



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>
Eprints ID : 16999

The contribution was presented at RFIA 2016 :
<http://rfia2016.iut-auvergne.com/>

To cite this version : Quéau, Yvain and Durix, Bastien and Lucas, Tom and Boumaza, Jade and Durou, Jean-Denis and Lauze, François *Fusion de données RVB-D par stéréophotométrie colorée*. (2016) In: Congrès Francophone de Reconnaissance des Formes et Intelligence Artificielle (RFIA 2016), 27 June 2016 - 1 July 2016 (Clermont-Ferrand, France).

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Fusion de données RVB-D par stéréophotométrie colorée

Yvain QUÉAU¹

yvain.queau@enseeiht.fr

Bastien DURIX¹

bastien.durix@enseeiht.fr

Tom LUCAS¹

tomlucas@live.fr

Jade BOUMAZA¹

jade.boumaza@gmail.com

Jean-Denis DUROU¹

durou@irit.fr

François LAUZE²

francois@di.ku.dk

¹ IRIT-ENSEEIHT, Université de Toulouse

² Department of Computer Science, University of Copenhagen, Denmark

Résumé

Nous proposons une nouvelle approche de la stéréophotométrie, où nous éclairons la scène à l'aide de sources ponctuelles anisotropes de type LED. Cette approche, qui est à la fois différentielle et variationnelle, estime directement et de façon robuste la profondeur, sans estimation préalable des normales et de l'albédo. Nous montrons comment utiliser cette nouvelle approche pour améliorer le relief fourni par un capteur RVB-D, en équipant le capteur de trois LEDs colorées, ce qui nous permet d'appliquer la stéréophotométrie colorée à l'image RVB.

Mots Clef

Reconstruction 3D, stéréophotométrie, sources ponctuelles, capteurs RVB-D, méthodes variationnelles.

Abstract

We propose a new approach of photometric stereo with anisotropic pointwise sources such as LEDs. This approach is both differential and variational, in order to recover the depth directly and in a robust manner, without estimating the normals and the albedo in a first step. We then show how to use this approach in order to refine the depth map obtained with a RGB-D sensor, by integrating on the sensor three colored LEDs allowing us to apply colored photometric stereo on the RGB image.

Keywords

3D-reconstruction, Photometric stereo, pointwise sources, RGB-D sensors, variational methods.

1 Introduction

La stéréophotométrie [13] est une technique qui consiste à estimer le relief d'une surface à partir de $m \geq 3$ photographies prises sous le même angle, mais sous m éclairages directionnels non coplanaires. Le cas où $m = 3$ est

particulièrement intéressant, car il rend la reconstruction 3D possible en utilisant une seule image RVB. En effet, en éclairant une scène *simultanément* par trois sources lumineuses colorées en rouge, vert et bleu, convenablement positionnées, on obtient en une seule prise de vue l'équivalent de trois photographies prises sous trois éclairages différents, ce qui permet de reconstruire en temps réel le relief de surfaces déformables [8]. Toutefois, la qualité des reconstructions 3D ainsi obtenues peut laisser à désirer, car il est difficile de garantir une certaine robustesse au bruit et aux données aberrantes avec aussi peu d'images. D'autre part, les capteurs RVB-D fournissent non seulement une image RVB, mais aussi une estimation, plutôt grossière, de la profondeur D (*depth*). Dans cet article, nous montrons comment obtenir une reconstruction 3D de meilleure qualité en fusionnant ces deux estimations grossières du relief.

Stéréophotométrie avec des éclairages ponctuels. En pratique, le caractère directionnel de l'éclairage est difficile à garantir et très peu représentatif du flux lumineux émis par les sources lumineuses à bas coût de type LED. De plus, la résolution de la stéréophotométrie avec des sources ponctuelles est difficile, car l'éclairage dépend explicitement des coordonnées 3D des points de la surface, qui sont justement les inconnues. La manière usuelle de résoudre le problème dans ces conditions consiste à estimer itérativement les normales, connaissant la profondeur, puis la profondeur, connaissant les normales [3, 5, 10]. Une autre possibilité est d'estimer directement la profondeur, par déformation itérative du maillage représentant la surface [14]. Nous adoptons le formalisme introduit récemment dans [12], qui est à la fois variationnel et différentiel. Celui-ci permet également d'estimer directement la profondeur, tout en garantissant une certaine robustesse et en facilitant la prise en compte de données RVB et d'éclairages ponctuels et anisotropes (notre montage expérimental impose en effet de respecter l'ensemble de ces contraintes).

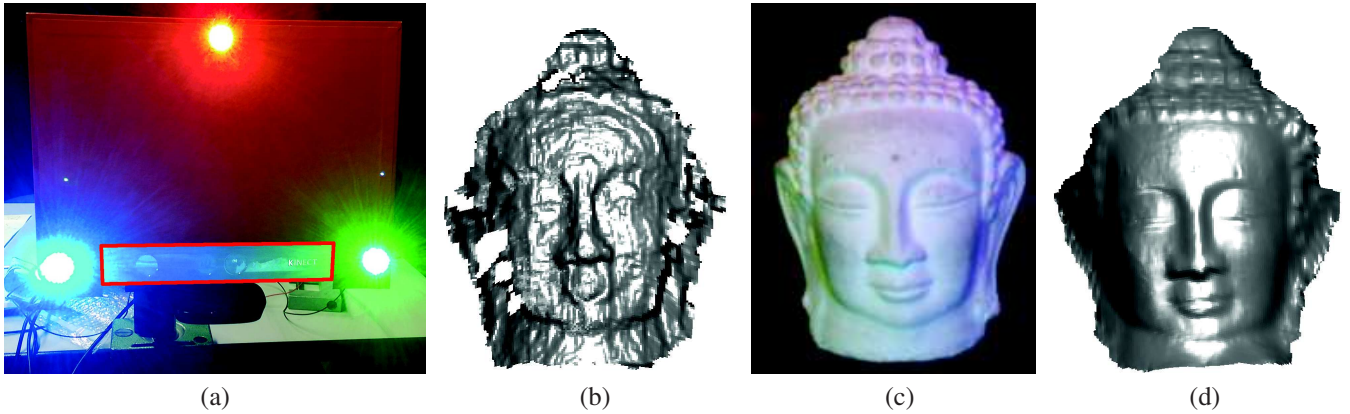


FIGURE 1 – (a) Photographie, prise depuis la scène à reconstruire, d’un capteur RVB-D (mis en évidence en rouge) équipé de trois LEDs colorées en rouge, vert et bleu. (b) Rééclairage de la reconstruction 3D fournie par le capteur de profondeur, qui est bruitée et comporte des données manquantes. (c) Image RVB fournie par le capteur, grâce à laquelle nous pouvons appliquer la stéréophotométrie colorée. (d) Rééclairage du résultat de la fusion de données RVB-D selon l’approche proposée.

Fusion de données RVB-D. Le problème de la fusion des données issues d’un capteur de profondeur standard et d’un capteur RVB standard a fait l’objet de plusieurs études récentes. Dans [11, 15], il est suggéré d’utiliser les données RVB, qui possèdent une meilleure résolution et sont généralement moins bruitées, afin d’affiner la « fonction de profondeur » (*depth map*). Au vu de la capacité des techniques photométriques à retrouver les détails les plus fins d’une surface, il a également été proposé de « guider » cette fusion par l’introduction d’un modèle photométrique. Dans [9, 16], la technique du *shape-from-shading* (SfS) est utilisée en conditions non contrôlées sur les canaux RVB : la donnée de profondeur est alors vue comme un a priori permettant de limiter les ambiguïtés inhérentes au SfS. Cependant, nous pouvons raisonnablement espérer que le contrôle de l’éclairage permette d’améliorer la qualité des reconstructions 3D, puisqu’il rend possible l’utilisation de la stéréophotométrie, qui constitue un problème bien posé, au contraire du SfS. Cette approche a été suivie dans [1], où il est proposé d’utiliser trois éclairages colorés pour que la stéréophotométrie colorée [8] puisse jouer le rôle de méthode d’affinage. L’éclairage y est toutefois supposé directionnel, ce qui limite l’utilisation de la technique en conditions réelles. Nous supposons au contraire que l’éclairage est ponctuel, hypothèse qui est plus adaptée à la réalisation d’un système à faible coût (cf. figure 1).

Structure de l’article. Nous proposons une approche variationnelle de la stéréophotométrie avec des éclairages ponctuels. Les limites de l’approche classique de la stéréophotométrie sont rappelées dans la partie 2. Nous présentons ensuite, dans la partie 3, un modèle différentiel pour la stéréophotométrie, plus adapté au caractère ponctuel de l’éclairage. L’inversion de ce modèle par une approche variationnelle est décrite dans la partie 4, et nous montrons dans la partie 5 comment étendre cette approche pour fusionner les données issues d’un capteur RVB-D.

2 Approche classique de la stéréophotométrie

La stéréophotométrie est généralement présentée comme un problème inverse, qui vise à résoudre un système d’équations modélisant les niveaux de gris, en fonction du relief et de la réflectance de la surface observée, des éclairages et de l’appareil photographique. Si les éclairages et l’appareil photographique sont étalonnés, les seules inconnues du problème sont le relief et la réflectance.

2.1 Inversion du modèle lambertien

Il est généralement supposé que l’appareil photographique fournit des images en niveaux de gris et que la surface est lambertienne, sa réflectance étant alors caractérisée par l’albédo $\rho > 0$. Les m éclairages sont supposés parallèles et uniformes, et sont donc caractérisés par des vecteurs $\mathbf{s}^i \in \mathbb{R}^3$, $i \in \{1, \dots, m\}$, orientés vers la source lumineuse, et dont la norme est proportionnelle à la densité du flux lumineux. La stéréophotométrie est alors formulée comme l’estimation de l’albédo ρ et de la normale \mathbf{n} , au point \mathbf{x} de la surface conjugué d’un pixel \mathbf{p} , par inversion du système d’équations suivant :

$$I^i(\mathbf{p}) = \underbrace{\rho(\mathbf{x})\mathbf{n}(\mathbf{x})}_{\mathbf{m}(\mathbf{x})} \cdot \mathbf{s}^i, \quad i \in \{1, \dots, m\} \quad (1)$$

où $I^i(\mathbf{p})$ désigne le niveau de gris au pixel \mathbf{p} dans la i -ème image, $\rho(\mathbf{x})$ l’albédo au point \mathbf{x} conjugué de \mathbf{p} , et $\mathbf{n}(\mathbf{x})$ la normale à la surface en \mathbf{x} ($\|\mathbf{n}(\mathbf{x})\| = 1$).

Dès lors que $m \geq 3$ éclairages *non coplanaires* sont utilisés, le système d’équations (1) admet une unique solution $\mathbf{m}(\mathbf{x})$ en moindres carrés, que l’on peut exprimer de manière analytique au moyen de la pseudo-inverse de la matrice obtenue par concaténation des vecteurs d’éclairage \mathbf{s}^i , et de laquelle on extrait facilement la normale et l’albédo :

$$\mathbf{n}(\mathbf{x}) = \frac{\mathbf{m}(\mathbf{x})}{\|\mathbf{m}(\mathbf{x})\|} \quad \rho(\mathbf{x}) = \|\mathbf{m}(\mathbf{x})\| \quad (2)$$

Le champ de normales ainsi estimé peut finalement être *intégré* en une fonction de profondeur. Toutefois, les conditions au bord, la forme du domaine Ω , et les éventuelles discontinuités de profondeur, constituent des difficultés non négligeables lors de cette dernière étape [7].

2.2 Limites de l'approche classique

Le recours à la variable auxiliaire $\mathbf{m} = \rho \mathbf{n}$ permet de linéariser la stéréophotométrie, puisque le vecteur \mathbf{m} n'est pas contraint à être unitaire. Toutefois, cette « astuce » présente les défauts suivants :

- Elle n'est adaptée qu'aux images en niveaux de gris (surfaces et sources « blanches »). Pour des images à plusieurs canaux, la définition de l'albédo dépend du canal de couleur considéré [6], alors que la normale n'en dépend pas. Les estimations de l'albédo et du relief sont donc généralement découplées [2], et l'intérêt de la linéarisation est perdu.
- Elle permet d'estimer non pas la fonction de profondeur, mais la normale à la surface. Une deuxième étape, l'intégration des normales, est donc requise [7]. Par conséquent, toute erreur sur l'estimation des normales sera propagée durant l'intégration. Ceci peut être évité en ayant recours à la formulation différentielle de la stéréophotométrie, qui vise à estimer directement la profondeur [12].
- Enfin, elle est limitée à des éclairages directionnels : pour une source ponctuelle telle qu'une LED, les vecteurs \mathbf{s}^i dépendent des coordonnées 3D du point \mathbf{x} de la surface éclairée. Or, ces coordonnées constituent justement les inconnues du problème de reconstruction 3D. L'utilisation de l'approche classique requiert donc d'alterner les phases d'estimation des normales, connaissant la profondeur, et d'intégration des normales [3, 10]. Là encore, la formulation différentielle s'avère intéressante, car elle ne fait intervenir que la profondeur.

Considérant que (i) la plupart des appareils photographiques fournissent des images en couleur et non en niveaux de gris, (ii) la finalité de la reconstruction 3D est l'estimation de coordonnées tridimensionnelles et non de normales, (iii) garantir le caractère directionnel de l'éclairage nécessite de recourir à des montages optiques complexes, tandis que les sources ponctuelles de type LED sont très peu chères et faciles à utiliser, nous proposons dans cet article une approche permettant de dépasser ces limitations.

3 Un nouveau modèle différentiel pour la stéréophotométrie

L'approche classique de la stéréophotométrie, rappelée dans la partie précédente, est adaptée à des éclairages directionnels et à des images en niveaux de gris. Nous proposons maintenant un nouveau modèle pour la stéréophotométrie, qui tient compte du caractère ponctuel de l'éclairage émis par une source lumineuse de type LED à spectre coloré.

3.1 Modèle d'éclairage

Rappelons que, pour la photographie d'une surface lambertienne blanche éclairée par une source ponctuelle de type LED, le niveau de couleur $I_\star(\mathbf{p})$, $\star \in \{R, V, B\}$, peut être modélisé de la façon suivante [6] :

$$I_\star(\mathbf{p}) = \frac{\gamma\beta\rho(\mathbf{x})}{\pi} \mathbf{s}_\star(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) \quad (3)$$

où β est un paramètre dépendant, entre autres, de l'ouverture de l'objectif et du temps de pose, γ est le facteur de réponse du capteur, supposée linéaire, et $\rho(\mathbf{x})$ est l'albédo de la surface au point \mathbf{x} . Le vecteur d'éclairage $\mathbf{s}_\star(\mathbf{x})$ en \mathbf{x} , pour une source localisée en \mathbf{x}_s , de coefficient d'anisotropie μ , s'écrit :

$$\mathbf{s}_\star(\mathbf{x}) = \frac{\phi_\star \cos^\mu \theta}{\|\mathbf{x}_s - \mathbf{x}\|^2 \|\mathbf{x}_s - \mathbf{x}\|} \mathbf{x}_s - \mathbf{x} \quad (4)$$

où θ désigne l'angle entre la direction d'éclairage $\mathbf{x} - \mathbf{x}_s$ et la direction principale \mathbf{n}_s de la LED, et où l'intensité ϕ_\star est définie comme suit :

$$\phi_\star = \int_{\lambda=0}^{+\infty} c_\star(\lambda) \phi(\lambda) d\lambda \quad (5)$$

où $\phi(\lambda)$ caractérise le spectre d'émission de la LED et $c_\star(\lambda)$ le spectre de transmission du filtre coloré $\star \in \{R, V, B\}$. Nous supposons que les paramètres \mathbf{x}_s , \mathbf{n}_s et ϕ_\star sont étalonnés selon la procédure décrite dans [6], et que μ est donné par le fabricant.

3.2 Modèle photométrique différentiel

Afin d'obtenir directement, par stéréophotométrie, une fonction de profondeur $z : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ caractérisant le relief de la surface photographiée, nous choisissons la formulation différentielle de la stéréophotométrie. Pour ce faire, il nous faut préciser le lien entre la normale à la surface et le gradient de la fonction de profondeur.

Nous utilisons le modèle d'appareil photographique en « trou d'épingle », représenté par la *matrice des paramètres internes* :

$$\mathbf{K} = \begin{bmatrix} f_u & 0 & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (6)$$

où $\mathbf{p}_0 = [u_0, v_0]^\top$ est la position du *point principal*, et les paramètres f_u et f_v , exprimés en pixels (px), représentent le produit de la distance focale (en mm) par les dimensions caractéristiques d'un pixel dans les directions horizontale et verticale (en $px.mm^{-1}$). Ces paramètres sont estimés durant la phase d'étalonnage. Le problème de la reconstruction 3D consiste à estimer :

$$\mathbf{x} = z \left[\frac{u - u_0}{f_u}, \frac{v - v_0}{f_v}, 1 \right]^\top \quad (7)$$

c'est-à-dire à déterminer la valeur de z associée à chaque pixel. Nous notons :

$$\begin{aligned} z : \Omega \subset \mathbb{R}^2 &\rightarrow \mathbb{R} \\ \mathbf{p} = [u, v]^\top &\mapsto z(\mathbf{p}) \end{aligned} \quad (8)$$

la fonction de profondeur associant à tout pixel la valeur de z correspondante, et $\nabla z(\mathbf{p}) = [\partial_u z(\mathbf{p}), \partial_v z(\mathbf{p})]^\top$ le gradient de cette fonction au pixel \mathbf{p} .

On montre facilement que la normale $\mathbf{n}(\mathbf{x})$ s'écrit :

$$\mathbf{n}(\mathbf{x}) = \frac{1}{d_z(\mathbf{p})} \underbrace{\begin{bmatrix} f_u \partial_u z(\mathbf{p}) \\ f_v \partial_v z(\mathbf{p}) \\ -z(\mathbf{p}) - [\mathbf{p} - \mathbf{p}_0]^\top \cdot \nabla z(\mathbf{p}) \end{bmatrix}}_{\bar{\mathbf{n}}(\mathbf{p})} \quad (9)$$

où $d_z(\mathbf{p}) = \|\bar{\mathbf{n}}(\mathbf{p})\|$ est le facteur de normalisation garantissant que la normale est unitaire, dont la dépendance en z est non linéaire.

En remplaçant dans (3) la normale $\mathbf{n}(\mathbf{x})$ par sa définition différentielle (9), et en développant le produit scalaire, on obtient en quelques lignes, pour chaque canal de chaque image I_\star^i , $(i, \star) \in \{1, \dots, m\} \times \{R, V, B\}$:

$$I_\star^i(\mathbf{p}) = \frac{\gamma\beta\rho(\mathbf{x})}{\pi d_z(\mathbf{p})} [\mathbf{t}_\star^i(\mathbf{x}) \cdot \nabla z(\mathbf{p}) - s_{\star,3}^i z(\mathbf{p})] \quad (10)$$

avec :

$$\mathbf{t}_\star^i(\mathbf{x}) = \begin{bmatrix} f_u s_{\star,1}^i(\mathbf{x}) - (u - u_0) s_{\star,3}^i(\mathbf{x}) \\ f_v s_{\star,2}^i(\mathbf{x}) - (v - v_0) s_{\star,3}^i(\mathbf{x}) \end{bmatrix} \quad (11)$$

où nous notons $\mathbf{s}_\star^i = [s_{\star,1}^i, s_{\star,2}^i, s_{\star,3}^i]^\top$ les trois composantes du vecteur d'éclairage \mathbf{s}_\star^i .

Le système d'équations (10) est un système d'EDPs non linéaires en z . Il présente l'avantage appréciable de faire intervenir explicitement la fonction de profondeur z , dont la connaissance est requise pour calculer l'éclairage incident en un point \mathbf{x} de la surface. Par rapport à (3), cette formulation présente cependant deux difficultés : elle est *non linéaire*, et elle ne peut être résolue localement, c'est-à-dire en un point \mathbf{x} indépendamment de ses voisins.

Nous montrons dans le prochain paragraphe comment éliminer la non linéarité due au dénominateur, ainsi que les paramètres γ et β et l'albédo $\rho(\mathbf{x})$, afin d'obtenir un système d'EDPs *quasi-linéaires* en z .

3.3 Élimination des non linéarités

La première difficulté du système d'EDPs (10) est la non linéarité due au facteur de normalisation $d_z(\mathbf{p})$, dont le rôle est de rendre la normale unitaire. L'absence de connaissance a priori sur l'albédo $\rho(\mathbf{x})$ et sur les paramètres γ et β constitue une autre difficulté. Toutefois, tous ces facteurs sont indépendants du numéro i de l'éclairage. Nous pouvons donc déduire d'une paire d'équations du type (10) vérifiées par les deux mêmes canaux de couleur (I_\star^i, I_\star^j), $i \neq j$, issus de deux images RVB I^i et I^j , les égalités suivantes :

$$\begin{aligned} \frac{\gamma\beta\rho(\mathbf{x})}{\pi d_z(\mathbf{p})} &= \frac{I_\star^i(\mathbf{p})}{\mathbf{t}_\star^i(\mathbf{x}) \cdot \nabla z(\mathbf{p}) - s_{\star,3}^i z(\mathbf{p})} \\ &= \frac{I_\star^j(\mathbf{p})}{\mathbf{t}_\star^j(\mathbf{x}) \cdot \nabla z(\mathbf{p}) - s_{\star,3}^j z(\mathbf{p})} \end{aligned} \quad (12)$$

Remarquons que des identifications similaires ont déjà été suggérées pour la stéréophotométrie. Il a notamment été montré dans [12] qu'elles permettaient de résoudre le problème de la stéréophotométrie pour une surface colorée éclairée par des sources blanches : nous complétons donc cette étude dans le présent article, en traitant le cas inverse d'une surface blanche éclairée par des sources colorées.

Nous tirons des égalités (12) l'EDP en z suivante :

$$\begin{aligned} &\underbrace{[I_\star^i(\mathbf{p})\mathbf{t}_\star^j(\mathbf{x}) - I_\star^j(\mathbf{p})\mathbf{t}_\star^i(\mathbf{x})]}_{\mathbf{a}_z^{i,j}(\mathbf{p}) \in \mathbb{R}^2} \cdot \nabla z(\mathbf{p}) \\ &= \underbrace{[I_\star^i(\mathbf{p})s_{\star,3}^j(\mathbf{x}) - I_\star^j(\mathbf{p})s_{\star,3}^i(\mathbf{x})]}_{b_z^{i,j}(\mathbf{p}) \in \mathbb{R}} z(\mathbf{p}) \end{aligned} \quad (13)$$

Les champs \mathbf{s} et \mathbf{t} dépendent de z , car $\mathbf{s}(\mathbf{x}) = \mathbf{s}(\mathbf{x}(z))$ et $\mathbf{t}(\mathbf{x}) = \mathbf{t}(\mathbf{x}(z))$, z étant la troisième composante de \mathbf{x} (cf. (7)). Ils ne dépendent toutefois pas de ∇z : l'EDP (13) est donc une EDP *quasi-linéaire*.

En formant le champ matriciel $\mathbf{A}_z : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}^{3 \binom{m}{2} \times 2}$, et le champ vectoriel $\mathbf{b}_z : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}^{3 \binom{m}{2}}$, où chaque ligne est constituée, respectivement, de l'un des vecteurs lignes $\mathbf{a}_z^{i,j}(\mathbf{p})^\top$, et de l'un des scalaires $b_z^{i,j}(\mathbf{p})$, nous obtenons le système d'EDPs quasi-linéaires en z suivant :

$$\mathbf{A}_z(\mathbf{p})\nabla z(\mathbf{p}) = \mathbf{b}_z(\mathbf{p}) \quad (14)$$

Cette formulation différentielle est plus simple que la formulation différentielle originale (10) : la non linéarité la plus importante, qui est due au facteur de normalisation d_z , a été éliminée. Il nous reste toutefois à proposer une méthode de résolution tenant compte du caractère redondant de ce système d'EDPs, qui sont en général incompatibles à cause des écarts au modèle, qui sont inévitables lorsque les données sont réelles. Cela fait l'objet de la prochaine partie.

4 Résolution variationnelle du système d'EDPs (14)

La résolution *exacte* du système (14) n'est généralement ni souhaitable (à cause des écarts au modèle comme le bruit ou les ombres), ni possible (les $3 \binom{m}{2}$ équations sont en général incompatibles, à cause justement des écarts au modèle). Le formalisme variationnel est bien adapté à la prise en compte de la redondance d'information et des écarts au modèle. Nous montrons dans cette partie comment résoudre le système d'EDPs (14) au sens d'un critère en moindres carrés.

Plutôt que de chercher une solution exacte du système (14), ce qui ne serait possible qu'avec $m = 2$ images en niveaux de gris, nous souhaitons modéliser le problème de la reconstruction 3D comme un problème variationnel. Nous faisons l'hypothèse que chaque égalité (14) n'est vérifiée qu'à un bruit additif près. Si nous supposons ce bruit gaussien et de moyenne nulle, la solution optimale (au sens du

maximum de vraisemblance) est alors obtenue en minimisant la somme, sur l'ensemble du domaine de reconstruction Ω , des résidus de (14) au carré :

$$\min_z \|\mathbf{A}_z \nabla z - \mathbf{b}_z\|_{L^2(\Omega)}^2 \quad (15)$$

En réalité, toute hypothèse sur le bruit devrait être formulée sur les images, et non sur le modèle (14), qui a été obtenu à partir de ratios entre images, cf. (12) : si le bruit sur les images est gaussien, alors le bruit sur leurs ratios suit une loi de Cauchy. Par conséquent, la solution en moindres carrés (15) est le meilleur estimateur linéaire non biaisé, mais il ne s'agit pas de la solution optimale, qui nécessiterait le recours à des estimateurs non convexes. Ceci est toutefois laissé en perspective.

La non linéarité du problème en moindres carrés (15) provient de la dépendance en z des champs \mathbf{A} et \mathbf{b} . Or, d'après (13), (11) et (4), cette dépendance est due à l'atténuation du flux lumineux en $1/\|\mathbf{x}_s - \mathbf{x}\|^2$ (sources ponctuelles) et en $\cos^\mu \theta$ (sources anisotropes). Si l'objet observé est suffisamment « petit » par rapport à sa distance aux sources lumineuses, alors la variation du résidu $\|\mathbf{A}_z \nabla z - \mathbf{b}_z\|_{L^2(\Omega)}^2$ est principalement due au gradient. Ce raisonnement nous invite, malgré l'absence de garantie sur la convexité du problème (15), à choisir un schéma numérique de type point fixe, qui découple l'estimation de z de celles de \mathbf{A}_z et \mathbf{b}_z . Commençons par réécrire le problème (15) sous la forme d'un problème « contraint » :

$$\begin{aligned} \min_z \|\mathbf{A} \nabla z - \mathbf{b}\|_{L^2(\Omega)}^2 \\ \text{s.c. } \begin{bmatrix} \mathbf{A} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_z \\ \mathbf{b}_z \end{bmatrix} \end{aligned} \quad (16)$$

Nous résolvons ce problème par un schéma itératif de type point fixe. Une itération de ce schéma s'écrit :

$$z^{(k+1)} = \underset{z}{\operatorname{argmin}} \|\mathbf{A}^{(k)} \nabla z - \mathbf{b}^{(k)}\|_{L^2(\Omega)}^2 \quad (17)$$

$$\begin{bmatrix} \mathbf{A}^{(k+1)} \\ \mathbf{b}^{(k+1)} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{z^{(k+1)}} \\ \mathbf{b}_{z^{(k+1)}} \end{bmatrix} \quad (18)$$

Le problème (17) est un problème variationnel en moindres carrés linéaires. En discrétisant l'opérateur gradient par des différences finies (dans l'implémentation actuelle, nous utilisons des différences finies avant d'ordre 1, avec une condition au bord de type Neumann), nous pouvons réécrire (17) comme un problème d'optimisation discrète, dont la condition nécessaire d'optimalité est un simple système linéaire ayant pour inconnues les valeurs de z :

$$\underbrace{(\overline{\mathbf{A}\mathbf{D}})^\top \overline{\mathbf{A}\mathbf{D}}}_{\overline{\mathbf{A}}} \mathbf{z} = (\overline{\mathbf{A}\mathbf{D}})^\top \overline{\mathbf{b}} \quad (19)$$

où $\overline{\mathbf{A}}$, $\overline{\mathbf{b}}$ et \mathbf{D} désignent, respectivement, les discrétisations des champs \mathbf{A} , \mathbf{b} et de l'opérateur gradient, et où le vecteur colonne \mathbf{z} est construit par concaténation des valeurs de z .

La matrice $\overline{\mathbf{A}}$ du système linéaire (19) est une matrice symétrique, *semi-définie positive*, dont le noyau est de dimension 1. La déficience de rang de $\overline{\mathbf{A}}$, qui provient de ce que $\mathbf{A} \nabla z - \mathbf{b} = \mathbf{A} \nabla (z + k) - \mathbf{b}$ pour toute constante k , ne serait pas survenue avec le problème en moindres carrés non linéaires initial (15), puisque les champs \mathbf{A}_z et \mathbf{b}_z ne sont pas invariants par translation de z . Néanmoins, l'utilisation d'un algorithme de descente de type gradient conjugué garantit une convergence rapide vers une des solutions de (19) (à chaque itération ($k + 1$), la valeur précédente $z^{(k)}$ est choisie comme initialisation).

Cet algorithme permet de résoudre le problème de la stéréophotométrie avec des sources ponctuelles. Nous présentons sur la figure 2 un résultat de reconstruction 3D obtenu à partir de $m = 8$ images d'une statuette en plâtre, éclairée successivement par $m = 8$ LEDs étalonnées selon la procédure décrite dans [6].



FIGURE 2 – Reconstruction 3D par stéréophotométrie. En haut : trois (parmi $m = 8$) images d'une statuette en plâtre, éclairée par $m = 8$ sources ponctuelles. En bas : résultat de l'approche variationnelle proposée.

Malgré tout, le temps d'acquisition des $m = 8$ images, qui est non négligeable, empêche toute utilisation en temps réel. Pour une telle utilisation, il est nécessaire d'utiliser un nombre d'images minimal, au risque de perdre la robustesse de la méthode.

Nous allons voir qu'il est très facile d'étendre cette approche pour estimer, à partir de deux estimations de la profondeur, certes grossières, mais obtenues en temps réel (respectivement, par stéréophotométrie à une seule image RVB et par un capteur de profondeur), une reconstruction 3D de qualité comparable à celle que permet d'obtenir la stéréophotométrie avec un grand nombre d'images.

5 Fusion de données RVB-D

Nous nous intéressons maintenant plus spécifiquement aux capteurs de type RVB-D. La donnée de profondeur issue de tels capteurs est généralement très bruitée. Nous suggérons, pour en améliorer la qualité, une extension des systèmes RVB-D très simple à réaliser et peu coûteuse, qui consiste à adjoindre au système trois sources de lumière colorées (cf. figure 1-a). L'approche variationnelle de la partie 4 peut alors être utilisée pour fusionner la reconstruction 3D issue du capteur de profondeur avec celle obtenue par stéréophotométrie à partir d'une seule image RVB [8].

5.1 Modèle variationnel L^2 - L^2

Soit z_0 la fonction de profondeur fournie par le capteur, que nous cherchons à débruiter. Nous pourrions chercher à débruiter z_0 par une approche variationnelle, en minimisant la somme d'un terme d'attache aux données de la forme $\lambda \|z - z_0\|_{L^2(\Omega)}^2$, $\lambda > 0$, et d'un terme de régularisation spatiale de la forme $\|\nabla z\|_{L^2(\Omega)}^2$ (Sobolev) ou $\|\nabla z\|_{L^2(\Omega)}$ (variation totale). Cependant, un tel modèle ne tiendrait pas compte du fait que nous disposons également, en plus de la donnée de profondeur z_0 , de l'image RVB.

Or, cette image peut être utilisée par la technique de stéréophotométrie colorée : chacun des canaux R, V et B peut être vu comme une image en niveaux de gris, obtenue sous un éclairage que nous supposons ponctuel et uniquement dû à la LED de même couleur. Les trois niveaux de couleur en chaque pixel \mathbf{p} définissent alors une contrainte photométrique de la forme (14), où $\mathbf{A}_z(\mathbf{p}) \in \mathbb{R}^{3 \times 2}$ et $\mathbf{b}_z(\mathbf{p}) \in \mathbb{R}^3$ (cf. partie 3). Nous suggérons d'utiliser cette contrainte photométrique comme terme d'attache aux données quadratique (L^2) dans un problème variationnel où la connaissance a priori z_0 est prise en compte par un terme de régularisation quadratique (L^2) :

$$\min_z \underbrace{\|\mathbf{A}_z \nabla z - \mathbf{b}_z\|_{L^2(\Omega)}^2 + \|\sqrt{\lambda}(z - z_0)\|_{L^2(\Omega)}^2}_{\mathcal{E}(z)} \quad (20)$$

La complétion des données manquantes est réalisée en attribuant à λ une dépendance spatiale : nous fixons $\lambda(u, v) > 0$ là où une donnée de profondeur $z_0(u, v)$ est disponible, et $\lambda(u, v) = 0$ ailleurs. Le terme photométrique se comporte donc à la fois comme un terme de « débruitage » et d'« interpolation » (*inpainting*).

Contrairement aux modèles de débruitage usuels, le modèle (20) ne comporte pas de terme explicite de régularisation spatiale. Dans le cas du *shape-from-shading*, il est nécessaire d'introduire un tel terme car le problème est mal posé. Cela n'est toutefois pas nécessaire en stéréophotométrie calibrée, puisque le problème devient bien posé.

La résolution du problème (20) n'est guère différente de celle de (15) : l'ajout du terme de régularisation par l'a priori issu du capteur de profondeur ne modifie que les coefficients diagonaux de la matrice $\overline{\mathbf{A}}$ définie en (19). De plus, le terme de régularisation n'est pas invariant par translation : la déficience de rang est donc éliminée dès lors que $\lambda(u, v) > 0$ en au moins un pixel (u, v) .

La figure 3 montre l'évolution du relief estimé au cours des itérations, sur les données de la figure 1. La figure 4 montre que le réglage de la valeur du paramètre λ revient à choisir un « compromis » entre la donnée de profondeur z_0 , qui est bruitée, et la reconstruction 3D par stéréophotométrie qui, à cause du faible nombre $m = 3$ d'images, est elle aussi peu fiable. Enfin, notre méthode est comparée sur la figure 5 à la méthode d'Or-El et al. [9].

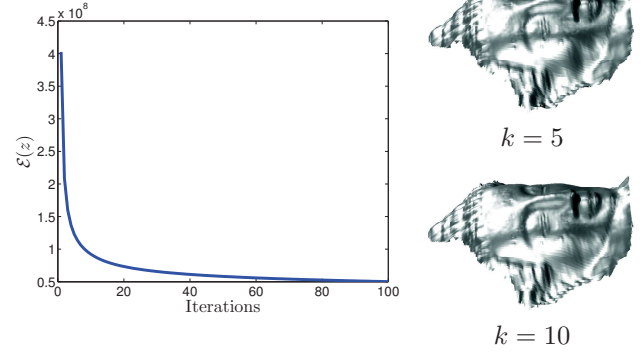


FIGURE 3 – À gauche : graphe de l'énergie $\mathcal{E}(z)$ définie en (20), en fonction des itérations, pour $\lambda = 10^4$. À droite : relief estimé après k itérations. Même si la convergence du schéma vers la solution de (20) est plutôt lente, en pratique 5 à 10 itérations (≈ 2 s de calcul en utilisant des codes Matlab non optimisés) suffisent à obtenir un résultat qualitativement satisfaisant.

5.2 Modèle robuste L^1 - L^1

L'approche en moindres carrés, bien que très rapide, n'est pas robuste aux données aberrantes. Par exemple, en présence d'ombres portées, d'ombres propres ou de reflets brillants, le niveau de couleur mesuré n'est pas fidèle au modèle (3). D'autre part, la profondeur issue du capteur RVB-D peut, en certains points, être manquante, ou aberrante à cause d'un mauvais alignement de l'image RVB avec la fonction de profondeur. Nous proposons de traiter les deux types de données aberrantes (photométriques et de profondeur) comme des écarts *parcimonieux*, qui peuvent être pris en compte en remplaçant la norme $L^2(\Omega)$ dans (20) par la norme $L^1(\Omega)$:

$$\min_z \|\mathbf{A}_z \nabla z - \mathbf{b}_z\|_{L^1(\Omega)} + \|\sqrt{\lambda}(z - z_0)\|_{L^1(\Omega)} \quad (21)$$

Le problème (21) est nettement plus difficile à résoudre que son équivalent quadratique (20). Cependant, l'introduction d'opérateurs proximaux permet de décomposer (21) en une suite de problèmes quadratiques plus faciles à résoudre. En introduisant une variable auxiliaire \bar{z} , nous réécrivons (21) de la façon suivante :

$$\min_{z, \bar{z}} \left\{ \|\mathbf{A}_z \nabla z - \mathbf{b}_z\|_{L^1(\Omega)} + \frac{1}{2\theta} \|z - \bar{z}\|_{L^2(\Omega)}^2 + \|\sqrt{\lambda}(\bar{z} - z_0)\|_{L^1(\Omega)} \right\} \quad (22)$$

où θ est un paramètre fixé par l'utilisateur (dans les tests, nous avons choisi $\theta = \frac{1}{2\lambda}$).

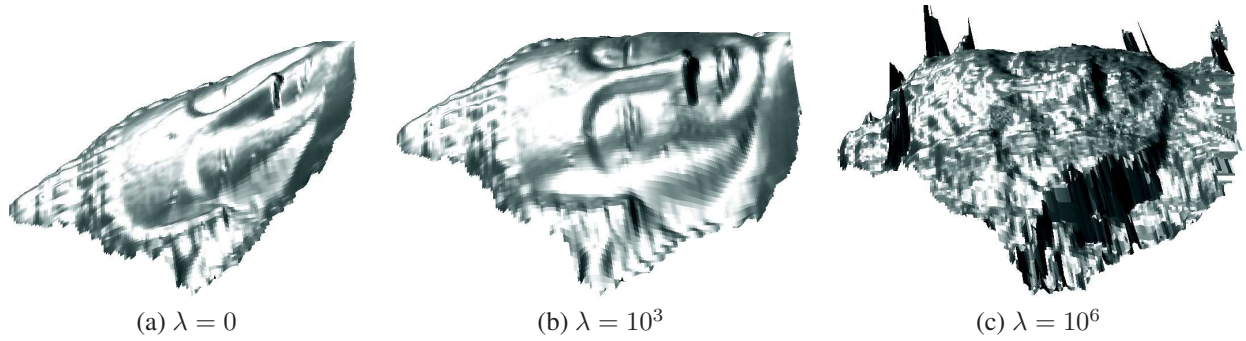


FIGURE 4 – Reconstructions 3D L^2 - L^2 obtenues pour plusieurs valeurs du paramètre λ . (a) Contrairement aux cas où de nombreuses images sont disponibles (cf. figure 2), la stéréophotométrie à une seule image RVB comporte un biais. (b) En augmentant le poids λ de l’a priori de profondeur, ce biais disparaît. (c) Toutefois, si λ est trop important, on retrouve un relief proche de celui fourni par le capteur de profondeur (cf. figure 5-a), qui est très bruité.

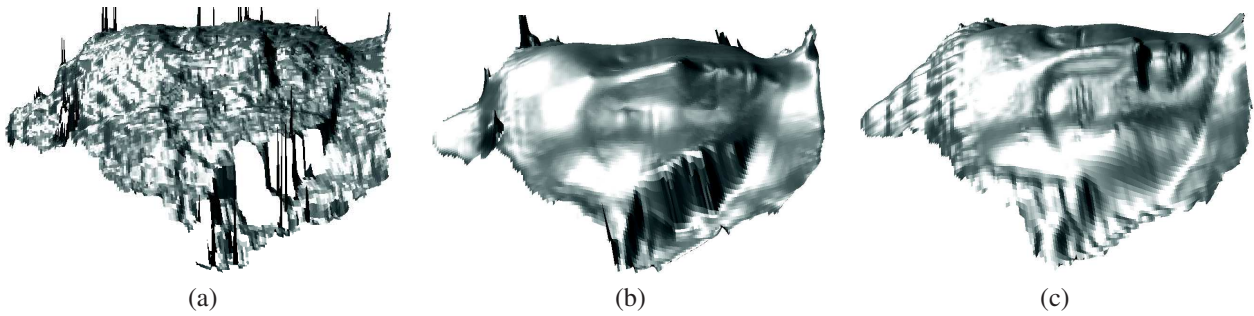


FIGURE 5 – (a) Relief z_0 issu du capteur RVB-D. (b) Fusion de données RVB-D par *shape-from-shading* non calibré [9]. (c) Reconstruction 3D L^2 - L^2 obtenu pour $\lambda = 10^4$, valeur qui donne visuellement le meilleur résultat.

Nous résolvons le problème (22) par un schéma d’optimisation alternée. L’estimation de z , à \bar{z} fixé, se ramène par un schéma ADMM [4] au problème (20). L’estimation de \bar{z} , à z fixé, admet la solution explicite suivante :

$$\bar{z} = \begin{cases} z^{(i+1)} + \frac{\lambda}{2\theta} & \text{si } 2\theta(z - z_0) \leq -\lambda \\ z^{(i+1)} - \frac{\lambda}{2\theta} & \text{si } 2\theta(z - z_0) \geq \lambda \\ z_0 & \text{sinon} \end{cases} \quad (23)$$

La figure 6 montre que le passage de la norme L^2 à la norme L^1 permet effectivement d’améliorer les résultats, en particulier au niveau des discontinuités de profondeur. Nous avons fixé $\lambda = 1$ pour la reconstruction L^1 - L^1 (nous avons vérifié expérimentalement que λ devait être plus faible qu’avec l’approche L^2 - L^2). Malgré le fait que le terme $\|\mathbf{A}_z \nabla z - \mathbf{b}_z\|_{L^1(\Omega)}$ soit à la fois non lisse et non convexe, ce qui ne permet pas de garantir la convergence de l’algorithme ADMM, nous n’avons pas constaté d’instabilité particulière, et la convergence a été atteinte en une dizaine d’itérations ($\approx 1 \text{ min}$).

6 Conclusion et perspectives

Nous avons proposé une reformulation de la stéréophotométrie sous la forme d’un système d’EDPs quasi-linéaires, et une approche variationnelle permettant de résoudre ce système. Nous avons également montré comment étendre cette approche pour prendre en compte une information a

priori sur la profondeur, obtenue par exemple avec un capteur RVB-D. Enfin, nous avons expérimentalement montré l’intérêt de notre approche en proposant une modification très simple des capteurs RVB-D : en équipant ces capteurs de LEDs colorées, elle permet d’améliorer sensiblement la qualité du relief fourni par de tels capteurs.

Dans cette application, nous avons cependant supposé que les sources lumineuses étaient parfaitement monochromatiques, afin de garantir que l’éclairage dans le canal \star n’est dû qu’à la source de couleur \star . Cette hypothèse n’étant probablement pas parfaitement vérifiée, nous pensons que la prise en compte du spectre de chaque source permettrait de corriger en partie le biais dans la stéréophotométrie à une seule image RVB (cf. figure 4-a). La méthode d’étalonnage de sources lumineuses à spectre coloré présentée dans [6] pourrait être utilisée pour estimer partiellement ces spectres.

Références

- [1] R. ANDERSON, B. STENGER et R. CIPOLLA : Augmenting depth camera output using photometric stereo. *In MVA*, pages 369–372, 2011.
- [2] S. BARSKY et M. PETROU : The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Trans. PAMI*, 25(10):1239–1252, 2003.

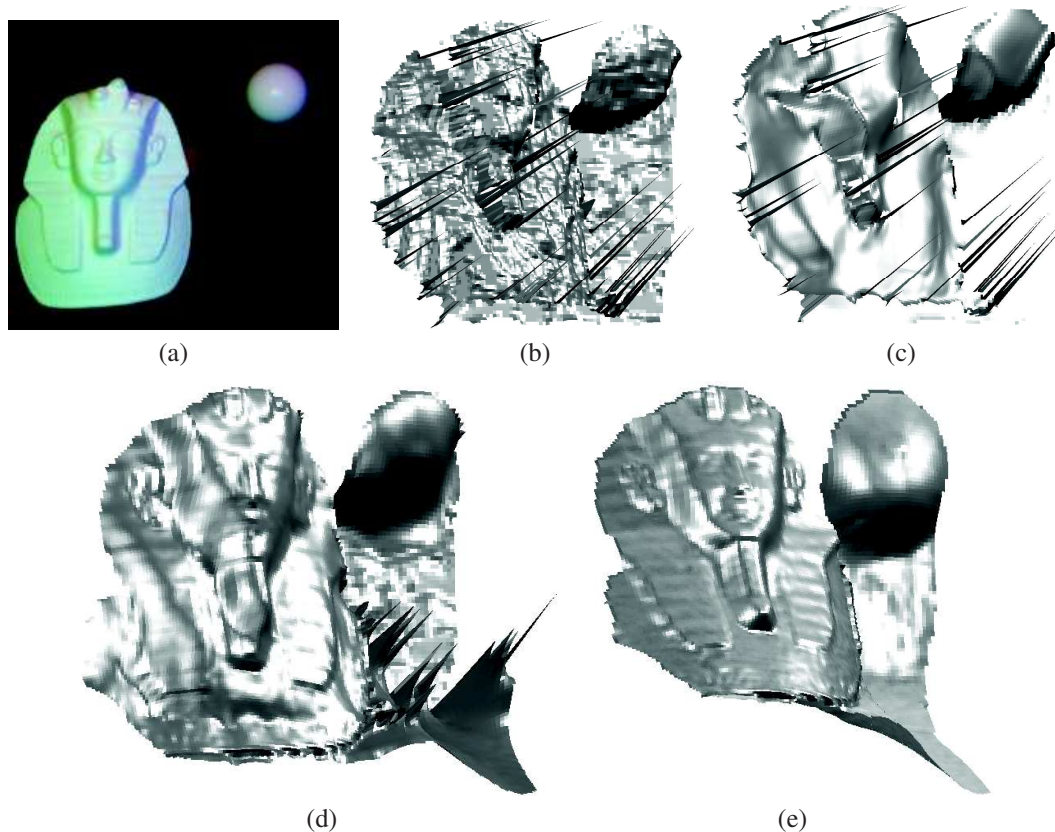


FIGURE 6 – (a) Une image RVB obtenue avec notre système. (b) Fonction de profondeur z_0 associée. (c) Fusion de données RVB-D non calibrée [9] : le relief est « sur-lissé ». (d) Fusion de données RVB-D L^2-L^2 : des artéfacts sont visibles au niveau des discontinuités de profondeur. (e) Résultat de l'approche robuste L^1-L^1 : les artéfacts sont fortement atténués.

- [3] A. BONY, B. BRINGIER et M. KHOUDER : Tridimensional reconstruction by photometric stereo with near spot light sources. *In EUSIPCO*, 2013.
- [4] S. BOYD, N. PARIKH, E. CHU, B. PELEATO et J. ECKSTEIN : Distributed optimization and statistical learning via the alternating direction method of multipliers. *Founds. Trends Mach. Learn.*, 3(1):1–122, 2011.
- [5] T. COLLINS et A. BARTOLI : 3D Reconstruction in Laparoscopy with Close-Range Photometric Stereo. *In MICCAI*, pages 634–642, 2012.
- [6] B. DURIX, Y. QUÉAU, J. BOUMAZA, T. LUCAS, J.-D. DUROU et F. LAUZE : Étalonnage de sources lumineuses de type LED. *In RFIA*, 2016.
- [7] J.-D. DUROU, J.-F. AUJOL et F. COURTEILLE : Integrating the normal field of a surface in the presence of discontinuities. *In EMMCVPR*, pages 261–273. 2009.
- [8] C. HERNÁNDEZ, G. VOGIATZIS, G. J. BROSTOW, B. STENGER et R. CIPOLLA : Non-rigid Photometric Stereo with Colored Lights. *In ICCV*, 2007.
- [9] R. OR-EL, G. ROSMAN, A. WETZLER, R. KIMMEL et A. M. BRUCKSTEIN : RGBD-Fusion: Real-Time High Precision Depth Recovery. *In CVPR*, 2015.
- [10] T. PAPADHIMITRI et P. FAVARO : Uncalibrated Near-Light Photometric Stereo. *In BMVC*, 2014.
- [11] J. PARK, H. KIM, Y. W. TAI, M. S. BROWN et I. S. KWEON : High-quality depth map upsampling and completion for RGB-D cameras. *IEEE Trans. Image Process.*, 23(12):5559–5572, 2014.
- [12] Y. QUÉAU, R. MECCA et J.-D. DUROU : Unbiased photometric stereo for colored surfaces: A variational approach. *In CVPR*, 2016.
- [13] R. J. WOODHAM : Photometric method for determining surface orientation from multiple images. *Opt. Eng.*, 19(1):134–144, 1980.
- [14] W. XIE, C. DAI et C. C. L. WANG : Photometric Stereo With Near Point Lighting: A Solution by Mesh Deformation. *In CVPR*, 2015.
- [15] J. YANG, X. YE, K. LI, C. HOU et Y. WANG : Color-guided depth recovery from RGB-D data using an adaptive autoregressive model. *IEEE Trans. Image Process.*, 23(8):3443–3458, 2014.
- [16] M. ZOLLHÖFER, A. DAI, M. INNMANN, C. WU, M. STAMMINGER, C. THEOBALT et M. NIESSNER : Shading-based refinement on volumetric signed distance functions. *ACM Trans. Graph.*, 34(4):1–14, 2015.