Omega_i} \left[\mathcal{H}_\epsilon\left(\partial_x \rho_i(\mathbf{p})\right) + \mathcal{H}_\epsilon\left(\partial_y \rho_i(\mathbf{p})\right)\right] \\ & + \varrho \sum_{1\leq i<j\leq n} \sum_{\mathbf{p}_i\in\Omega_i} \sum_{\mathbf{p}_j\in\Omega_j} C_{i,j}(\mathbf{p}_i,\mathbf{p}_j)\,\mathcal{H}_\epsilon\left(\rho_i(\mathbf{p}_i) - \rho_j(\mathbf{p}_j)\right) \end{aligned} \tag{4.29}$$

Dans le problème (4.29), le premier terme vise à vérifier la loi de Lambert (au sens de la fonction de Huber), le deuxième impose la régularité de l'albédo (au sens de la variation totale anisotrope), tandis que le dernier garantit la cohérence entre les fonctions d'albédo estimées (au sens de la fonction de Huber). Enfin, γ et ϱ sont des hyper-paramètres qui permettent de contrôler, respectivement, la régularité de l'albédo et la cohérence inter-vues.

4.5 Résolution par majoration-minimisation alternée

Pour résoudre le problème (4.29), nous utilisons une méthode de majoration-minimisation alternée. Comme le montre le schéma de la figure 4.1, cet algorithme fonctionne de la façon suivante : à partir d'une estimation $(\boldsymbol{\rho}^{(k)}, \boldsymbol{\sigma}^{(k)})$ de la solution à l'itération (k), les fonctions d'albédo et les vecteurs d'éclairage sont successivement mis à jour comme suit :

$$\boldsymbol{\rho}^{(k+1)} = \operatorname*{argmin}_{\boldsymbol{\rho}} \varepsilon_{\boldsymbol{\rho}}^{(k)}(\boldsymbol{\rho}) \tag{4.30}$$

$$\boldsymbol{\sigma}^{(k+1)} = \operatorname*{argmin}_{\boldsymbol{\sigma}} \varepsilon_{\boldsymbol{\sigma}}^{(k)}(\boldsymbol{\sigma}) \tag{4.31}$$

où $\varepsilon_{\boldsymbol{\rho}}^{(k)}$ et $\varepsilon_{\boldsymbol{\sigma}}^{(k)}$ sont des majorants locaux quadratiques de $\varepsilon(\cdot, \boldsymbol{\sigma}^{(k)})$ et $\varepsilon(\boldsymbol{\rho}^{(k+1)}, \cdot)$ autour de $\boldsymbol{\rho}^{(k)}$ et $\boldsymbol{\sigma}^{(k)}$, respectivement. Ces itérations sont répétées jusqu'à la convergence.

1. Dans la partie expérimentale, nous utilisons $\epsilon = 10^{-4}$.

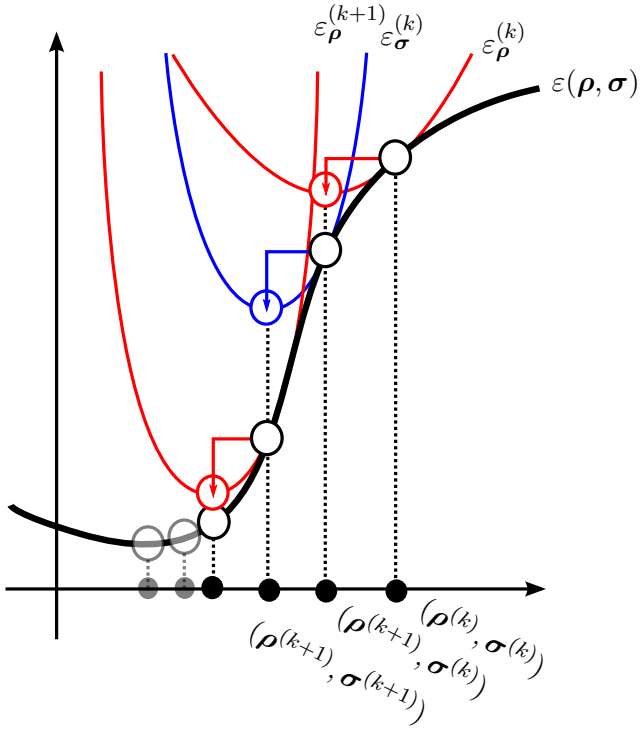


FIGURE 4.1 – Résolution par majoration-minimisation alternée. Les énergies partielles gelées $\varepsilon(\cdot,\boldsymbol{\sigma})$ et $\varepsilon(\boldsymbol{\rho},\cdot)$ sont localement majorées par les fonctions quadratiques $\varepsilon_{\boldsymbol{\rho}}$ (en rouge) et $\varepsilon_{\boldsymbol{\sigma}}$ (en bleu). Ces majorants quadratiques sont ensuite minimisés (globalement). Le processus est répété jusqu'à la convergence.

Nous pouvons remarquer que la fonction :

$$\overline{\mathcal{H}}_\epsilon(x;x_0) = \begin{cases} \dfrac{x^2}{2\,\epsilon} & \text{si } |x_0| \leq \epsilon \\ \dfrac{x^2}{2\,|x_0|} + \dfrac{|x_0|}{2} - \dfrac{\epsilon}{2} & \text{si } |x_0| > \epsilon \end{cases} \tag{4.32}$$

qui vérifie $\overline{\mathcal{H}}_\epsilon(x_0;x_0) = \mathcal{H}_\epsilon(x_0)$, se trouve être un majorant quadratique approprié de $\mathcal{H}_\epsilon$ au voisinage de x_0, $\forall\, x_0 \in \mathbb{R}$. Ceci est facile à vérifier si $|x_0| \leq \epsilon$, de par la définition (4.28) de $\mathcal{H}_\epsilon$. Si $|x_0| > \epsilon$, la différence $\overline{\mathcal{H}}_\epsilon(x;x_0) - \mathcal{H}_\epsilon(x)$ s'écrit :

$$\begin{cases} \dfrac{(|x_0| - \epsilon)\,(|x_0|\,\epsilon - x^2)}{2\,|x_0|\,\epsilon} & \text{si } |x| \leq \epsilon \\ \dfrac{(|x| - |x_0|)^2}{2\,|x_0|} & \text{si } |x| > \epsilon \end{cases} \tag{4.33}$$

qui est positive dans tous les cas.

Par conséquent, la fonction :

$$\begin{aligned}\varepsilon_{\boldsymbol{\rho}}^{(k)}(\boldsymbol{\rho}) := &\sum_{i=1}^{n}\sum_{\mathbf{p}\in\Omega_i}\overline{\mathcal{H}}_\epsilon\left(\rho_i(\mathbf{p})\,\boldsymbol{\sigma}_i^{(k)\top}\boldsymbol{\nu}_i(\mathbf{p}) - I_i(\mathbf{p}); e_i^{(k),(k)}\right)\\ &+\gamma\sum_{i=1}^{n}\sum_{\mathbf{p}\in\Omega_i}\left[\overline{\mathcal{H}}_\epsilon\left(\partial_x\rho_i(\mathbf{p}); \partial_x\rho_i^{(k)}(\mathbf{p})\right) + \overline{\mathcal{H}}_\epsilon\left(\partial_y\rho_i(\mathbf{p}); \partial_y\rho_i^{(k)}(\mathbf{p})\right)\right]\\ &+\varrho\sum_{1\le i<j\le n}\sum_{\mathbf{p}_i\in\Omega_i}\sum_{\mathbf{p}_j\in\Omega_j} C_{i,j}(\mathbf{p}_i,\mathbf{p}_j)\overline{\mathcal{H}}_\epsilon\left(\rho_i(\mathbf{p}_i) - \rho_j(\mathbf{p}_j); \rho_i^{(k)}(\mathbf{p}_i) - \rho_j^{(k)}(\mathbf{p}_j)\right)\end{aligned} \tag{4.34}$$

où :

$$e_i^{(k_1),(k_2)} = \boldsymbol{\rho}_i^{(k_1)}(\mathbf{p})\,\boldsymbol{\sigma}_i^{(k_2)\top}\boldsymbol{\nu}_i(\mathbf{p}) - I_i(\mathbf{p}) \tag{4.35}$$

est un majorant local quadratique de $\varepsilon(\cdot, \boldsymbol{\sigma}^{(k)})$ au voisinage de $\boldsymbol{\rho}^{(k)}$. Cette fonction est donc adaptée à la mise à jour (4.30). De la même manière, la fonction :

$$\begin{aligned}\varepsilon_{\sigma}^{(k)}(\boldsymbol{\sigma}) := &\sum_{i=1}^{n}\sum_{\mathbf{p}\in\Omega_i}\overline{\mathcal{H}}_\epsilon\left(\boldsymbol{\rho}_i^{(k+1)}(\mathbf{p})\boldsymbol{\sigma}_i^{\top}\boldsymbol{\nu}_i(\mathbf{p}) - I_i(\mathbf{p}); e_i^{(k+1),(k)}\right)\\ &+\gamma\sum_{i=1}^{n}\sum_{\mathbf{p}\in\Omega_i}\left[\mathcal{H}_\epsilon\left(\partial_x\boldsymbol{\rho}_i^{(k+1)}(\mathbf{p})\right) + \mathcal{H}_\epsilon\left(\partial_y\boldsymbol{\rho}_i^{(k+1)}(\mathbf{p})\right)\right]\\ &+\varrho\sum_{1\le i<j\le n}\sum_{\mathbf{p}_i\in\Omega_i}\sum_{\mathbf{p}_j\in\Omega_j}\left[C_{i,j}(\mathbf{p}_i,\mathbf{p}_j)\mathcal{H}_\epsilon\left(\boldsymbol{\rho}_i^{(k+1)}(\mathbf{p}_i) - \boldsymbol{\rho}_j^{(k+1)}(\mathbf{p}_j)\right)\right]\end{aligned} \tag{4.36}$$

est un majorant local quadratique de $\varepsilon(\boldsymbol{\rho}^{(k+1)}, \cdot)$ au voisinage de $\boldsymbol{\sigma}^{(k)}$. Cette fonction est donc adaptée à la mise à jour (4.31).

Le problème (4.30) se ramène alors à un problème en moindres carrés linéaires, que nous résolvons par la méthode des gradients conjugués appliquée aux équations normales. Le problème (4.31), quant à lui, se décompose en une série de n problèmes en moindres carrés linéaires, indépendants et de petite taille, qui peuvent être résolus au moyen de la pseudo-inverse.

Les étapes d'optimisation (4.30) et (4.31) sont itérées jusqu'à la convergence ou jusqu'à ce qu'un nombre maximal d'itérations soit atteint. Nous initialisons les variables avec la solution triviale du problème non régularisé ($\gamma = \varrho = 0$). Cette solution est obtenue en supposant l'éclairage diffus (cf. (4.10)) et en assimilant les fonctions d'albédo aux images d'entrée. Dans les tests, nous avons observé que 50 itérations suffisaient toujours pour atteindre une solution stable, c'est-à-dire pour que l'écart relatif entre deux valeurs successives $\varepsilon(\boldsymbol{\rho}^{(k)}, \boldsymbol{\sigma}^{(k)})$ et $\varepsilon(\boldsymbol{\rho}^{(k+1)}, \boldsymbol{\sigma}^{(k+1)})$ de l'énergie passe en dessous de 10^{-3}.

La preuve de convergence de ce schéma numérique dépasse l'objectif de ce travail. Une telle preuve pourrait certainement être inspirée de [136], qui utilise une méthode similaire de majoration-minimisation alternée, appelée *alternating reweighted least-squares*. Notons toutefois que le taux de convergence semble être sous-linéaire (cf. figure 4.2), ce qui laisse penser que des stratégies numériques plus rapides pourraient sans doute être trouvées.

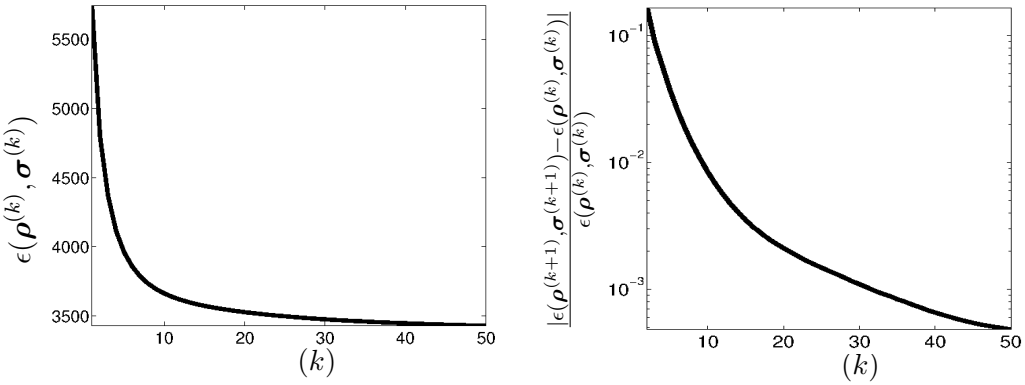


FIGURE 4.2 – À gauche : évolution de l'énergie $\varepsilon(\boldsymbol{\rho}^{(k)}, \boldsymbol{\sigma}^{(k)})$ définie en (4.29), en fonction du nombre d'itérations (k), pour le test de la figure 4.6. À droite : écart relatif entre deux valeurs successives de l'énergie. L'algorithme s'arrête lorsque cette valeur devient inférieure à 10^{-3}, ce qui se produit en moins de 50 itérations, et prend environ 3 minutes sur un processeur i7, avec un code Matlab non optimisé pour $n = 13$ images de taille 540×960.

4.6 Résultats expérimentaux

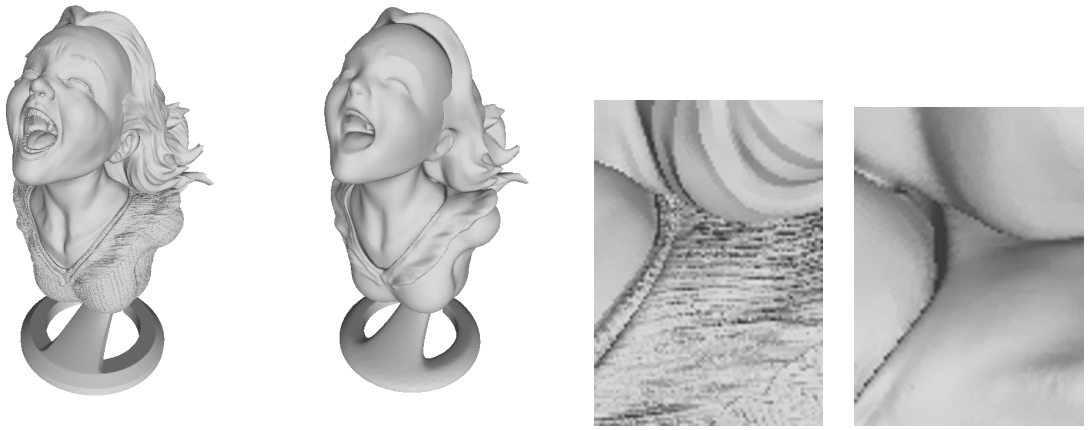
Dans ce paragraphe, nous évaluons la méthode d'estimation proposée sur des données de synthèse, puis sur des données réelles. Dans un premier temps, nous effectuons une comparaison quantitative de nos résultats avec deux méthodes utilisant une seule image : la décomposition *cartoon* + texture de [97] et la décomposition intrinsèque de [52].

4.6.1 Évaluation quantitative sur des données de synthèse

Nous commençons par évaluer notre méthode d'estimation de l'albédo en utilisant $n = 13$ images de taille 540×960 d'un objet dont la géométrie, parfaitement connue, est représentée sur la figure 4.3-a.

Deux scénarios sont testés. Sur la figure 4.4, une réflectance purement lambertienne, constante par morceaux, a été plaquée sur la surface de l'objet. L'éclairage est ici de type « *skydome* », c'est-à-dire quasiment diffus. Les effets d'ombrage sont donc très limités. Il est normal qu'une méthode d'estimation n'utilisant pas explicitement de modèle de réflectance, comme la décomposition *cartoon* + texture de [97], appliquée à chaque image, fournisse de bons résultats. De plus, l'albédo étant parfaitement constant par morceaux, la décomposition intrinsèque de l'image fondée sur un a priori de parcimonie, comme celle de [52], appliquée à chaque image, fonctionne également très bien.

Sur la figure 4.5, une réflectance plus complexe (non uniforme) est plaquée sur la chemise. La chevelure comporte des reflets brillants et l'éclairage diffus est remplacé par une seule source lumineuse étendue, ce qui provoque des effets d'ombrage plus importants. Il devient plus difficile de retirer l'ombrage sans un modèle explicite de réflectance, comme cela est le cas de l'approche *cartoon* + texture, tandis que la décomposition intrinsèque se révèle peu robuste aux reflets brillants.



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

FIGURE 4.3 – (a) Surface 3D utilisée dans nos tests. (b) La même surface, après lissage, permet d'évaluer l'impact d'une connaissance imprécise du relief. (c)-(d) Zooms sur une zone proche du cou de (a) et (b), respectivement.

Dans chaque cas, les méthodes [97] et [52] auxquelles nous nous comparons sont appliquées indépendamment à chacune des $n = 13$ images. De fait, les fonctions d'albédo, qui sont estimées séparément, n'ont aucune chance d'être cohérentes, ce qui peut se révéler problématique si ces fonctions sont destinées à être plaquées pour simuler, par exemple, le rééclairage de la surface. En revanche, notre approche estime de manière simultanée et cohérente les n fonctions d'albédo. Comme nous disposons de la vérité terrain de l'albédo, nous pouvons évaluer ces résultats en calculant la RMSE de chaque méthode sur l'ensemble des $n = 13$ images. Les valeurs obtenues sont rassemblées dans la table 4.1.

	Canal	*Cartoon* + texture [97]	Décomposition intrinsèque [52]	Nos résultats
Scénario 1 (cf. figure 4.4)	R	0,62	0,26	**0,07**
	V	0,23	0,14	**0,04**
	B	0,38	0,24	**0,07**
Scénario 2 (cf. figure 4.5)	R	0,60	0,29	**0,22**
	V	0,32	0,22	**0,13**
	B	0,24	0,21	**0,12**

TABLE 4.1 – RMSE sur l'albédo estimé (valeurs comprises entre 0 et 1), dans chaque canal, sur l'ensemble des images. Pour chacun des deux scénarios testés, notre approche surpasse les méthodes fonctionnant image par image.

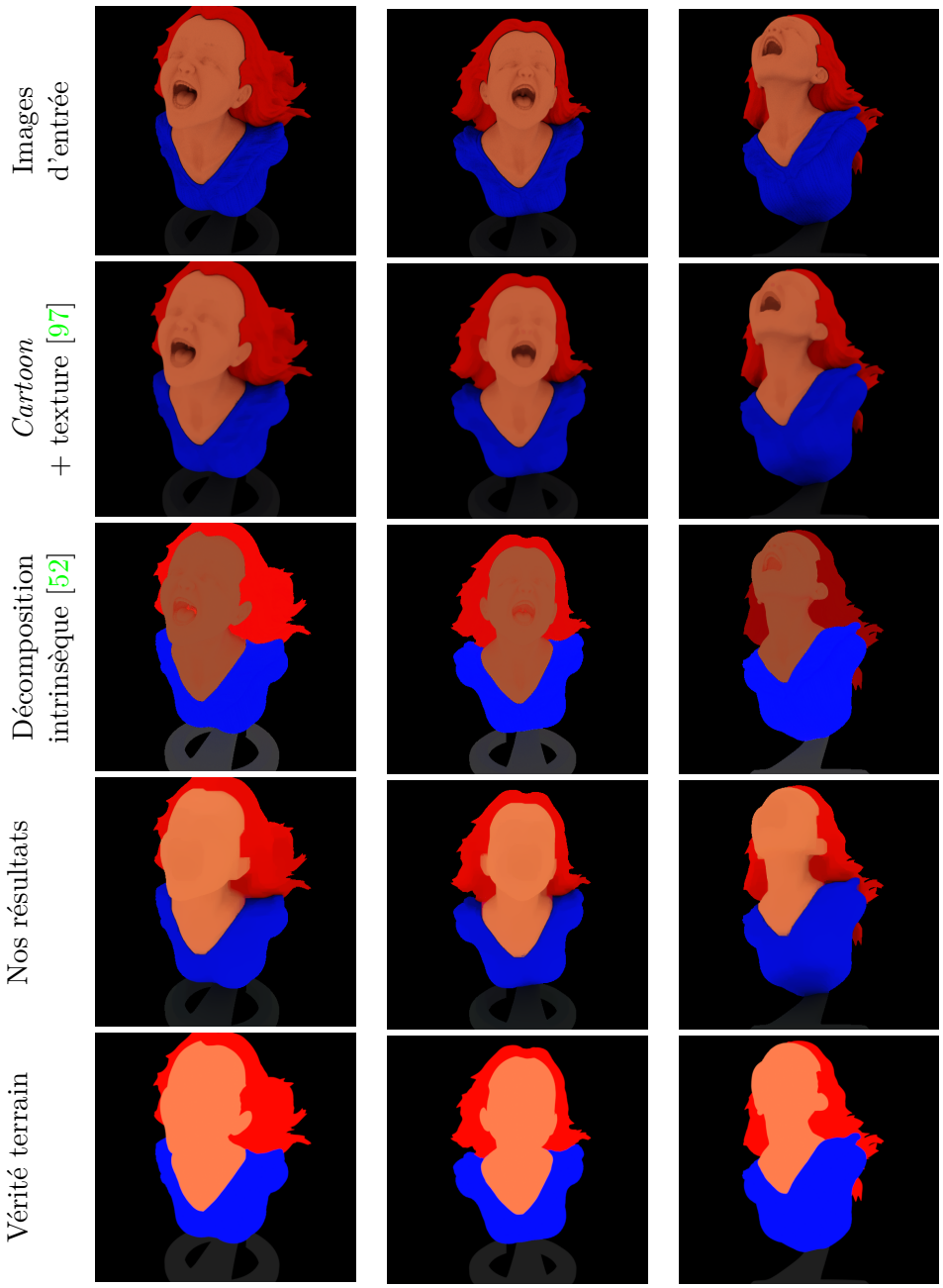


FIGURE 4.4 – Première ligne : trois (parmi $n = 13$) vues de l'objet de la figure 4.3-a, d'albédo constant par morceaux, éclairé par une source lumineuse de type « *skydome* ». Deuxième ligne : estimation de l'albédo par décomposition *cartoon* + texture [97]. Troisième ligne : estimation de l'albédo par décomposition intrinsèque [52]. Quatrième ligne : estimation de l'albédo par l'approche proposée ($\gamma = 8$; $\varrho = 1000$). Cinquième ligne : vérité terrain.

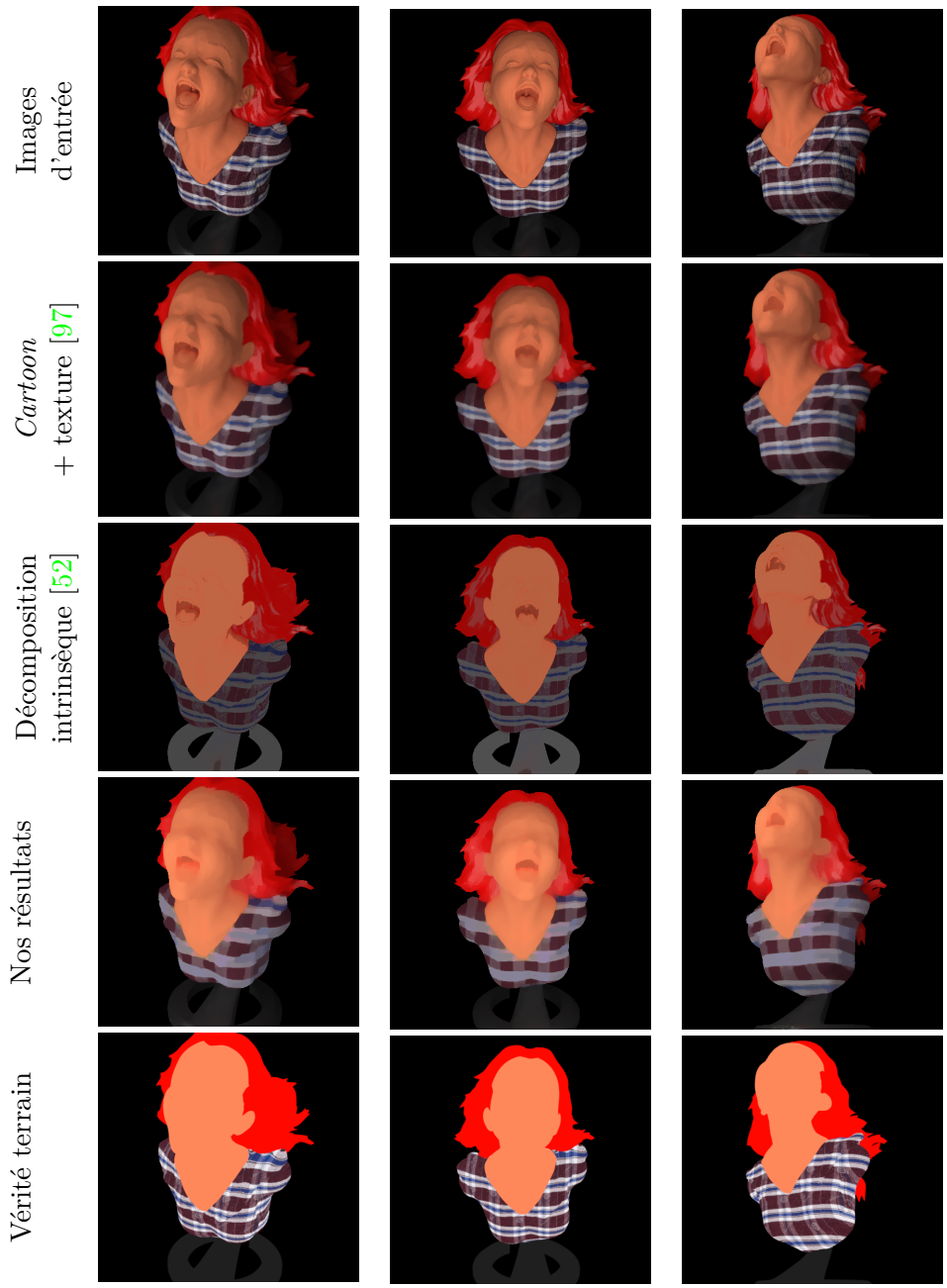


FIGURE 4.5 – Première ligne : trois (parmi $n = 13$) vues de l'objet de la figure 4.3-a, de réflectance plus complexe, éclairé par une source lumineuse étendue. Deuxième ligne : estimation de l'albédo par décomposition *cartoon* + texture [97]. Troisième ligne : estimation de l'albédo par décomposition intrinsèque [52]. Quatrième ligne : estimation de l'albédo par l'approche proposée ($\gamma = 2,5$; $\varrho = 1000$). Cinquième ligne : vérité terrain.

Les résultats qualitatifs des figures 4.4 et 4.5, ainsi que les évaluations quantitatives regroupées dans la table 4.1, nous permettent de faire les observations suivantes :

- L'utilisation d'un modèle explicite de formation de l'image permet de dépasser la décomposition *cartoon* + texture. En effet, la composante *cartoon*, sur la deuxième ligne de la figure 4.5, est beaucoup moins uniforme que l'albédo estimé par les deux autres méthodes (cf. les troisième et quatrième lignes de la figure 4.5). L'ombrage est flouté, mais pas réellement éliminé. Ce défaut pourrait être gommé en augmentant le poids du terme de régularisation, mais cela ferait perdre les détails de certaines parties telles que la chemise, dans l'exemple de la figure 4.5.

- Dans un contexte multi-vues, l'estimation simultanée des fonctions d'albédo permet de rendre ces fonctions cohérentes et de rendre l'estimation robuste en présence de reflets brillants. Lorsque les fonctions d'albédo sont estimées séparément, des incohérences apparaissent, ce qui explique les valeurs élevées de la RMSE dans la table 4.1. Ce défaut est clairement visible au niveau des cheveux, sur la troisième ligne de la figure 4.4. En comparaison, nos résultats confirment l'intuition initiale selon laquelle l'estimation de l'albédo doit bénéficier doublement du contexte multi-vues, qui permet non seulement d'estimer une géométrie grossière de la surface, mais également de contraindre l'albédo en un point de la surface à être identique dans toutes les images où il est visible. De plus, puisque les reflets brillants « se déplacent » lorsque le point de vue change, ces derniers n'apparaissent pas aux mêmes endroits de la surface dans les différentes images. La prise en compte de données multi-vues permet donc d'accroître la robustesse aux reflets brillants. Cela se confirme sur les résultats de la figure 4.5 : au niveau de la chevelure, le résultat de la quatrième ligne est bien moins dégradé par les reflets brillants que les résultats des deux autres méthodes.

- Un a priori de parcimonie de l'albédo doit être préféré à la variation totale. Comme nous utilisons un terme de régularisation fondé sur la variation totale, qui favorise les fonctions d'albédo constantes par morceaux, les très bons résultats de la figure 4.4 étaient prévisibles. Il reste néanmoins une trace de pénombre dans l'albédo estimé, en particulier au niveau du cou. Puisque nous connaissons la géométrie de l'objet, il semble que nous pourrions corriger cet effet de pénombre, mais cela nécessiterait de connaître également l'éclairage, ce qui n'est pas le cas avec un éclairage naturel. De plus, nous devrions tenir compte non seulement de l'éclairage primaire, mais également des réflexions de la lumière sur les différents éléments de la scène, comme le font les moteurs de rendu pour simuler une image. À titre de comparaison, l'approche fondée sur la parcimonie [52] élimine la pénombre sans pour autant qu'il soit nécessaire de modéliser les réflexions secondaires. Cette méthode devrait également permettre de mieux gommer l'ombrage du visage, dans un cas comme celui de la figure 4.5, sans dégrader les structures fines de la chemise autant que la variation totale. Pour résumer, la relative simplicité de notre solution numérique, qui vient de ce que l'a priori de Potts a été remplacé par la variation totale (cf. paragraphe 4.4.5), a un prix. Une perspective intéressante serait d'élaborer une stratégie numérique permettant de revenir au problème initial (4.24), bien que celui-ci soit non différentiable et non convexe.

4.6.2 Influence d'une imprécision sur le relief estimé

Dans les expériences précédentes, la géométrie était parfaitement connue, ce qui n'est pas le cas en pratique, à cause des inévitables erreurs commises lors de la reconstruction 3D par SfM/MVS. C'est pourquoi il est nécessaire d'évaluer le comportement de notre approche lorsque le relief estimé est imprécis. Pour le prochain test, nous utilisons la surface de la figure 4.3-b (dont un zoom est représenté sur la figure 4.3-d), qui a été obtenue en lissant la surface 3D initiale de la figure 4.3-a (dont un zoom est représenté sur la figure 4.3-c), grâce à un outil du logiciel `meshlab`. Au vu des résultats de la figure 4.6, il semble que notre méthode soit robuste aux imprécisions du relief de l'objet. Cela prouve qu'elle est effectivement adaptée aux cas d'usage auxquels nous la destinons.

FIGURE 4.6 – Mêmes tests que ceux de la figure 4.5, avec une version imprécise du relief (représenté sur les figures 4.3-b et 4.3-d), pour $\gamma = 2,5$ et $\varrho = 1000$. Les résultats sont qualitativement similaires à ceux de la figure 4.5, obtenus avec le relief exact. La RMSE évaluée dans chaque canal RVB vaut, respectivement : 0,24 ; 0,14 ; 0,13. Ces valeurs sont légèrement supérieures aux valeurs obtenues avec le relief exact (cf. table 4.1).

Sur la figure 4.7, nous évaluons qualitativement notre méthode, appliquée à la sortie d'un pipeline de SfM/MVS sur un jeu de données réelles. Nous disposons donc d'une estimation des poses des caméras et d'une géométrie grossière de la scène. Ce test confirme que la méthode proposée est effectivement robuste aux imprécisions sur le relief estimé. Dans l'ensemble, les reflets brillants sont correctement éliminés, et les fonctions d'albédo présentent bien un aspect *cartoon*. Cependant, l'albédo est sous-estimé sur les ailes du nez et au niveau du menton. En effet, puisque l'éclairage est fixe, ces parties de la surface se situent dans chaque image à l'intérieur de l'ombre propre. Une première solution à ce défaut consiste à augmenter le poids du terme de régularisation, au risque de perdre certains détails. Une autre solution est de contrôler davantage l'éclairage de la scène, afin qu'aucun point de la surface ne se situe à l'intérieur de l'ombre propre dans l'ensemble des vues. Nous reviendrons sur ce point dans le prochain paragraphe.

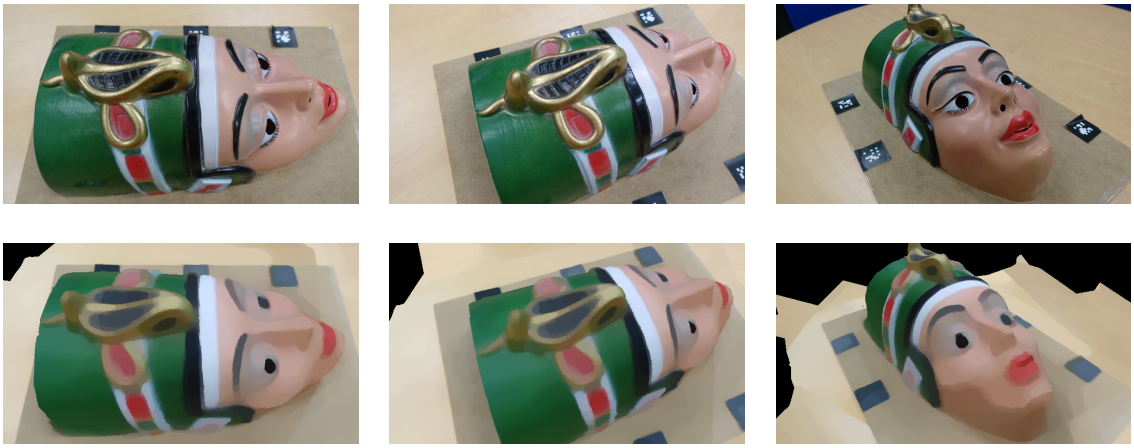


FIGURE 4.7 – Test sur un jeu de données réelles. Première ligne : trois des $n = 8$ vues de la scène. Deuxième ligne : fonctions d'albédo estimées par l'approche proposée ($\gamma = 2$; $\varrho = 1000$). La géométrie et les poses des caméras sont estimées par un pipeline SfM/MVS.

4.6.3 Influence des hyper-paramètres

Dans les tests précédents, nous avons choisi arbitrairement les valeurs des hyper-paramètres γ et ϱ, de manière à obtenir les « meilleurs » résultats. Bien entendu, il est nécessaire de discuter un tel réglage, qui peut se révéler fastidieux.

Afin d'illustrer l'influence de ces paramètres, intéressons-nous d'abord à ce qui se passe en l'absence de termes de régularisation et de cohérence inter-vues, c'est-à-dire si $\gamma = \varrho = 0$ dans (4.29). Dans ce cas, seul le terme photométrique est optimisé, ce qui revient à effectuer l'estimation par la recherche du maximum de vraisemblance. Dans le cas où l'éclairage est constant, le problème devient dégénéré. L'éclairage estimé est diffus, c'est-à-dire de la forme (4.10), et les fonctions d'albédo sont assimilées aux images. L'éclairage est donc « absorbé » par les fonctions d'albédo, ce que nous essayons précisément d'éviter.

Pour éviter ce phénomène, le terme de lissage doit être activé en imposant $\gamma > 0$. Si nous imposons toujours $\varrho = 0$, alors le problème variationnel (4.29) se réduit à un ensemble de n problèmes indépendants de restauration d'images, qui sont similaires au problème du débruitage ℓ^1-TV, à ceci près que le terme d'attache aux données est physiquement réaliste. Toutefois, comme le terme photométrique est invariant à une transformation du type $(\rho_i, \boldsymbol{\sigma}_i) := (\chi_i \rho_i, \boldsymbol{\sigma}_i / \chi_i)$, $\chi_i > 0$, chaque albédo ρ_i n'est estimé qu'à un facteur d'échelle près, ce qui interdit la cohérence des n fonctions d'albédo, comme cela est le cas des méthodes mono-vue d'estimation de l'albédo que nous avons également testées.

Ce problème est résolu si nous activons le terme de cohérence inter-vues, c'est-à-dire si nous donnons à ϱ une valeur strictement positive. Dans ce cas, l'ambiguïté précédente $\{\rho_i, \boldsymbol{\sigma}_i\}_i := \{\chi \rho_i, \boldsymbol{\sigma}_i / \chi\}$, $\chi > 0$, existe toujours, mais cette ambiguïté devient globale, c'est-à-dire indépendante de i. Pour la résoudre, il suffit par exemple de fixer arbitrairement l'albédo en un point, ou de normaliser les valeurs estimées.

Il est donc nécessaire de faire en sorte que les deux hyper-paramètres γ et ϱ soient strictement positifs. Le choix de la valeur de ϱ n'est pas réellement critique. En effet, le terme de cohérence inter-vues, qui est contrôlé par ce paramètre, provient de la relaxation d'une contrainte « dure » (pour s'en convaincre, il suffit de comparer les problèmes (4.24) et (4.29)). Il faut donc choisir ϱ « suffisamment grand », de manière à ce que le terme de régularisation du problème (4.29) s'approche de la contrainte dure du problème (4.24). Dans tous nos tests, nous n'avons rencontré aucun problème particulier avec $\varrho = 1000$. Bien entendu, si les correspondances entre images n'ont pas été évaluées correctement par le SfM, cette valeur doit être réduite, mais les méthodes de SfM telles que celle de la suite logicielle AliceVision [7] sont aujourd'hui suffisamment matures pour pouvoir fournir des correspondances fiables.

Le choix de la valeur de γ semble plus critique. Ceci est mis en évidence par les courbes de la figure 4.8 qui montrent, canal par canal, comment évolue la RMSE (calculée à la convergence sur le même jeu de données que celui de la figure 4.5) en fonction de γ. Ces courbes montrent que la valeur « optimale » de γ est difficile à définir : dans cet exemple, une valeur élevée de γ permet de réduire considérablement la RMSE sur le visage et la chevelure, qui sont à dominante rouge, en leur donnant un aspect uniforme (cf. la dernière ligne de la figure 4.9). Cependant, une valeur beaucoup plus faible de γ est requise pour préserver les détails fins de la chemise, qui comporte davantage de vert et de bleu (cf. la première ligne de la figure 4.9).

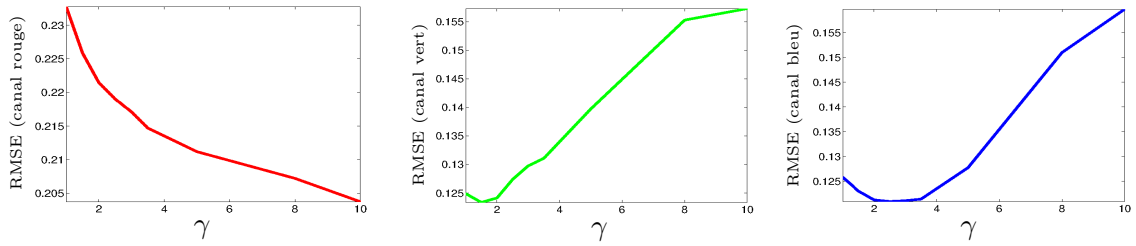


FIGURE 4.8 – Étude quantitative de l'influence du paramètre γ menée sur le même jeu de données que celui de la figure 4.5, avec $\varrho = 1000$.

Une situation où le réglage de γ est beaucoup plus facile est celle où l'éclairage n'est pas fixe, mais varie fortement d'une image à l'autre. Comme nous l'avions évoqué dans le paragraphe 4.1.1, le problème de l'estimation jointe de l'albédo et de l'éclairage est alors sur-contraint, ce qui rend la régularisation superflue. Cette nouvelle approche dépasse le cadre de nos travaux, mais il semble que ce soit une piste prometteuse. D'ailleurs, la combinaison entre reconstruction 3D multi-vues et stéréophotométrie fait actuellement l'objet de recherches actives dans la communauté de la vision par ordinateur [122, 129].

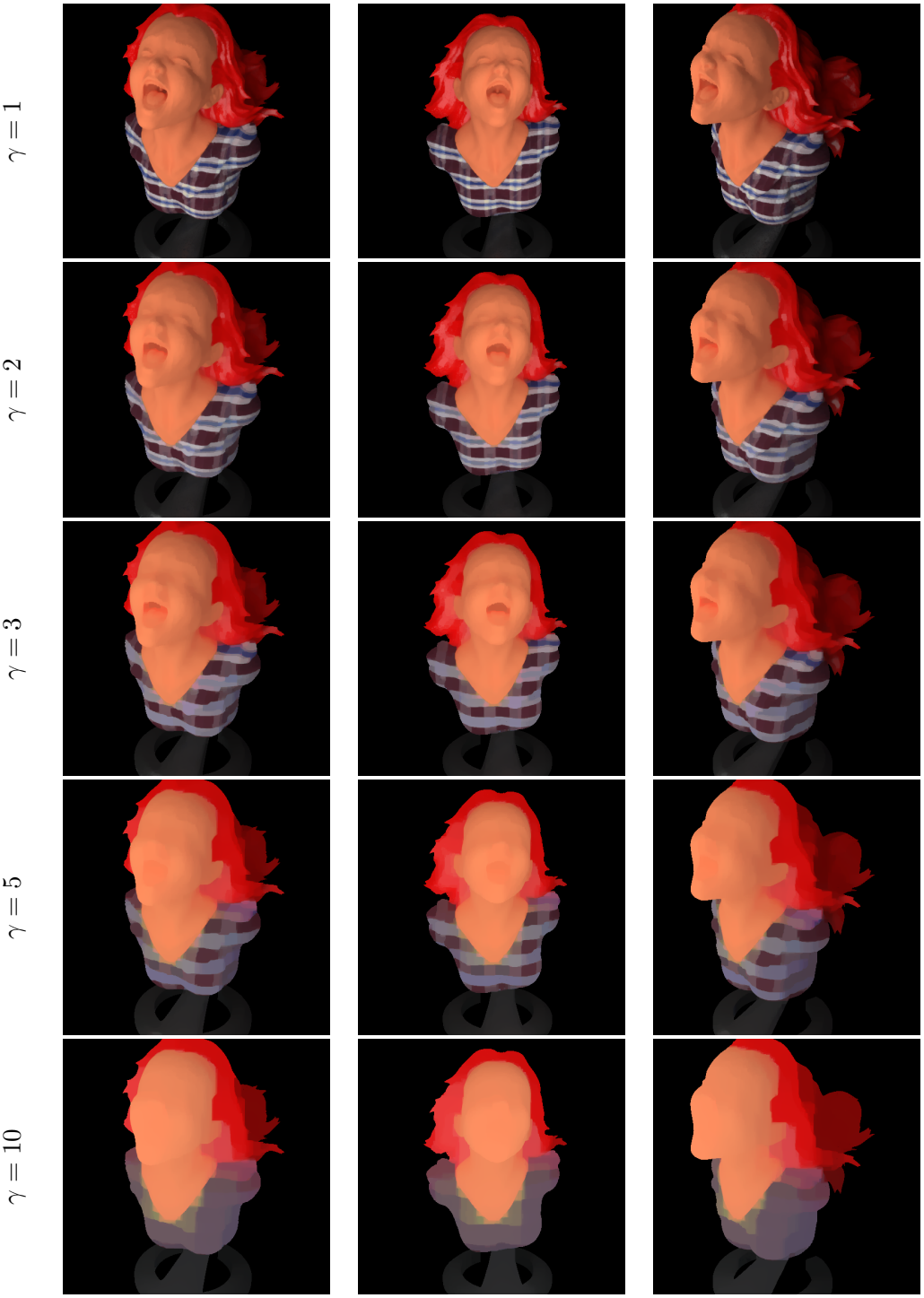


FIGURE 4.9 – Influence qualitative du paramètre γ sur le même jeu de données que celui de la figure 4.5, avec $\varrho = 1000$.

Conclusion

Dans ce chapitre, nous avons proposé une nouvelle approche variationnelle permettant d'estimer l'albédo et l'éclairage à partir d'images prises sous différents angles. Nous avons montré pourquoi il était préférable de paramétrer l'albédo par les coordonnées image. L'utilisation d'un raisonnement bayésien nous a permis d'élaborer un modèle variationnel comportant différents termes : un terme d'attache aux données en norme ℓ^1 traduisant la loi de Lambert, un terme de régularisation de type Potts et une contrainte garantissant la cohérence inter-vues. La résolution numérique proposée utilise un algorithme de majoration-minimisation alternée sur une relaxation bi-convexe du problème original. Les résultats expérimentaux permettent de valider la robustesse de l'estimation à la présence de reflets brillants, et la cohérence entre les différentes fonctions d'albédo estimées.

L'analyse critique des résultats nous a quand même permis de mettre en évidence plusieurs limites de cette nouvelle approche, qui constituent autant de futurs axes de recherche. D'une part, nous pourrions éviter la relaxation du terme de régularisation de Potts, afin de mieux garantir l'aspect « constant par morceaux » des fonctions de réflectance estimées. D'autre part, si nous avons choisi de paramétrer l'albédo par les coordonnées image pour des raisons numériques, il semble quand même plus naturel de définir ce paramétrage sur la surface elle-même, au risque de rendre le problème nettement plus complexe. Enfin, le moyen le plus simple pour estimer l'albédo et l'éclairage sans ambigüité est de modifier l'éclairage entre les différentes prises de vue. L'extension de nos travaux à la « stéréophotométrie multi-vues » semble effectivement très prometteuse.

Quoi qu'il en soit, la connaissance de l'albédo et de l'éclairage est une hypothèse requise par le *shape-from-shading*. L'estimation de ces deux entités à partir d'une connaissance grossière du relief permet donc d'étendre considérablement le champ d'utilisation des algorithmes de SfS. Il semble à présent naturel de combiner les différents éléments présentés dans ce mémoire afin d'estimer simultanément, à partir d'un ensemble d'images, les trois éléments constitutifs d'une scène que sont le relief, la réflectance et l'éclairage.

Conclusion

Résumé des contributions

Dans ce mémoire, nous avons présenté plusieurs contributions, théoriques ou numériques, qui visent à fusionner les techniques géométriques et les techniques photométriques de reconstruction 3D.

Dans le chapitre 1, nous avons décrit le pipeline classique de reconstruction 3D par photogrammétrie. S'il permet en général d'obtenir de bons résultats, ce pipeline présente certaines limitations, de par l'aspect purement géométrique des techniques qu'il utilise, à savoir le *structure-from-motion* (SfM) et la stéréoscopie multi-vues (MVS). La reconstruction 3D d'une scène peu ou pas texturée, pour laquelle le MVS ne permet pas de restituer les détails du relief, apparaît comme l'une des principales faiblesses de ce pipeline.

Le *shape-from-shading* (SfS), qui est une figure emblématique des techniques photométriques, est parfaitement adapté à la reconstruction 3D de scènes non texturées. Cependant, cette technique nécessite de bien contrôler l'éclairage, ce qui limite son champ d'application à des photographies prises en laboratoire. Dans le chapitre 2, nous avons montré comment mettre en œuvre le SfS dans des conditions opératoires réalistes, c'est-à-dire sous un éclairage naturel. Si la résolution proposée semble numériquement stable, les ambiguïtés inhérentes au problème sont levées grâce à un terme de régularisation, qui peut prendre la forme d'un lissage par « minimisation de la surface totale », ou d'un a priori sur la profondeur. Ce dernier terme de régularisation nous a permis de concevoir un premier algorithme de reconstruction 3D combinant les techniques SfM, MVS et SfS.

Cette approche n'est toutefois pas optimale, à cause de son caractère séquentiel. En effet, le contexte multi-vues sert à obtenir une approximation grossière de la fonction de profondeur, que l'ombrage sert à affiner a posteriori. Dans le chapitre 3, nous avons montré comment opérer une réelle fusion entre les techniques MVS et SfS. La stéréoscopie multi-vues ainsi modifiée profite des informations apportées par l'ombrage, ce qui lui permet d'effectuer la reconstruction 3D d'objets peu ou pas texturés.

Néanmoins, la mise en œuvre du SfS ne va pas de soi. En effet, l'algorithme du chapitre 3 n'est utilisable que sous les hypothèses classiques du SfS, à savoir que la réflectance de la surface et l'éclairage doivent être connus. Dans le chapitre 4, nous avons montré comment relâcher ces hypothèses. Les fonctions de profondeur estimées par MVS, qui sont grossières mais denses, nous permettent d'estimer l'albédo de la surface, supposée lambertienne, ainsi que l'éclairage. Grâce à ces estimations, le champ d'application de la technique de reconstruction 3D proposée est notablement élargi.

Vers un pipeline complet de scannage 3D

Les différentes contributions présentées dans ce mémoire semblent pouvoir contribuer à la création d'un pipeline complet de scannage 3D. À partir des fonctions de profondeur estimées par MVS, nous pouvons estimer une fonction d'albédo et un éclairage par image, pour ensuite affiner la profondeur par SfS. La réelle fusion de ces deux étapes semble plus efficace qu'une simple approche séquentielle.

En effet, la première étape (3.14) de l'algorithme de reconstruction 3D proposé dans le chapitre 3 fournit une estimation grossière de la géométrie par MVS. Cela nous permet, grâce à l'algorithme présenté dans le chapitre 4, d'estimer l'albédo de la surface et l'éclairage. Nous sommes alors en mesure de mettre en œuvre les étapes (3.15) et (3.16) de l'algorithme de reconstruction 3D du chapitre 3, qui visent à affiner la profondeur par SfS. Cette séquence est ensuite répétée jusqu'à la convergence.

Nous pressentons donc que l'approche résumée par l'algorithme 2 permet d'estimer à la fois le relief et la réflectance d'une scène 3D, ainsi que son éclairage, à partir d'un ensemble d'images et de la connaissance des paramètres intrinsèques et des poses de la caméra. Afin de tester cette approche, nous avons plaqué une texture non uniforme sur le modèle 3D de la figure 3.3. Les images qui en résultent, ainsi que des résultats préliminaires, sont représentés sur les figures 5.1 et 5.3.

Ces premiers résultats suffisent à valider le bien-fondé de l'ensemble de notre approche. En effet, la fonction de profondeur que nous obtenons est bien plus précise que celle de la suite logicielle AliceVision, même après une étape de post-traitement (cf. la deuxième ligne de la figure 5.1). Notons tout de même que le nombre d'images utilisées ($n = 6$) est faible, en comparaison de ce qui est généralement requis par le MVS. Effectivement, les résultats s'améliorent lorsque le nombre d'images utilisées augmente ($n = 12$, cf. figure 5.2).

Algorithme 2 : Algorithme générique de reconstruction 3D (albédo non uniforme).

Entrées : images, poses des caméras

Sorties : pour chaque image : fonction d'albédo, fonction de profondeur, éclairage

répéter

- Estimation grossière de la profondeur (3.14)
- Estimation de l'albédo (4.30)
- Estimation de l'éclairage (4.31)
- Affinage de la fonction de profondeur (3.15) + (3.16)

jusqu'à *convergence*;

Le maillage obtenu par fusion des fonctions de profondeur est débarrassé des aberrations du relief. Les défauts résiduels du relief disparaissent après plaquage de la texture, comme le montre la figure 5.2. Le modèle 3D obtenu ne restitue cependant que très grossièrement le relief. Les détails apparents sur le modèle 3D complet proviennent en réalité d'une illusion de type « trompe-l'œil », puisque les images d'origine tiennent lieu de texture.

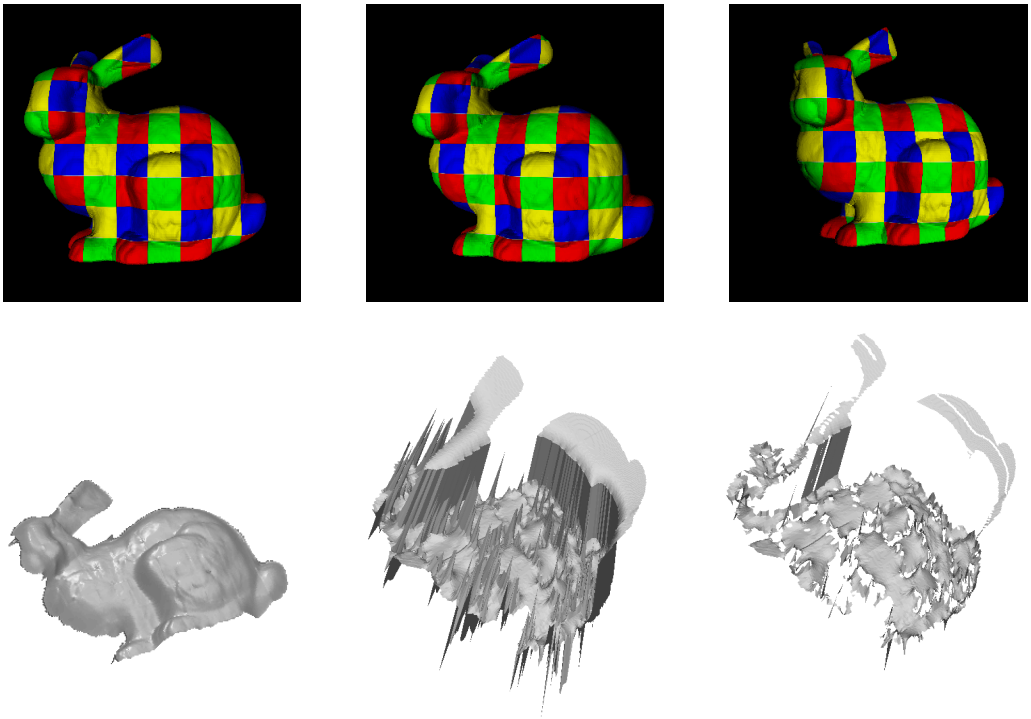


FIGURE 5.1 – Estimation du relief à partir d'un ensemble de $n = 6$ images de synthèse. Première ligne : trois des $n = 6$ images utilisées. Deuxième ligne : à gauche, fonction de profondeur obtenue par notre approche ; au centre et à droite, fonctions de profondeur obtenues avec la suite logicielle AliceVision, après une étape de post-traitement, respectivement avant et après filtrage.

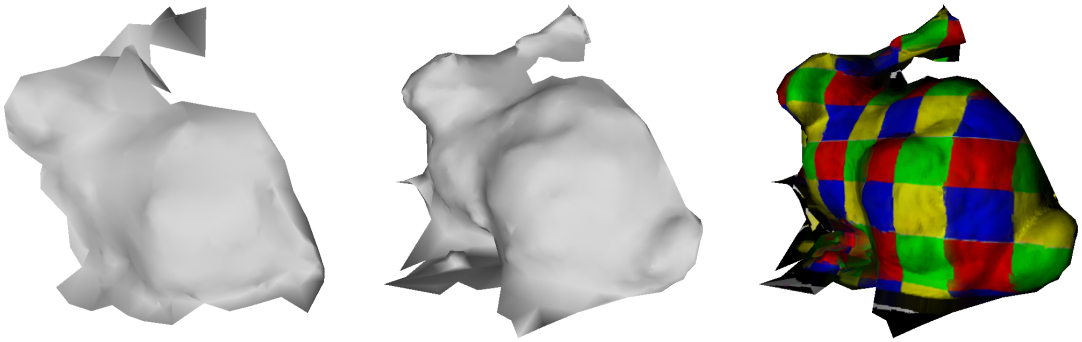


FIGURE 5.2 – À gauche : maillage 3D fourni par AliceVision à partir de $n = 6$ images. Au centre et à droite : maillage 3D et modèle 3D complet fournis par AliceVision avec $n = 12$ images. Même s'il est globalement restitué, le relief est excessivement lissé. L'ombrage présent dans la texture donne l'illusion que les détails géométriques ont été reconstruits.

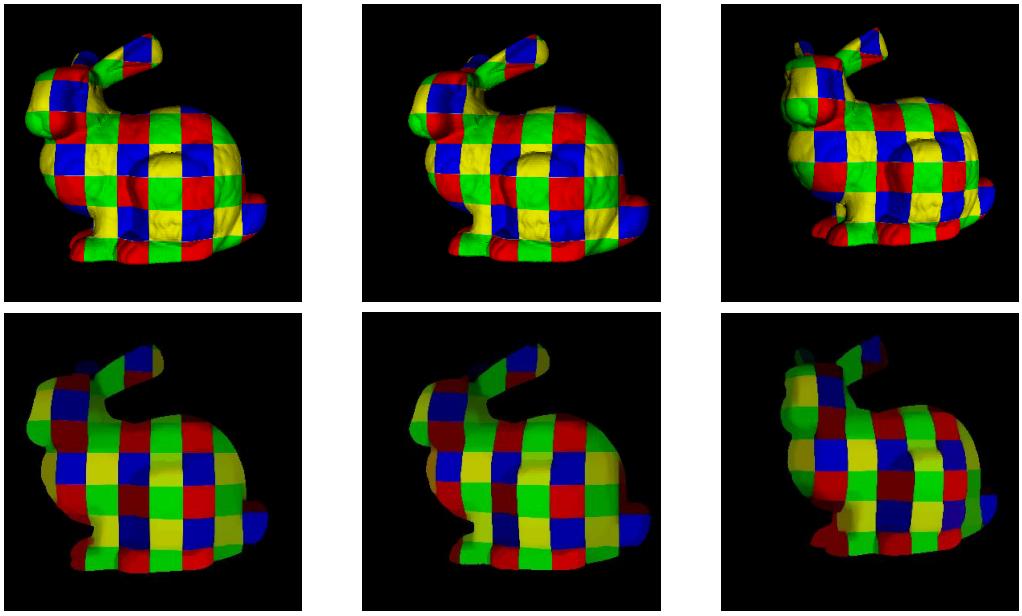


FIGURE 5.3 – Estimation de l'albédo à partir d'un ensemble de $n = 6$ images de synthèse. Première ligne : trois des $n = 6$ images utilisées. Deuxième ligne : fonctions d'albédo estimées par notre approche, correspondant aux différentes vues.

A contrario, les fonctions d'albédo estimées par notre approche (cf. figure 5.3) ne contiennent quasiment plus d'ombrage. En les plaquant sur le relief estimé, au lieu de plaquer les images, le rééclairage de la scène n'en sera que plus réaliste. Les solutions actuelles de rééclairage de scènes 3D à partir de photographies se limitent à de petits objets et reposent sur de nombreuses connaissances a priori [167], ou bien nécessitent un environnement très contrôlé [143, chapitre 11]. L'approche que nous proposons est intermédiaire : elle se fonde sur un modèle physique, en l'occurrence celui de la réflexion de la lumière par une surface lambertienne, tout en s'affranchissant des contraintes imposant aux photographies d'être prises en laboratoire.

Il est quand même flagrant, sur les résultats de la figure 5.3, que les différents carreaux d'une même couleur ne sont pas restitués sous une même apparence. La séparation entre albédo et ombrage n'est pas parfaite, à cause de la configuration de la texture de type « Arlequin ». En effet, l'a priori utilisé dans le modèle (4.20) ne permet d'uniformiser l'albédo que sur des zones connexes de couleur uniforme, ce qui est le cas de chaque carreau du motif d'Arlequin, mais pas d'un ensemble de plusieurs carreaux de même couleur. Cela explique pourquoi le rouge des carreaux situés sur la tête du lapin semble beaucoup plus sombre que celui des carreaux situés sur son arrière-train.

Stéréophotométrie multi-vues

Pour améliorer l'estimation de l'albédo, il est nécessaire de s'inspirer de la stéréophotométrie, c'est-à-dire de faire varier l'éclairage entre les prises de vue. Nous avons déjà décrit cette technique de reconstruction 3D photométrique, qui présente la particularité d'être bien posée. Cependant, les conditions opératoires usuelles de la post-production, qui ont défini le cadre des recherches menées dans cette thèse, ne permettent pas de réellement contrôler l'éclairage. Pour cette raison, la piste de la stéréophotométrie a été délaissée, mais elle constitue malgré tout une de nos principales perspectives de recherche.

À titre d'exemple, revenons sur l'algorithme présenté dans le chapitre 4, qui estime l'albédo et l'éclairage d'une scène 3D à partir d'une géométrie grossière du relief. L'énergie à minimiser (4.24) comporte un terme d'attache aux données traduisant la loi de Lambert, un terme de régularisation par variation totale et un terme de cohérence inter-vues. La figure 5.4 présente les résultats obtenus dans le cas où l'éclairage varie d'une image à l'autre. En comparaison des résultats de la première ligne de la figure 4.9, la texture rayée de la chemise est correctement restituée, et l'ombrage est quasiment éliminé du visage, bien que le poids $\gamma = 1$ du terme de régularisation soit très faible [1]. Ce test montre que sans a priori sur l'éclairage, le seul moyen d'éviter le recours à un a priori sur la réflectance, donc le réglage d'hyper-paramètres, est de faire varier l'éclairage entre les prises de vue.

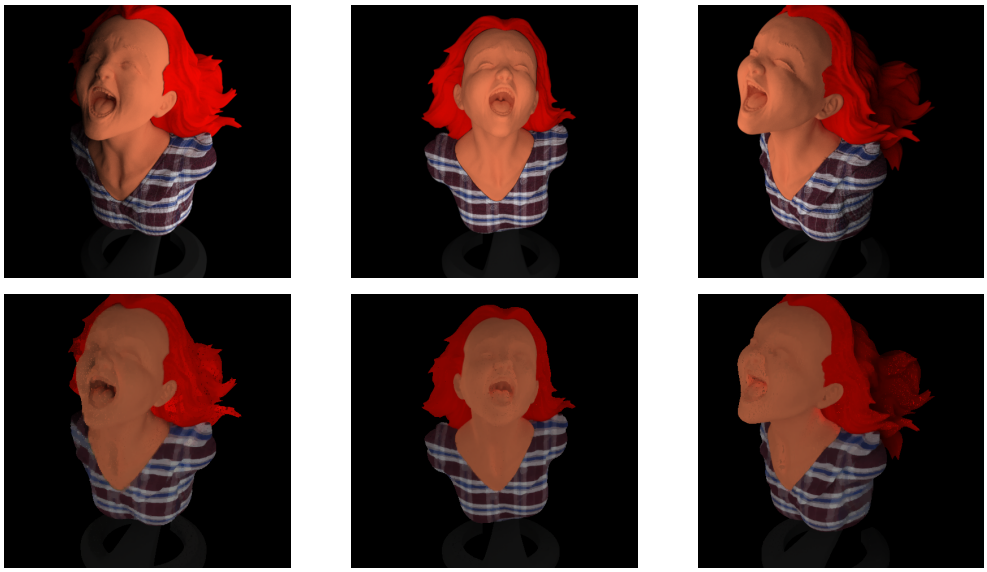


FIGURE 5.4 – Première ligne : trois (parmi $n = 13$) vues de l'objet de la figure 4.3-a calculées sous des éclairages différents. Deuxième ligne : fonctions de réflectance estimées par notre approche. En comparaison des résultats de la première ligne de la figure 4.9, obtenus avec les mêmes poids ($\gamma = 1$ et $\varrho = 1000$), mais avec un éclairage fixe, la texture rayée de la chemise est correctement restituée, et l'ombrage est quasiment éliminé du visage.

1. Si certains pixels ne possèdent pas de correspondant, le problème est mal posé. Le terme de régularisation est donc nécessaire, ce qui signifie que nous ne devons pas annuler son poids.

Matériaux non lambertiens

Dans ce mémoire, nous nous sommes restreints aux surfaces lambertiennes, pour lesquelles la réflectance est entièrement caractérisée par l'albédo. Or, il pourrait être profitable de prendre en compte des réflectances plus complexes. Cela permettrait non seulement de simuler le rééclairage d'une scène 3D de manière plus photo-réaliste, mais également d'améliorer la reconstruction 3D d'objets non parfaitement lambertiens. En effet, les taches brillantes sont porteuses d'informations sur la géométrie de la scène, puisque celle-ci peut être assimilée localement à un miroir. L'estimation de BRDF complexes [11], qui a fait l'objet de travaux récents [38, 131], constitue effectivement une autre de nos perspectives.

L'estimation de la réflectance dans une scène complexe soulève la problématique de la segmentation d'une image en différents matériaux. Cette segmentation peut être menée en utilisant une seule image, comme le font des travaux récents [5], dont certains recourent à l'apprentissage profond [33], les différentes segmentations étant fusionnées a posteriori. Une autre solution consiste à segmenter l'ensemble des images simultanément [40]. Une fois le matériau segmenté, le contexte multi-vues fournit plusieurs échantillons de sa réflectance. L'utilisation d'une base de données permet ensuite d'identifier ce matériau dans de nouvelles images, c'est-à-dire de résoudre un problème de classification [19], plutôt que d'utiliser un modèle de réflectance potentiellement complexe. Cela pourrait nous autoriser à élargir cette étude aux matériaux transparents et transluminescents, que nous avons délibérement mis de côté dans ce mémoire.

Bibliographie

[1] E. H. Adelson et A. P. Pentland : *Perception as Bayesian inference*, chapitre The perception of shading and reflectance, pages 409–423. Cambridge University Press, 1996. 3

[2] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz et R. Szeliski : Building Rome in a Day. *In Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 72–79, 2009. 10

[3] M. Aggarwal, H. Hua et N. Ahuja : On cosine-fourth and vignetting effects in real lenses. *In Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, volume 1, pages 472–479, 2001. 28

[4] A. H. Ahmed et A. A. Farag : Shape from shading under various imaging conditions. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007. 36

[5] Y. Aksoy, T.-H. Oh, S. Paris, M. Pollefeys et W. Matusik : Semantic soft segmentation. *ACM Transactions on Graphics*, 37(4):72:1–72:13, 2018. 98

[6] P. Alcantarilla, J. Nuevo et A. Bartoli : Fast explicit diffusion for accelerated features in nonlinear scale spaces. *In Proceedings of the British Machine Vision Conference (BMVC)*, 2013. 12

[7] AliceVision. https://alicevision.org/. 2, 21, 22, 23, 69, 90

[8] J.-F. Aujol, G. Aubert, L. Blanc-Féraud et A. Chambolle : Image decomposition into a bounded variation component and an oscillating component. *Journal of Mathematical Imaging and Vision (JMIV)*, 22:71–88, 2005. 73

[9] J.-F. Aujol, G. Gilboa, T. Chan et S. Osher : Structure-Texture Image Decomposition – Modeling, Algorithms, and Parameter Selection. *International Journal of Computer Vision (IJCV)*, 67(1):111–136, 2006. 73

[10] Autodesk: Arnold. https://www.arnoldrenderer.com/. 1

[11] D. Azinović, T.-M. Li, A. Kaplanyan et M. Niessner : Inverse path tracing for joint material and lighting estimation. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 98

[12] S Bakshi et Y.-H. Yang : Shape from shading for non-Lambertian surfaces. *In Proceedings of the International Conference on Image Processing (ICIP)*, volume 2, pages 130–134, 1994. 36

[13] J. T. Barron et J. Malik : Shape, illumination, and reflectance from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 37(8):1670–1687, 2015. 30, 38, 39, 45, 47, 48, 74

[14] A. Bartoli, Y. Gérard, F. Chadebecq, T. Collins et D. Pizarro : Shape-from-template. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 37(10):2099–2118, 2015. 2

[15] R. Basri et D. P. Jacobs : Lambertian reflectances and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 25(2):218–233, 2003. 37

[16] R. Basri, D. P. Jacobs et I. Kemelmacher : Photometric stereo with general, unknown lighting. *International Journal of Computer Vision (IJCV)*, 72(3):239–257, 2007. 34, 72

[17] H. Bay, T. Tuytelaars et L. Van Gool : Surf: Speeded up robust features. *In Proceedings of the European Conference on Computer Vision (ECCV)*, pages 404–417, 2006. 12

[18] P. N. Belhumeur, D. J. Kriegman et A. L. Yuille : The bas-relief ambiguity. *International Journal of Computer Vision (IJCV)*, 35(1):33–44, 1999. 30

[19] S. Bell, K. Bala et N. Snavely : Intrinsic images in the wild. *ACM Transactions on Graphics*, 33(4):159:1–159:12, 2014. 74, 98

[20] S. Belongie, J. Malik et J. Puzicha : Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, (4):509–522, 2002. 12

[21] A. Blake, A. Zisserman et G. Knowles : Surface descriptions from stereo and shading. *Image and Vision Computing (IVC)*, 3(4):183–191, 1985. 58

[22] A. Bousseau, S. Paris et F. Durand : User assisted intrinsic images. *ACM Transactions on Graphics*, 28(5):130:1–130:10, 2009. 74

[23] S. Boyd, N. Parikh, E. Chu, B. Peleato et J. Eckstein : Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011. 43, 44

[24] Y. Boykov et V. Kolmogorov : An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 26(9):1124–1137, 2004. 21

[25] M. Brady et A. L. Yuille : An extremum principle for shape from contour. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 6(3):288–301, 1984. 2

[26] M. Breuss, E. Cristiani, J.-D. Durou, M. Falcone et O. Vogel : Perspective shape from shading: Ambiguity analysis and numerical approximations. *SIAM Journal on Imaging Sciences*, 5(1):311–342, 2012. 35, 36

[27] D. H. Brown : Decentering distortion of lenses. *Journal of Photogrammetric Engineering and Remote Sensing*, 32(3):444–462, 1966. 6

[28] M. Brown et D. G Lowe : Unsupervised 3D object recognition and reconstruction in unordered datasets. *In Proceeedings of the Fifth International Conference on 3-D Digital Imaging and Modeling (3DIM'05)*, pages 56–63, 2005. 17

[29] M. Brown et D. G. Lowe : Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision (IJCV)*, 74(1):59–73, 2007. 12

[30] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha et P. Fua : Brief: Computing a local binary descriptor very fast. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 34(7):1281–1298, 2012. 13

[31] A. Chambolle : A uniqueness result in the theory of stereo vision: coupling shape from shading and binocular information allows unambiguous depth reconstruction. *Annales de l'IHP - Analyse non linéaire*, 11(1):1–16, 1994. 58

[32] Q. Chen et V. Koltun : A simple model for intrinsic image decomposition with depth cues. *In Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 241–248, 2013. 74

[33] M. Cimpoi, S. Maji et A. Vedaldi : Deep filter banks for texture recognition and segmentation. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3828–3836, 2015. 98

[34] F. Courteille, A. Crouzil, J.-D. Durou et P. Gurdjos : Shape from Shading for the Digitization of Curved Documents. *Machine Vision and Applications*, 18(5):301–316, 2007. 33

[35] E. Cristiani et M. Falcone : Fast semi-Lagrangian schemes for the Eikonal equation and applications. *SIAM Journal on Numerical Analysis*, 45(5):1979–2011, 2007. 33

[36] J. E. Cryer, P.-S. Tsai et M. Shah : Integration of shape from shading and stereo. *Pattern Recognition*, 28(7):1033–1043, 1995. 58

[37] P. Debevec et J. Malik : Recovering high dynamic range radiance maps from photographs. *In Proceedings of the Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 369–378, 1997. 28

[38] V. Deschaintre, M. Aittala, F. Durand, G. Drettakis et A. Bousseau : Single-Image SVBRDF Capture with a Rendering-Aware Deep Network. *ACM Transactions on Graphics*, 37(128), 2018. 98

[39] A. Dick, P. Torr et R. Cipolla : Automatic 3D modelling of architecture. *In Proceedings of the British Machine Vision Conference (BMVC)*, volume 1, pages 372–381, 2000. 1

[40] A. Djelouah, J.-S. Franco, E. Boyer, F. Le Clerc et P. Pérez : Sparse Multi-View Consistency for Object Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 37(9):1890–1903, 2015. 98

[41] J.-D. Durou : *Shape from shading - Éclairages, réflexions et perspectives*. Habilitation à diriger des recherches, Université Paul Sabatier, Toulouse, France, 2007. 25, 26, 27, 28, 33

[42] J.-D. Durou : Reconstruction 3D à partir des ombrages. *In Problèmes inverses en imagerie et en vision*, volume 2 de *Traité Signal et Image, IC2*, chapitre 10, pages 353–390. Lavoisier, 2009. 29

[43] J.-D. Durou, M. Falcone et M. Sagona : Numerical methods for shape-from-shading: A new survey with benchmarks. *Computer Vision and Image Understanding (CVIU)*, 109(1):22–43, 2008. 33, 45

[44] M. Falcone et M. Sagona : An algorithm for the global solution of the shape-from-shading model. *In Proceedings of the International Conference on Image Analysis and Processing (ICIAP)*, pages 596–603, 1997. 33

[45] R. T. Frankot et R. Chellappa : A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 10(4):439–451, 1988. 33

[46] D. Frolova, D. Simakov et R. Basri : Accuracy of spherical harmonic approximations for images of Lambertian objects under far and near lighting. *In Proceedings of the European Conference on Computer Vision (ECCV)*, pages 574–587, 2004. 37

[47] Y. Furukawa et C. Hernàndez : Multi-view stereo: A tutorial. *Foundations and Trends in Computer Graphics and Vision*, 9(1–2), 2013. 2, 19, 20, 59

[48] Y. Furukawa et J. Ponce : Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 32(8):1362–1376, 2010. 20

[49] E. Garces, A. Munoz, J. Lopez-Moreno et D. Gutierrez : Intrinsic Images by Clustering. *Computer Graphics Forum*, 31(4):1415–1424, 2012. 73

[50] P. Gargallo, E. Prados et P. Sturm : Minimizing the reprojection error in surface reconstruction from images. *In Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2007. 19

[51] P. Gargallo et P. Sturm : Bayesian 3D modeling from images using multiple depth maps. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 885–891, 2005. 56

[52] P. Gehler, C. Rother, M. Kiefel, L. Zhang et B. Schölkopf : Recovering Intrinsic Images with a Global Sparsity Prior on Reflectance. *In Advances in Neural Information Processing Systems*, pages 765–773, 2011. 73, 83, 84, 85, 86, 87

[53] J. Geng : Structured-light 3D surface imaging: a tutorial. *Advances in Optics and Photonics*, 3(2):128–160, 2011. 2

[54] R. Glowinski et A Marroco : Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. *Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique*, 9(R2):41–76, 1975. 43

[55] M. Goesele, B. Curless et S. M. Seitz : Multi-view stereo revisited. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 2402–2409, 2006. 20

[56] G. H. Golub et C. F. Van Loan : *Matrix Computations (quatrième édition)*. Johns Hopkins University Press, 2013. 75

[57] G. Graber, J. Balzer, S. Soatto et T. Pock : Efficient minimal-surface regularization of perspective depth maps in variational stereo. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 511–520, 2015. 43, 51, 59, 60

[58] M. D. Grossberg et S. K. Nayar : Modeling the space of camera response functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 26:1272–1282, 2004. 28

[59] B. Haefner, Y. Quéau, T. Möllenhoff et D. Cremers : Fight ill-posedness with ill-posedness: Single-shot variational depth super-resolution from shading. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 164–174, 2018. 38

[60] T. Haines et R. Wilson : Integrating stereo with shape-from-shading derived orientation information. *In Proceedings of the British Machine Vision Conference (BMVC)*, volume 2, pages 910–919, 2007. 58

[61] Y. Han, J.-Y. Lee et I. S. Kweon : High quality shape from a single RGB-D image under uncalibrated natural illumination. *In Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1617–1624, 2013. 38, 46, 49, 52

[62] S. M. Haque, A. Chatterjee et V. M. Govindu : High quality photometric reconstruction using a depth camera. *In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2283–2290, 2014. 38

[63] C. Harris et M. Stephens : A combined corner and edge detector. *In Proceedings of the Alvey Vision Conference*, pages 147–151, 1988. 12

[64] R. Hartley : In defense of the eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 19(6):580–593, 1997. 16

[65] R. Hartley et A. Zisserman : *Multiple view geometry in computer vision*. Cambridge University Press, 2003. 5, 11, 16

[66] R. I. Hartley et P. Sturm : Triangulation. *In Proceedings of the International Conference on Computer Analysis of Images and Patterns*, pages 190–197, 1995. 17

[67] M. Hassaballah, A. A. Abdelmgeid et H A. Alshazly : *Image Features Detection, Description and Matching*, pages 11–45. Springer, 2016. 12

[68] C. Hernández : *Stereo and Silhouette Fusion for 3D Object Modeling from Uncalibrated Images Under Circular Motion*. Thèse de doctorat, École Nationale Supérieure des Télécommunications, 2004. 20

[69] C. Hernández et F. Schmitt : Silhouette and stereo fusion for 3D object modeling. *Computer Vision and Image Understanding (CVIU)*, 96(3):367–392, 2004. 2

[70] C. Hernández, G. Vogiatzis, G. J. Brostow, B. Stenger et R. Cipolla : Non-rigid Photometric Stereo with Colored Lights. *In Proceedings of the International Conference on Computer Vision (ICCV)*, pages 2104–2114, 2007. 34

[71] H. Hirschmuller : Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):328–341, 2008. 20

[72] M. Hong, Z.-Q. Luo et M. Razaviyayn : Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. *SIAM Journal on Optimization (SIOPT)*, 26(1):337–364, 2016. 44

[73] R. Horaud et O. Monga : *Vision par ordinateur: outils fondamentaux.* Traité des nouvelles technologies, Série informatique. Éditions Hermès, 1995. 11

[74] B. K. P. Horn : *Shape From Shading: A Method for Obtaining the Shape of a Smooth Opaque Object From One View.* Thèse de doctorat, MIT, Department of Electrical Engineering and Computer Science, 1970. 2, 30, 33

[75] B. K. P. Horn et M. J. Brooks : The Variational Approach to Shape From Shading. *Computer Vision, Graphics, and Image Processing (CVGIP)*, 33(2):174–208, 1986. 33, 41

[76] B. K. P. Horn et M. J. Brooks : *Shape from Shading.* MIT Press, 1989. 33

[77] D. R. Hougen et N. Ahuja : Adaptive polynomial modelling of the reflectance map for shape estimation from stereo and shading. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 991–994, 1994. 38

[78] R. Huang et W. A. P. Smith : Shape-from-shading under complex natural illumination. *In Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 13–16, 2011. 38

[79] Y. Iwahori, H. Sugie et N. Ishii : Reconstructing shape from shading images under point light source illumination. *In Proceedings of the IEEE International Conference on Pattern Recognition (ICPR)*, volume 1, pages 83–87, 1990. 36

[80] M. Jancosek et T. Pajdla : Multi-view reconstruction preserving weakly-supported surfaces. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3121–3128, 2011. 51

[81] M. Jancosek et T. Pajdla : Exploiting visibility information in surface reconstruction to preserve weakly supported surfaces. *International Scholarly Research Notices*, 2014:1–20, 2014. 21

[82] H. Jin, D. Cremers, D. Wang, A. Yezzi, E. Prados et S. Soatto : 3-D Reconstruction of Shaded Objects from Multiple Images Under Unknown Illumination. *International Journal of Computer Vision (IJCV)*, 76(3):245–256, 2008. 58, 74

[83] M. K. Johnson et E. H Adelson : Shape estimation in natural illumination. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2553–2560, 2011. 34, 38

[84] Y.-C. Ju, M. Breuss, A. Bruhn et S. Galliani : Shape from Shading for Rough Surfaces: Analysis of the Oren-Nayar Model. *In Proceedings of the British Machine Vision Conference (BMVC)*, 2012. 36

[85] Y.-C. Ju, S. Tozza, M. Breuss, A. Bruhn et A. Kleefeld : Generalised Perspective Shape from Shading with Oren-Nayar Reflectance. *In Proceedings of the British Machine Vision Conference (BMVC)*, 2013. 36

[86] J. T. Kajiya : The rendering equation. *Proceedings of the Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 20(4):143–150, 1986. 29

[87] T. P. Kersten et M. Lindstaedt : Image-Based Low-Cost Systems for Automatic 3D Recording and Modelling of Archaeological Finds and Objects. *In Proceedings of the Euro-Mediterranean Conference*, pages 1–10, 2012. 1

[88] K. Kim, A. Torii et M. Okutomi : Multi-view Inverse Rendering Under Arbitrary Illumination and Albedo. *In Proceedings of the European Conference on Computer Vision (ECCV)*, pages 750–767, 2016. 74

[89] R. Kimmel et A. Bruckstein : Tracking level sets by level sets: a method for solving the shape from shading problem. *Computer Vision and Image Understanding (CVIU)*, 62(1):47–58, 1995. 33

[90] R. Kimmel, K. Siddiqi, B. Kimia et A. Bruckstein : Shape from shading: Level set propagation and viscosity solutions. *International Journal of Computer Vision (IJCV)*, 16:107–133, 1995. 33

[91] R. Kozera : Uniqueness in shape from shading revisited. *Journal of Mathematical Imaging and Vision (JMIV)*, 7(2):123–138, 1997. 30

[92] P.-Y. Laffont, A. Bousseau et G. Drettakis : Rich Intrinsic Image Decomposition of Outdoor Scenes from Multiple Views. *IEEE Transactions on Visualization and Computer Graphics*, 19(2):210–224, 2013. 74

[93] P.-Y. Laffont, A. Bousseau, S. Paris, F. Durand et G. Drettakis : Coherent intrinsic images from photo collections. *ACM Transactions on Graphics*, 31:202:1–202:11, 2012. 74

[94] E. H. Land et J. J. McCann : Lightness and retinex theory. *Journal of the Optical Society of America*, 61:1–11, 1971. 73, 78

[95] M. S. Langer et S. W. Zucker : Shape-from-shading on a cloudy day. *Journal of the Optical Society of America*, 11(2):467–478, 1994. 36

[96] F. Langguth, K. Sunkavalli, S. Hadap et M. Goesele : Shading-aware Multi-view Stereo. *In Proceedings of the European Conference on Computer Vision (ECCV)*, pages 469–485, 2016. 58, 74

[97] V. Le Guen : Cartoon + Texture Image Decomposition by the TV-L1 Model. *Image Processing On Line*, 4:204–219, 2014. 73, 83, 84, 85, 86

[98] Y. G. Leclerc et A. F. Bobick : The direct computation of height from shading. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 552–558, 1991. 38

[99] K. M. Lee et C. C. J. Kuo : Shape from shading with perspective projection. *Computer Vision, Graphics, and Image Processing (CVGIP)*, 59(2):202–212, 1994. 35

[100] V. Lepetit, F. Moreno-Noguer et P. Fua : EPnP: An Accurate O(n) Solution to the PnP Problem. *International Journal of Computer Vision (IJCV)*, 81:155–166, 2009. 18

[101] B. Levy, S. Petitjean, N. Ray et J. Maillot : Least squares conformal maps for automatic texture atlas generation. *ACM Transactions on Graphics*, 21:362–371, 07 2002. 21

[102] G. Li et T. K. Pong : Global convergence of splitting methods for nonconvex composite optimization. *SIAM Journal on Optimization*, 25(4):2434–2460, 2015. 44

[103] P.-L. Lions, E. Rouy et A Tourin : Shape-from-shading, viscosity solutions and edges. *Numerische Mathematik*, 64(1):323–353, 1993. 33

[104] D. C. Liu et J. Nocedal : On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1-3):503–528, 1989. 44, 64

[105] H.C. Longuet-Higgins : A computer algorithm for reconstructing a scene from two projections. *In Readings in Computer Vision*, pages 61–62. Morgan Kaufmann, 1987. 16

[106] D. G. Lowe : Object recognition from local scale-invariant features. *In Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 1150–1157, 1999. 12, 13

[107] R. Maier, K. Kim, D. Cremers, J. Kautz et M. Niessner : Intrinsic3D: High-Quality 3D Reconstruction by Joint Appearance and Geometry Optimization with Spatially-Varying Lighting. *In Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 3133–3141, 2017. 58, 69, 74

[108] D. W. Marquardt : An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963. 19

[109] D. Maurer, Y. C. Ju, M. Breuss et A. Bruhn : Combining shape from shading and stereo: A joint variational method for estimating depth, illumination and albedo. *International Journal of Computer Vision (IJCV)*, 126(12):1342–1366, 2018. 58, 59, 74

[110] R. Mecca, Y. Quéau, F. Logothetis et R. Cipolla : A single-lobe photometric stereo approach for heterogeneous material. *SIAM Journal on Imaging Sciences*, 9(4):1858–1888, 2016. 34

[111] J. Mélou, Y. Quéau, F Castan et J.-D. Durou : A splitting-based algorithm for multi-view stereopsis of textureless objects. *In Proceedings of the International Conference on Scale Space and Variational Methods in Computer Vision (SSVM)*, pages 51–63, 2019. 4, 55

[112] J. Mélou, Y. Quéau, F Castan et J.-D. Durou : Une solution numérique générique pour la reconstruction 3D d'objets non texturés. *In Actes des journées francophones des jeunes chercheurs en vision par ordinateur (ORASIS)*, 2019. 4, 55

[113] J. Mélou, Y. Quéau, J.-D. Durou, F. Castan et D Cremers : Beyond Multi-view Stereo: Shading-Reflectance Decomposition. *In Proceedings of the International Conference on Scale Space and Variational Methods in Computer Vision (SSVM)*, pages 694–705, 2017. 4, 71

[114] J. Mélou, Y. Quéau, J.-D. Durou, F. Castan et D Cremers : Estimation de la réflectance à partir de données multi-vues. *In Actes des journées francophones des jeunes chercheurs en vision par ordinateur (ORASIS)*, 2017. 4, 71

[115] J. Mélou, Y. Quéau, J.-D. Durou, F. Castan et D Cremers : Variational reflectance estimation from multi-view images. *Journal of Mathematical Imaging and Vision (JMIV)*, 60(9):1527–1546, 2018. 4, 71

[116] R. Montes Soldado et C. Ureña Almagro : An Overview of BRDF Models. Technical Report, Universidad de Granada, Espagne, 2012. 29

[117] G. Morin : *3D models for...* Habilitation à diriger des recherches, Université de Toulouse, 2014. 1

[118] P. Moulon : *Positionnement robuste et précis de réseaux d'images*. Thèse de doctorat, Université de Paris-Est, 2014. 2, 11

[119] P. Moulon, P. Monasse et R. Marlet : openMVG: An open multiple view geometry library. https://github.com/openMVG/openMVG, 2014. 11

[120] D. Mumford : Bayesian rationale for the variational formulation. *In Geometry-Driven Diffusion in Computer Vision*, pages 135–146. Springer, 1994. 77

[121] A. Nadian-Ghomsheh, Y. Hassanian et K. Navi : Intrinsic Image Decomposition via Structure-Preserving Image Smoothing and Material Recognition. *PLoS ONE*, 11(12):1–22, 2016. 73, 74

[122] G. Nam, J. Lee, D. Gutierrez et M. Kim : Practical SVBRDF Acquisition of 3D Objects with Unstructured Flash Photography. *ACM Transactions on Graphics*, 37(6), 2018. 90

[123] S. K. Nayar et Y. Nakagawa : Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 16(8):824–831, 1994. 2

[124] D. Nehab, S. Rusinkiewicz, J. Davis et R. Ramamoorthi : Efficiently combining positions and normals for precise 3D geometry. *ACM Transactions on Graphics*, 24(3):536–543, 2005. 56

[125] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg et T. Limperis : Geometrical considerations and nomenclature for reflectance. NBS Monograph 160, National Bureau of Standards, Washington D.C., États-Unis, 1977. 26

[126] J. Oliensis : Uniqueness in Shape from Shading. *International Journal of Computer Vision (IJCV)*, 6(2):75–104, 1991. 30

[127] R. Or-El, G. Rosman, A. Wetzler, R. Kimmel et A. Bruckstein : RGBD-Fusion: Real-Time High Precision Depth Recovery. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5407–5416, 2015. 38, 39, 41, 44

[128] M. Oren et S. K. Nayar : Generalization of Lambert's Reflectance Model. *In Proceedings of the Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 239–246, 1994. 29

[129] J. Park, S. N. Sinha, Y. Matsushita, Y. W. Tai et I. S. Kweon : Robust multi-view photometric stereo using planar mesh parameterization. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 39(8):1591–1604, 2017. 90

[130] A. P. Pentland : Local shading analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 6(2):170–187, 1984. 30, 31, 33

[131] J. Philip, M. Gharbi, T. Zhou, A. Efros et G. Drettakis : Multi-view relighting using a geometry-aware network. *ACM Transactions on Graphics*, 38(4), 2019. 98

[132] B. T. Phong : Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, 1975. 29

[133] Pixar: RenderMan. https://renderman.pixar.com/. 1

[134] E. Prados et O. Faugeras : A generic and provably convergent shape-from-shading method for orthographic and pinhole cameras. *International Journal of Computer Vision (IJCV)*, 65(1):97–125, 2005. 33, 35

[135] E. Prados et O. Faugeras : Shape from shading: a well-posed problem ? *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 870–877, 2005. 36

[136] Y. Quéau, B. Durix, T. Wu, D. Cremers, F. Lauze et J.-D. Durou : LED-based Photometric Stereo: Modeling, Calibration and Numerical Solution. *Journal of Mathematical Imaging and Vision (JMIV)*, 60(3):313–340, 2018. 73, 82

[137] Y. Quéau, J.-D. Durou et J.-F. Aujol : Normal integration: a survey. *Journal of Mathematical Imaging and Vision (JMIV)*, 60(4):576–593, 2018. 33

[138] Y. Quéau, R. Mecca et J.-D. Durou : Unbiased Photometric Stereo for Colored Surfaces: A Variational Approach. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4359–4368, 2016. 34

[139] Y. Quéau, J. Mélou, F. Castan, D. Cremers et J.-D. Durou : A variational approach to shape-from-shading under natural illumination. *In Proceedings of the International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, pages 342–357, 2017. 4, 25

[140] Y. Quéau, T. Wu et D. Cremers : Semi-calibrated near-light photometric stereo. *In Proceedings of the International Conference on Scale Space and Variational Methods in Computer Vision (SSVM)*, pages 656–668, 2017. 34

[141] Y. Quéau : *Reconstruction tridimensionnelle par stéréophotométrie.* Thèse de doctorat, Université de Toulouse, 2015. 2, 34

[142] R. Ramamoorthi et P. Hanrahan : An Efficient Representation for Irradiance Environment Maps. *In Proceedings of the Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 497–500, 2001. 37

[143] E. Reinhard, G. Ward, S. Pattanaik et P. Debevec : *High Dynamic Range Imaging.* Morgan Kaufmann, 2005. 28, 96

[144] X. Ren et J Malik : Learning a classification model for segmentation. *In Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV)*, 2003. 12

[145] S. R. Richter et S. Roth : Discriminative shape from shading in uncalibrated illumination. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1128–1136, 2015. 38

[146] M. A. Robertson, S. Borman et R. L. Stevenson : Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *Journal of Electronic Imaging*, 12:2003, 1999. 28

[147] P. Rodríguez-Gonzálvez, M. Rodríguez-Martín, L. F. Ramos et D. González-Aguilera : 3D reconstruction methods and quality assessment for visual inspection of welds. *Automation in Construction*, 79:49–58, 2017. 1

[148] E. Rouy et A. Tourin : A viscosity solutions approach to shape-from-shading. *SIAM Journal on Numerical Analysis*, 29(3):867–884, 1992. 33

[149] M. Schmidt : minFunc: unconstrained differentiable multivariate optimization in Matlab. http://www.cs.ubc.ca/~schmidtm/Software/minFunc.html, 2005. 44

[150] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein et R. Szeliski : Multi-view stereo evaluation web page. http://vision.middlebury.edu/mview/. 56

[151] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein et R. Szeliski : A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 519–528, 2006. 56

[152] S. A. Shafer et T. Kanade : Using shadows in finding surface orientations. *Computer Vision, Graphics, and Image Processing (CVGIP)*, 22(1):145–176, 1983. 2

[153] L. Shen et C. Yeo : Intrinsic images decomposition using a local and global sparse representation of reflectance. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 697–704, 2011. 73

[154] J. Song, H. Cho, J. Yoon et S. M. Yoon : Structure adaptive total variation minimization-based image decomposition. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(9):2164–2176, 2018. 73

[155] A. J. Stewart et M. S. Langer : Toward accurate recovery of shape from shading under diffuse lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 19(9):1020–1025, 1997. 36

[156] M. Storath et A. Weinmann : Fast partitioning of vector-valued images. *SIAM Journal on Imaging Sciences*, 7(3):1826–1852, 2014. 79

[157] C. Strecha, A. Bronstein, M. Bronstein et P. Fua : LDAHash: Improved Matching with Smaller Descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 34(1):66–78, 2012. 13

[158] J. Sun, M. Smith, L. Smith, L. Coutts, R. Dabis, C. Harland et J. Bamber : Reflectance of human skin using colour photometric stereo: with particular application to pigmented lesion analysis. *Skin Research and Technology*, 14(2):173–179, 2008. 1

[159] R. Szeliski : Fast shape from shading. *In Proceedings of the European Conference on Computer Vision (ECCV)*, pages 359–368, 1990. 33

[160] T. Taketomi, H. Uchiyama et S. Ikeda : Visual SLAM algorithms: a survey from 2010 to 2016. *IPSJ Transactions on Computer Vision and Applications*, 9(1):16, 2017. 17

[161] A. Tankus, N. Sochen et Y. Yeshurun : Shape-from-shading under perspective projection. *International Journal of Computer Vision (IJCV)*, 63(1):21–43, 2005. 35

[162] T. Tinne et M. Krystian : Local invariant feature detectors: A survey. *Foundations and Trends in Computer Graphics and Vision*, 3(3):177–280, 2008. 12

[163] R. Toldo, R. Gherardi, M. Farenzena et A. Fusiello : Hierarchical structure-and-motion recovery from uncalibrated images. *Computer Vision and Image Understanding (CVIU)*, 140:127–143, 2015. 17

[164] C. Tomasi et T. Kanade : Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision (IJCV)*, 9(2):137–154, 1992. 17

[165] S. Tozza et M. Falcone : Analysis and Approximation of Some Shape-from-Shading Models for Non-Lambertian Surfaces. *Journal of Mathematical Imaging and Vision (JMIV)*, 55(2):153–178, 2016. 36

[166] B. Triggs, P. McLauchlan, R. Hartley et A. Fitzgibbon : Bundle Adjustment - A Modern Synthesis. *In Vision Algorithms: Theory and Practice*, pages 298–372, 2000. 18

[167] Unity-Technologies: DeLightingTool. https://github.com/Unity-Technologies/DeLightingTool/. 96

[168] O. Vogel, M. Breuss, T. Leichtweis et J. Weickert : Fast shape-from-shading for Phong-type surfaces. *In Proceedings of the International Conference on Scale Space and Variational Methods in Computer Vision (SSVM)*, pages 733–744, 2009. 36

[169] O. Vogel, M. Breuss et J. Weickert : Perspective Shape from Shading with Non-Lambertian Reflectance. *In Pattern Recognition*, pages 517–526, 2008. 36

[170] A. Wendel, M. Maurer, G. Graber, T. Pock et H. Bischof : Dense reconstruction on-the-fly. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1450–1457, 2012. 43, 60

[171] A. P. Witkin : Recovering surface shape and orientation from texture. *Artificial Intelligence*, 17(1):17–45, 1981. 2

[172] R. J. Woodham : Photometric Method for Determining Surface Orientation from Multiple Images. *Optical Engineering*, 19(1):139–144, 1980. 34

[173] C. Wu, B. Wilburn, Y. Matsushita et C. Theobalt : High-quality shape from multiview stereo and shading under general illumination. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 969–976, 2011. 58

[174] C. Wu, M. Zollhöfer, M. Niessner, M. Stamminger, S. Izadi et C. Theobalt : Real-time shading-based refinement for consumer depth cameras. *ACM Transactions on Graphics*, 33(6):200:1–200:10, 2014. 38

[175] D. Xu, Q. Duan, J. Zheng, J. Zhang, J. Cai et T.-J. Cham : Recovering Surface Details under General Unknown Illumination Using Shading and Coarse Multi-view Stereo. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1526–1533, 2014. 56, 58

[176] G. Yu et J.-M. Morel : ASIFT: An Algorithm for Fully Affine Invariant Comparison. *Image Processing On Line (IPOL)*, 1:11–38, 2011. 12

[177] L. F. Yu, S. K. Yeung, Y. W. Tai et S. Lin : Shading-Based Shape Refinement of RGB-D Images. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1415–1422, 2013. 38

[178] R. Zhang, P.-S. Tsai, J. E. Cryer et M. Shah : Shape-from-shading: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 21(8):690–706, 1999. 33

[179] M. Zollhöfer, A. Dai, M. Innman, C. Wu, M. Stamminger, C. Theobalt et M. Niessner : Shading-based refinement on volumetric signed distance functions. *ACM Transactions on Graphics*, 34(4):96:1–96:14, 2015. 38, 51, 52, 58, 68

RÉSUMÉ

Les solutions opérationnelles de reconstruction 3D à partir de photographies se fondent généralement sur des méthodes multi-vues, dites géométriques, qui n'utilisent que très partiellement les informations photométriques contenues dans les images. L'utilisation de ces dernières doit permettre, à la fois, d'augmenter la finesse du relief reconstruit et d'estimer les caractéristiques photométriques de la scène que sont la réflectance de la surface et l'éclairage incident. Cette thèse montre comment une approche combinant les techniques multi-vues et les techniques photométriques peut effectivement améliorer un pipeline de reconstruction 3D. Nous commençons par présenter une façon de rendre la méthode du shape-from-shading opérationnelle en conditions d'éclairage naturel. Ceci nous permet de combiner les approches multi-vues et photométriques sous la forme d'un problème variationnel, pour lequel nous proposons un schéma numérique de résolution. Cependant, comme le shape-from-shading lui-même, la méthode proposée nécessite la connaissance a priori de la réflectance et de l'éclairage. Nous présentons une méthode d'estimation conjointe de ces deux grandeurs à partir d'un relief grossier de la scène. En guise de conclusion, nous présentons un algorithme qui, à partir des images utilisées par un pipeline de reconstruction 3D multi-vues, permet d'estimer une carte de réflectance et une carte de profondeur par vue, ainsi que l'éclairage de la scène.

ABSTRACT

Practical solutions for 3D-reconstruction from photographs are generally based on multi-view methods (called geometric methods) that only very partially use the photometric information contained in the images. The use of the latter should increase the fineness of the reconstructed relief and allow the estimate of the photometric characteristics of the scene, i.e. surface reflectance and incident lighting. This thesis shows how an approach combining multiview and photometric techniques can effectively improve a 3D-reconstruction pipeline. We start by presenting a way to make the shape-from-shading method operational in daylight conditions. This allows us to combine multiview and photometric approaches in the form of a variational problem, for which we propose a numerical solving scheme. However, like shape-from-shading itself, the proposed method requires an a priori knowledge of reflectance and lighting. We present a method for the joint estimation of both these quantities from a coarse relief of the scene. As a conclusion, we present an algorithm which, from the images used by a multi-view 3D-reconstruction pipeline, allows us to estimate a reflectance map and a depth map per view, as well as the lighting of the scene.