



Résolution du shape-from-shading par apprentissage

Jean-Denis Durou, Vincent Charvillat, Maxime Daramy, Pierre Gurdjos

► To cite this version:

Jean-Denis Durou, Vincent Charvillat, Maxime Daramy, Pierre Gurdjos. Résolution du shape-from-shading par apprentissage. ORASIS - Congrès des jeunes chercheurs en vision par ordinateur, Jun 2011, Praz-sur-Arly, France. 2011. <inria-00596354>

HAL Id: inria-00596354

<https://hal.inria.fr/inria-00596354>

Submitted on 27 May 2011

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

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

Résolution du shape-from-shading par apprentissage

Jean-Denis DUROU¹

Vincent CHARVILLAT²

Maxime DARAMY²

Pierre GURDJOS²

¹IRIT, Université Paul Sabatier, Toulouse

²IRIT, ENSEEIHT, Toulouse

durou@irit.fr, Vincent.Charvillat@enseeiht.fr

maxime.daramy@hotmail.fr, pgurdjos@enseeiht.fr

Résumé

Dans cet article, nous tentons de résoudre le problème du shape-from-shading par apprentissage. Un premier objet de relief connu est photographié sous un éclairage connu. Sa photographie est découpée en imagerie \mathcal{I}_a de taille 3×3 , associées à des normales N . Nous effectuons l'analyse en composantes principales (ACP) des données, afin de projeter les imagerie \mathcal{I}_a dans un espace de faible dimension n . Nous analysons ensuite la photographie d'un deuxième objet de relief inconnu, prise sous le même éclairage, que nous découpons en imagerie \mathcal{I}_t de même taille 3×3 . Pour chaque imagerie \mathcal{I}_t , nous cherchons l'imagerie \mathcal{I}_a la plus proche, et nous affectons à \mathcal{I}_t la normale N associée à \mathcal{I}_a . Cette méthode est testée sur images de synthèse et sur photographies réelles. Nous montrons son intérêt, mais aussi ses limites, en particulier face au problème bien connu de l'ambiguïté concave/convexe. Néanmoins, ce travail ouvre des perspectives pour la résolution de la stéréophotométrie.

Mots Clef

Reconstruction 3D, shape-from-shading, apprentissage.

Abstract

In this paper, we try to solve the shape-from-shading problem using machine learning techniques. A first object whose shape is known is imaged under a known lighting. Its photograph is cut into patches \mathcal{I}_a of size 3×3 , associated with normals N . We process a principal components analysis (PCA) on the data, in order to project the patches \mathcal{I}_a in a space of lower dimension n . Then, we analyse the photograph of a second object of unknown shape, taken under the same lighting, which we cut into patches \mathcal{I}_t of the same size 3×3 . For each patch \mathcal{I}_t , we search its closest patch \mathcal{I}_a and assign to \mathcal{I}_t the normal N associated to \mathcal{I}_a . This method is tested on synthetic, as well as on real images. We show its advantages, but also its drawbacks, particularly faced to the well-known problem of the concave/convex ambiguity. Nevertheless, this work opens some perspectives for the resolution of the photometric stereo problem.

Keywords

3D-reconstruction, Shape-from-shading, Machine Learning.

1 Modèles du shape-from-shading

Dans cet article, nous proposons une nouvelle méthode de résolution du shape-from-shading inspirée des techniques d'apprentissage artificiel. Dans la mesure où cette approche est nouvelle, il semble prudent de nous en tenir aux modèles les plus simples du shape-from-shading.

Modèle non différentiel Nous supposons que la scène est lambertienne, d'albédo uniforme égal à 1, et que l'éclairage est parallèle et uniforme, c'est-à-dire qu'il peut être décrit, en direction et en norme, par un vecteur \mathbf{S} (par convention, \mathbf{S} pointe vers la source lumineuse). Soit P un point de la surface de la scène et $\mathbf{N}(P)$ la normale sortante, de norme unitaire, en ce point. Sous ces hypothèses, l'image Q de P a pour niveau de gris :

$$I(Q) = K \mathbf{S} \cdot \mathbf{N}(P) \quad (1)$$

où K est une constante. En introduisant la version normalisée du vecteur d'éclairage $\mathbf{s} = \mathbf{S}/\|\mathbf{S}\|$ et le niveau de gris normalisé $i(Q) = I(Q)/(K \|\mathbf{S}\|)$, l'expression (1) du niveau de gris se réécrit :

$$i(Q) = \mathbf{s} \cdot \mathbf{N}(P) \quad (2)$$

Cette équation constitue un modèle particulièrement simple du shape-from-shading. Ce problème bien connu de vision par ordinateur, qui vise à calculer le relief (*shape*) d'une scène, à partir de l'ombrage (*shading*) lu sur une photographie, a été introduit par Horn en 1970 [12].

Le niveau de gris normalisé $i(Q)$ étant compris entre 0 et 1, il ressort de l'équation (2) que les valeurs extrêmes de $i(Q)$ correspondent à deux sortes de points particuliers :

- Les points Q où $i(Q) = 1$ sont tels que $\mathbf{N}(P)$ est parallèle à \mathbf{S} et de même sens. Ces points sont appelés des *points singuliers*.
- Les points Q où $i(Q) = 0$ sont tels que $\mathbf{N}(P)$ est orthogonal à \mathbf{S} . Ces points forment généralement des lignes dans l'image appelées *terminateurs*.

Si l'éclairage \mathbf{S} est connu, ce qui est généralement le cas lorsqu'on met en œuvre la technique du shape-from-shading, alors :

- En un point P dont l'image Q est un point singulier, la normale $\mathbf{N}(P)$ coïncide avec \mathbf{s} .
- En un point P dont l'image Q se situe sur un terminateur, la normale $\mathbf{N}(P)$ est orthogonale à \mathbf{s} . Si \mathbf{T} désigne la tangente au terminateur en Q , alors la projection \mathbf{n} de $\mathbf{N}(P)$ sur le plan image est orthogonale à \mathbf{T} (cf. figure 1), mais comme une ambiguïté subsiste sur le sens de \mathbf{n} , il existe deux valeurs possibles pour la normale $\mathbf{N}(P)$.
- Pour tous les autres points P , il existe une ambiguïté sur $\mathbf{N}(P)$. Nous pouvons seulement affirmer que l'angle θ entre $\mathbf{N}(P)$ et \mathbf{S} vaut $\theta = \arccos(i(Q))$ (cf. figure 2).

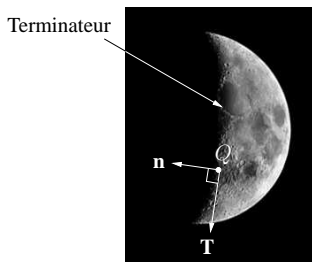


FIGURE 1 – Si l'image Q d'un point P se trouve sur un terminateur, alors la projection \mathbf{n} de $\mathbf{N}(P)$ sur le plan image est orthogonale à la tangente \mathbf{T} du terminateur en Q , mais nous ne pouvons rien affirmer sur le sens du vecteur \mathbf{n} .

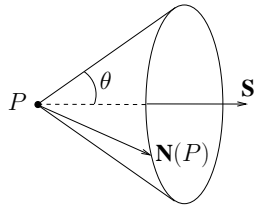


FIGURE 2 – En un point P dont l'image Q n'est ni un point singulier, ni située sur un terminateur, il existe une ambiguïté sur la normale $\mathbf{N}(P)$. Les différentes valeurs possibles de $\mathbf{N}(P)$ forment un cône d'axe \mathbf{S} et d'angle au sommet $\theta = \arccos(i(Q))$.

Modèle différentiel Pour un point image Q non singulier, il n'est pas possible d'attribuer au point P conjugué de Q n'importe laquelle des normales vérifiant l'équation (2). Si on impose à la surface recherchée d'être différentiable, alors $\mathbf{N}(P)$ doit varier continûment d'un point de l'image aux points voisins. Dans ce cas, il vaut mieux utiliser un modèle différentiel de shape-from-shading. Pour ce faire, utilisons un repère tridimensionnel $(Oxyz)$ dont l'axe Oz coïncide avec l'axe optique de l'appareil photographique (cf. figure 3). Un point P de la surface de la scène a pour coordonnées (x, y, z) . Si nous négligeons l'effet de la perspective, et que nous assimilons le grandissement

à 1, alors P a son image en $Q = (x, y)$, c'est-à-dire que l'image de la scène est obtenue par projection orthogonale sur le plan image. Dans ces conditions, le relief de la scène est le graphe $z = u(x, y)$ d'une fonction u , appelée *profondeur*, qui est la nouvelle inconnue du problème. Il est facile d'écrire les coordonnées de $\mathbf{N}(P)$ en fonction des dérivées partielles $p = \partial_x u$ et $q = \partial_y u$ de la fonction u :

$$\mathbf{N}(P) = \frac{1}{\sqrt{1 + p(x, y)^2 + q(x, y)^2}} [-p(x, y), -q(x, y), 1]^\top \quad (3)$$

En utilisant l'expression (3) de $\mathbf{N}(P)$, et en notant $\mathbf{s} = [s_1, s_2, s_3]^\top$, l'équation (2) se réécrit :

$$i(x, y) \sqrt{1 + \|\nabla u(x, y)\|^2} + [s_1, s_2]^\top \cdot \nabla u(x, y) = s_3 \quad (4)$$

où $\nabla u = [\partial_x u, \partial_y u]^\top$ désigne le gradient de u . L'équation (4), qui est une équation aux dérivées partielles du premier ordre, non linéaire, constitue un premier modèle différentiel du shape-from-shading, moins souvent rencontré que la version obtenue avec un *éclairage frontal*, c'est-à-dire lorsque $\mathbf{s} = [0, 0, 1]^\top$. Dans ce cas, l'équation (4) devient :

$$\|\nabla u(x, y)\|^2 = \frac{1}{i(x, y)^2} - 1 \quad (5)$$

Ce nouveau modèle différentiel particulièrement simple du shape-from-shading est la célèbre *équation eikonale*, dont la première mention explicite est due à Bruss [3].

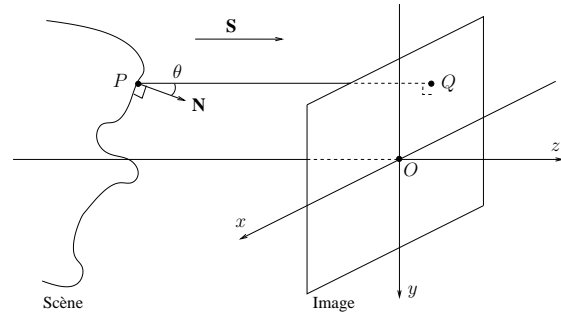


FIGURE 3 – En négligeant la perspective et en assimilant le grandissement transverse à 1, un point P de la scène de coordonnées (x, y, z) a son image au point $Q = (x, y)$ obtenu par projection orthogonale de P sur le plan image.

2 Objectifs et plan de l'article

Pour la grande majorité des points Q d'une image, il existe donc une ambiguïté sur la normale $\mathbf{N}(P)$ au point P de la scène conjugué de Q (cf. paragraphe 1). Cette caractéristique du shape-from-shading traduit son caractère fondamentalement *mal posé*. En particulier, il est impossible de discerner un relief de son « moulage » à la vue d'une seule image. Cette ambiguïté a été judicieusement mise en évidence par Pentland dans [16], grâce à la photographie de gauche de la figure 4, qui montre deux monticules de

condres des îles Hawaii. La photographie de droite est la même que celle de gauche, mais elle a été retournée. On perçoit pourtant un nouveau relief, constitué de deux cratères. En réalité, nous pouvons interpréter le relief de la scène photographiée de deux manières différentes. Cette ambiguïté est appelée l'*ambiguïté concave/convexe* parce que, parmi les deux reliefs possibles, les monticules sont convexes et les cratères sont concaves. L'observateur privilégie un de ces reliefs en faisant l'hypothèse, statistiquement plausible, que « la lumière vient d'en haut ».

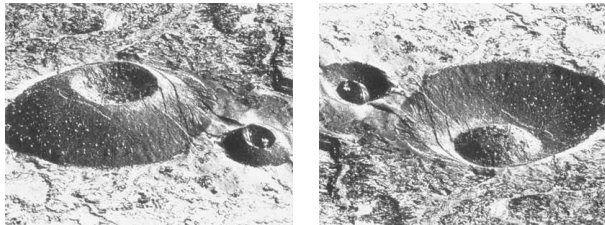


FIGURE 4 – À gauche : photographie de deux monticules de cendres des îles Hawaii. À droite : après retournement, on voit plutôt deux cratères. Cette illustration de l'ambiguïté concave/convexe provient de [16].

L'approche différentielle du shape-from-shading montre que les coordonnées de la normale $\mathbf{N}(P)$ ne sont pas indépendantes, puisqu'elles sont toutes liées aux dérivées partielles de u . Il semble donc que le modèle différentiel (2) soit mieux contraint que le modèle non différentiel (4). Il est vrai que, pour calculer le relief d'une scène à partir d'un champ de normales, il faut que ce champ soit intégrable ou, ce qui revient au même, que son rotationnel soit identiquement nul [6]. Utiliser un modèle différentiel de shape-from-shading revient donc à garantir que le champ de normales soit intégrable, et évite donc d'avoir à l'imposer lors de la phase d'intégration, comme cela est le cas dans [10].

Néanmoins, l'utilisation d'un modèle différentiel ne suffit pas, sans autre hypothèse, à faire du shape-from-shading un problème bien posé. Les premiers résultats d'unicité de la solution nécessitaient tous l'existence d'au moins un point singulier dans l'image et la connaissance d'une condition au bord de type Dirichlet (profondeur connue) ou Neumann (normale connue) le long d'une courbe fermée [3, 1, 15, 2]. Ces premiers résultats, limités aux solutions de classe C^2 , ont ensuite été étendus aux *solutions de viscosité*, qui sont des solutions différentiables presque partout [14, 9, 18]. Plus récemment, des modèles différentiels plus réalistes de shape-from-shading ont été proposés [17, 20, 4, 19].

Dans cet article, nous n'apportons aucun nouveau résultat concernant l'étude du nombre de solutions du shape-from-shading. Nous supposons que les hypothèses du paragraphe 1 sont valides, ce qui permet de simplifier la modélisation, et nous tentons de résoudre le problème par apprentissage artificiel. Très intuitivement, il semble qu'en ayant observé suffisamment de photographies de scènes de reliefs connus, qui constituent donc des *données d'appren-*

tissage, on puisse ensuite utiliser cette connaissance pour reconstruire le relief d'une scène à partir d'une seule photographie, que l'on considère comme une *donnée de test*.

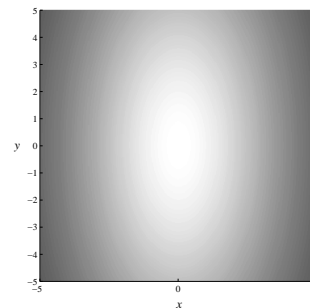


FIGURE 5 – Image de niveau de gris $i_t(x, y) = 1/\sqrt{1 + 16x^2 + 4y^2}$. Cette image provient de [8].

Bien sûr, cette approche est limitée par le caractère mal posé du shape-from-shading : si deux reliefs donnent la même image, il est évidemment impossible de déterminer le « bon » relief à la vue de cette seule image, que ce soit par apprentissage ou par toute autre méthode de résolution. À ce propos, l'ambiguïté concave/convexe n'est pas la seule ambiguïté du shape-from-shading. Dans [8], le cas de l'image de la figure 5, de niveau de gris normalisé $i_t(x, y) = 1/\sqrt{1 + 16x^2 + 4y^2}$, est étudié de près. La figure 6 montre trois exemples de solutions de l'équation eikonale (5) correspondant au niveau de gris $i_t(x, y)$. Le relief de gauche est le graphe de la fonction qui a permis de simuler l'image de la figure 5. Les deux autres reliefs appartiennent à une famille de surfaces dépendant continûment d'un paramètre réel, qui sont toutes solutions de cette même équation eikonale. Dans ces conditions, en un point Q donné de l'image de la figure 5, il n'est pas possible de déterminer la normale au point P conjugué de Q .

Néanmoins, en faisant l'hypothèse que la surface est localement sphérique et convexe, le problème peut devenir *artificiellement bien posé*. L'hypothèse de sphéricité locale a déjà été proposée par Pentland [16], mais dans une version légèrement différente de la nôtre où les sphères ne sont pas contraintes à être convexes, et qui donne des résultats plutôt décevants [5]. L'hypothèse d'une surface localement sphérique et convexe en tout point est très arbitraire, mais il semble statistiquement raisonnable de favoriser les reliefs convexes.

La suite de l'article est organisée comme suit : dans le paragraphe 3, nous introduisons un estimateur des normales qui découle des hypothèses de sphéricité et de convexité locales de la surface ; dans le paragraphe 4, nous montrons que cet estimateur équivaut à l'analyse en composantes principales (ACP) d'images d'hémisphères convexes ; en outre, nous illustrons notre méthode de résolution sur plusieurs exemples ; enfin, le paragraphe 5 résume les contributions de notre travail et propose un prolongement.

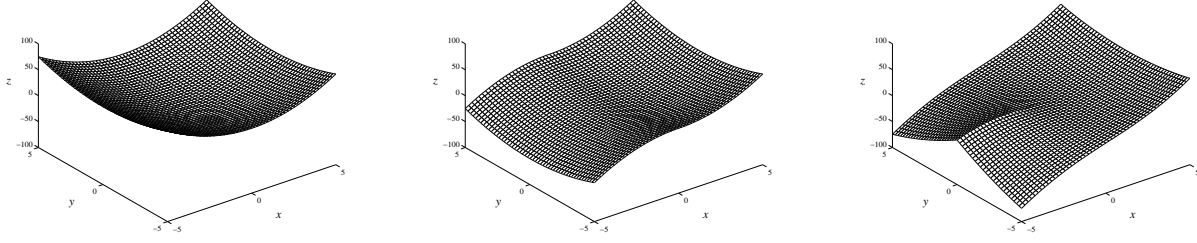


FIGURE 6 – Graphes de trois fonctions u solutions de l'équation eikonale (5) correspondant au niveau de gris $i_t(x, y)$ de l'image de la figure 5. Ces images proviennent de [8].

3 Estimateur des normales

Commençons par rappeler les hypothèses qui nous ont permis d'obtenir le modèle très simple de shape-from-shading du paragraphe 1. Nous nous intéressons à une scène différentiable, lambertienne, d'albédo égal à 1 (couleur blanche uniforme), située suffisamment loin de l'appareil photographique pour que la perspective soit négligeable. Cette scène est éclairée par un faisceau uniforme, parallèle à l'axe optique (éclairage frontal). Le modèle non différentiel à utiliser sous ces hypothèses est l'équation (2). Le modèle différentiel correspondant est l'équation eikonale (5). Nous nous intéressons à la famille d'hémisphères \mathcal{H}_r paramétrées par le rayon $r \in \mathbb{R}^+$, d'équations $z = u_r(x, y)$, avec :

$$u_r(x, y) = \sqrt{r^2 - x^2 - y^2} \quad (6)$$

L'équation (5) permet de calculer le niveau de gris i_r de l'image \mathcal{I}_r de \mathcal{H}_r :

$$i_r(x, y) = \sqrt{1 - \frac{x^2 + y^2}{r^2}} \quad (7)$$

Le seul point singulier de l'image \mathcal{I}_r se trouve à l'origine $O = (0, 0)$, tandis que le contour de la silhouette, c'est-à-dire le cercle d'équation $x^2 + y^2 = r^2$, constitue un terminateur.

Supposons que nous disposions d'une image \mathcal{I} de niveau de gris $i(x, y)$ ayant été obtenue dans les mêmes conditions opératoires que précédemment, et que le relief de la scène photographiée soit inconnu. Bien entendu, l'inconnue u doit vérifier l'équation eikonale (5). Néanmoins, pour un point Q_0 de \mathcal{I} de niveau de gris $i_0 \in [0, 1]$, il existe une infinité de points (x, y) de la famille d'images $(\mathcal{I}_r)_{r \in \mathbb{R}^+}$, formant un ensemble \mathcal{E}_{i_0} , tels que $i_r(x, y) = i_0$. D'après (7), ces points sont tels que :

$$x^2 + y^2 = r^2(1 - i_0^2) \quad (8)$$

Afin de rendre le problème (artificiellement) bien posé, cherchons dans \mathcal{E}_{i_0} les points tels que $\nabla i_r(x, y)$ soit égal au gradient $[\partial_x i_0, \partial_y i_0]^\top$ du niveau de gris de l'image \mathcal{I} en Q_0 . De (7), nous déduisons facilement l'expression de $\nabla i_r(x, y)$, qui est définie en tout point non singulier :

$$\nabla i_r(x, y) = -\frac{1}{r \sqrt{r^2 - x^2 - y^2}} [x, y]^\top \quad (9)$$

Nous cherchons donc les points (x, y) de \mathcal{E}_{i_0} tels que :

$$[x, y]^\top = -r \sqrt{r^2 - x^2 - y^2} [\partial_x i_0, \partial_y i_0]^\top \quad (10)$$

Les équations (8) et (10) forment un système de trois équations non linéaires à trois inconnues x, y et r . Nous déduisons de (8) que $r^2 - x^2 - y^2 = r^2 i_0^2$, ce qui nous permet de réécrire (10) :

$$[x, y]^\top = -r^2 i_0 [\partial_x i_0, \partial_y i_0]^\top \quad (11)$$

De (8) et de (11), nous tirons l'expression suivante de la solution en r :

$$\hat{r}^2 = \frac{1 - i_0^2}{i_0^2 (\partial_x i_0^2 + \partial_y i_0^2)} \quad (12)$$

En reportant cette expression dans (11), nous trouvons la solution en (x, y) :

$$[\hat{x}, \hat{y}]^\top = -\frac{1 - i_0^2}{i_0 (\partial_x i_0^2 + \partial_y i_0^2)} [\partial_x i_0, \partial_y i_0]^\top \quad (13)$$

Nous voyons donc qu'il existe un seul rayon \hat{r} , et un seul point (\hat{x}, \hat{y}) de l'image $\mathcal{I}_{\hat{r}}$, qui vérifient les contraintes (8) et (10). Or l'unicité de la solution est la caractéristique essentielle d'un problème bien posé. Nous voici maintenant en mesure de proposer un estimateur des normales. En effet, il suffit d'attribuer au point P_0 conjugué de Q_0 la normale en (\hat{x}, \hat{y}) à l'hémisphère $\mathcal{H}_{\hat{r}}$. De (6), (12) et (13), nous tirons facilement le gradient $[\hat{p}, \hat{q}]^\top = \nabla u_{\hat{r}}(\hat{x}, \hat{y})$:

$$[\hat{p}, \hat{q}]^\top = \frac{\sqrt{1 - i_0^2}}{i_0 \sqrt{\partial_x i_0^2 + \partial_y i_0^2}} [\partial_x i_0, \partial_y i_0]^\top \quad (14)$$

Remarquons que cet estimateur n'est pas défini en un point Q_0 tel que $i_0 = 0$, ce qui est normal car, d'après (5), la pente $\|\nabla u(x_0, y_0)\|$ est infinie en un tel point. Le cas où $[\partial_x i_0, \partial_y i_0]^\top = [0, 0]^\top$ peut sembler plus problématique, mais ce cas ne survient, en pratique, qu'aux points singuliers. Il s'avère donc que l'estimateur proposé est défini partout où il existe une ambiguïté sur la normale, c'est-à-dire juste là où il le fallait.

Une autre propriété de cet estimateur des normales est que l'image \mathcal{I}_s , simulée à partir des valeurs estimées \hat{p} et \hat{q} ,

coïncide avec l'image originale \mathcal{I} . En effet, nous tirons de (14) :

$$\hat{p}^2 + \hat{q}^2 = \frac{1}{i_0^2} - 1 \quad (15)$$

ce qui montre bien, d'après l'équation (5), que le niveau de gris de \mathcal{I}_s est le même que celui de \mathcal{I} .

Cependant, trouver des valeurs de p et q telles que l'image simulée à partir de l'équation (5) coïncide avec l'image de départ ne constitue pas, en soi, un résultat remarquable. En effet, nous savons qu'il existe, sauf en certains points, une infinité de couples (p, q) qui vérifient l'équation (5). Un résultat vraiment remarquable serait que le champ de normales obtenu soit intégrable. Or, pour un champ de normales intégrable, le théorème de Schwarz impose l'égalité :

$$\partial_y \hat{p} = \partial_x \hat{q} \quad (16)$$

Il est facile de montrer que les expressions (14) de \hat{p} et \hat{q} ne vérifient pas l'égalité (16). Il nous faudra donc coupler l'estimateur (14) avec une méthode d'intégration permettant au champ de normales de ne pas être intégrable.

Exemple L'estimateur des normales (14) a été appliqué au vase de la figure 7-a, dont l'image 7-b a été simulée en utilisant l'équation (5). La méthode d'intégration utilisée [6] est une variante de la méthode de Horn et Brooks [13]. Le relief obtenu, qui est représenté sur la figure 7-c, est qualitativement acceptable. Nous remarquons toutefois que le col du vase, où la courbure gaussienne est négative, est plutôt mal reconstruit. Cela semble normal, puisque l'hypothèse de sphéricité locale est clairement fautive en ce point. Plus précisément, la racine carrée de l'écart quadratique moyen (RMSE) entre la profondeur estimée et la profondeur réelle vaut $|\Delta u|_2 = 0,70$, à comparer aux valeurs 1,00, 0,38 et 0,62 obtenues par les trois méthodes de résolution « classiques » testées dans [7] sur le même exemple.

4 Estimation des normales par ACP

Dans ce paragraphe, nous abordons le même problème que dans le paragraphe 3 en utilisant une voie apparemment différente, mais nous verrons un peu plus loin que ces deux voies sont analogues.

Phase d'apprentissage Intéressons-nous pour commencer aux deux photographies de la figure 8, qui ont été prises dans des conditions opératoires identiques censées reproduire les hypothèses simplificatrices du paragraphe 1. Les deux objets ont été peints avec la même peinture blanc mat. Utilisons la terminologie de l'apprentissage artificiel. L'image 8-a, qui correspond à un relief connu, constitue une *donnée d'apprentissage*, tandis que l'image 8-b, dont nous ne connaissons pas le relief, sera considérée comme une *donnée de test*. Plus précisément, nous découpons l'image 8-a en imassettes disjointes \mathcal{I}_a de taille 3×3 (de plus grandes imassettes ont été testées, mais cela augmente le temps de calcul sans qu'il y ait une réelle amélioration des résultats). Ce faisant, chaque imasette \mathcal{I}_a peut être associée à la normale \mathbf{N} de son pixel central.

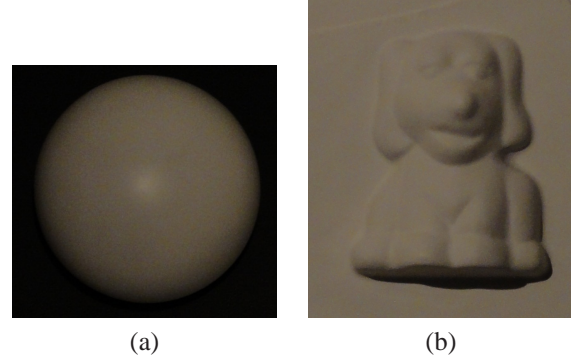


FIGURE 8 – (a) Donnée d'apprentissage. (b) Donnée de test.

Nous effectuons ensuite l'analyse en composantes principales des imassettes \mathcal{I}_a , qui sont des points de \mathbb{R}^9 . La base de vecteurs propres est constituée de 9 imassettes qui sont représentées sur la figure 9. Nous voyons immédiatement que le premier de ces vecteurs propres est le masque de convolution d'un filtre moyennneur, tandis que les deux vecteurs propres suivants (situés sur la première colonne de la figure 9) ressemblent beaucoup aux masques de convolution caractéristiques des filtres de Sobel, qui permettent de calculer le gradient du niveau de gris. La projection des imassettes sur le sous-espace engendré par les vecteurs propres associés aux $n = 3$ plus grandes valeurs propres permet de représenter les imassettes \mathcal{I}_a sous la forme d'un nuage de points de \mathbb{R}^3 . Il semble donc que les $n = 3$ composantes principales d'une imasette \mathcal{I}_a soient le niveau de gris moyen I_a de \mathcal{I}_a , ainsi que ses deux dérivées partielles d'ordre 1.

En guise de vérification, nous avons reporté sur la figure 10 trois nuages de points correspondant aux $n = 3$ composantes principales des imassettes de trois hémisphères \mathcal{H}_r (cf. paragraphe 3). Chacun de ces nuages semble former une nappe dans \mathbb{R}^3 , et les trois nappes semblent être tangentes entre elles au point $(1, 0, 0)$. Cela s'explique bien si nous interprétons ces $n = 3$ composantes principales comme le niveau de gris et ses deux dérivées partielles d'ordre 1. En effet, les équations (6) et (7) montrent que $u_r(x, y) = r i_r(x, y)$. En dérivant, il vient :

$$\|\nabla u_r(x, y)\|^2 = r^2 \|\nabla i_r(x, y)\|^2 \quad (17)$$

Des équations (17) et (5), on déduit donc :

$$i_r(x, y) = \frac{1}{\sqrt{1 + r^2 \|\nabla i_r(x, y)\|^2}} \quad (18)$$

Cette équation est bien conforme aux observations que le graphique de la figure 10 nous a inspirées. En effet, (18) montre que $i_r(x, y) = 1$ si $\nabla i_r(x, y) = [0, 0]^T$, quelle que soit la valeur de r . De plus, la nappe de couleur verte, qui est la plus « resserrée », correspond à la plus grande valeur de r , ce qui est bien cohérent avec l'équation (18).

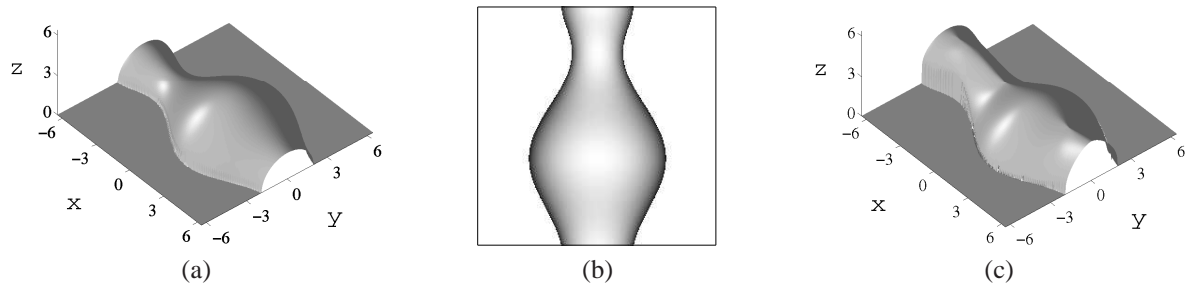


FIGURE 7 – (a) Surface représentant un vase. (b) Image de synthèse correspondante. (c) Relief reconstruit en appliquant l’estimateur des normales (14) à l’image (b), puis en intégrant ce champ de normales par la méthode décrite dans [6].

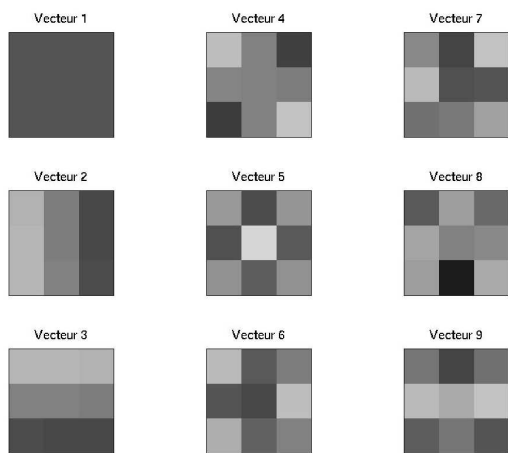


FIGURE 9 – Base de vecteurs propres des imagettes \mathcal{I}_a de taille 3×3 provenant de l’image de la figure 8-a.

Phase de test Nous découpons maintenant l’image de la figure 8-b en imagettes \mathcal{I}_t de même taille 3×3 que les imagettes \mathcal{I}_a , puis nous projetons ces points sur le sous-espace propre correspondant aux $n = 3$ composantes principales des données d’apprentissage. Les deux nuages sont représentés sur la figure 11 : les données d’apprentissage en rouge (on reconnaît la forme caractéristique du nuage associé à une hémisphère), les données de test en bleu. Nous attribuons au pixel central de chaque imagerie \mathcal{I}_t la normale \mathbf{N} associée à l’imagette \mathcal{I}_a la plus proche de \mathcal{I}_t .

La figure 12 montre deux images simulées à partir du champ de normales estimées par ce procédé. À gauche, c’est l’éclairage original (quasi-frontal) qui est utilisé. Nous ne sommes pas surpris de retrouver l’image originale, à un facteur près cependant, qui vient de ce que l’image de la figure 8-b n’est pas normalisée. Pour la figure de droite, nous utilisons un autre éclairage qui met en évidence le principal défaut de notre méthode de résolution : dans la mesure où les normales estimées ne sont pas intégrables,

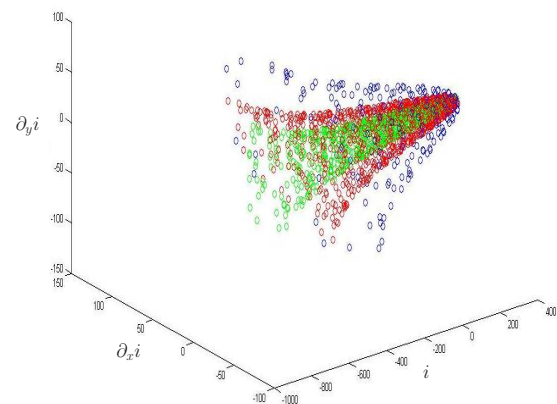


FIGURE 10 – Trois nuages de points correspondant aux $n = 3$ composantes principales des imagettes de trois hémisphères \mathcal{H}_r . Chaque nuage forme une nappe de \mathbb{R}^3 .

elles sont faiblement corrélées, ce qui se traduit par l’apparition d’un bruit important.

Bien entendu, il serait tentant d’utiliser comme ensemble d’apprentissage non pas une, mais plusieurs images d’hémisphères. Cela permettrait d’avoir un nuage de points d’apprentissage réellement tridimensionnel, contrairement au nuage rouge de la figure 11, qui ne constitue qu’une nappe de \mathbb{R}^3 . Néanmoins, dans la mesure où il ne s’agit pas d’images de synthèse mais de photographies, nous ne disposons que de l’image de la figure 8-a comme donnée d’apprentissage. L’utilisation d’une interpolation sophistiquée de type TPS (*Thin Plate Spline*) n’a pas permis d’améliorer les résultats de manière significative. Il semble que ce soit plutôt en densifiant le nuage de points d’apprentissage qu’une amélioration puisse survenir.

D’autres tests nous ont montré que la valeur $n = 3$ était optimale. Une valeur de n plus petite occasionne une dégradation notable des résultats, alors qu’une valeur plus élevée ne les améliore pas, tout en faisant croître inutilement le temps de calcul. Cela n’est pas très surprenant, car

l'approche que nous avons adoptée, par deux voies différentes, consiste à supposer que la surface est localement sphérique : dans le paragraphe 3, cette hypothèse est explicite ; dans le paragraphe 4, elle est implicite, puisque les données d'apprentissage sont des images d'hémisphères. Or, l'estimateur des normales (14) ne dépend que du niveau de gris et de son gradient, c'est-à-dire de $n = 3$ variables.

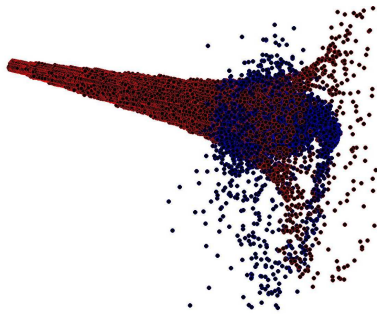


FIGURE 11 – Deux nuages de points de \mathbb{R}^3 : données d'apprentissage en rouge ; données de test en bleu.

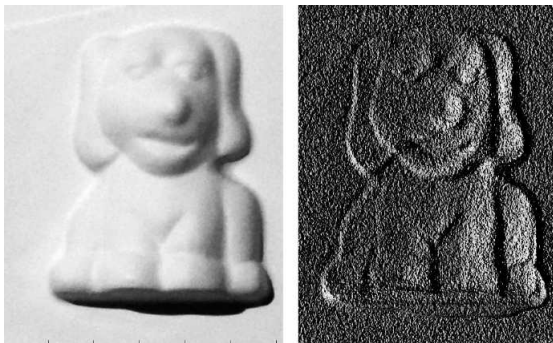


FIGURE 12 – Deux images simulées calculées à partir du champ de normales estimé par ACP, à partir de l'image de la figure 8-b. À gauche, l'éclairage réel a été utilisé. À droite, un nouvel éclairage montre que le champ de normales estimé n'est pas intégrable.

Problème des ombres On peut s'interroger sur la pertinence de notre approche, qui consiste à avoir recours à une ACP, puisque nous avons montré l'équivalence avec l'estimateur des normales présenté dans le paragraphe 3. Cela ne saute pas forcément aux yeux, mais cette deuxième voie permet de généraliser très facilement la première : par exemple, au lieu de chercher à adapter l'estimateur (14) à un autre éclairage, l'analyse en composantes principales fait cela très naturellement, pourvu que l'on dispose d'images d'hémisphères obtenues sous l'éclairage voulu. Cela nous a permis de mettre en évidence un autre défaut de cette nouvelle méthode de résolution du shape-from-shading : elle n'est pas conçue pour traiter correctement les ombres. L'exemple de la figure 13 montre deux images

de synthèse éclairées de biais, de telle sorte que l'image de droite, qui constitue la donnée de test, comporte une ombre. Trois paires de flèches colorées matérialisent la « similarité photométrique » (même niveau de gris et même gradient du niveau de gris) entre trois points de l'image de test 13-b et trois points de l'image d'apprentissage 13-a. Manifestement, la paire bleue est erronée à cause de l'ombre. Par intégration du champ de normales estimé par ACP, nous obtenons le relief de la figure 14, qui ressemble globalement à une gaussienne. Ce relief a été volontairement inversé, afin de montrer l'effet de l'ombre, qui se traduit par une petite excroissance.

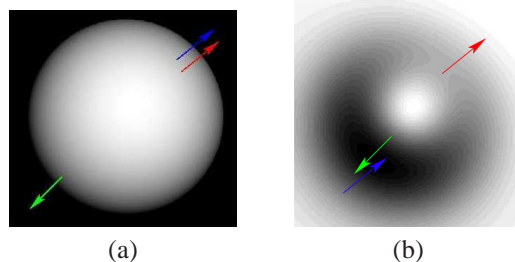


FIGURE 13 – Éclairage de biais : l'image de test (b) comporte une ombre, mais pas l'image d'apprentissage (a).

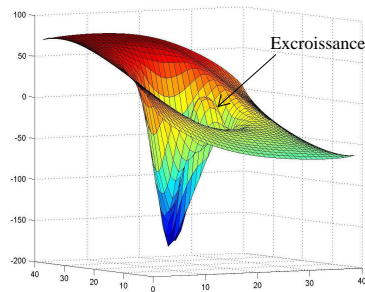


FIGURE 14 – Relief reconstruit par apprentissage sur l'image 13-a, à partir de l'image 13-b.

5 Conclusion et perspectives

Dans cet article, nous avons tenté de résoudre le shape-from-shading, dans le cas le plus simple où son modèle se ramène à l'équation eikonale. Nous avons proposé un estimateur des normales qui traduit une hypothèse très forte, à savoir que la surface est localement sphérique et convexe. Nous avons montré que cet estimateur pouvait être interprété comme le résultat de l'analyse en composantes principales (ACP) d'une famille d'images d'hémisphères. Enfin, nous l'avons testé sur un certain nombre d'images de synthèse et d'images réelles. Les principales qualités de l'estimateur des normales proposé sont qu'il est facile à mettre en œuvre et que l'image simulée à partir des normales estimées coïncide avec l'image initiale. Son princi-

pal défaut est de ne pas garantir que le champ de normales estimé soit intégrable.

Une extension naturelle de notre travail consiste à l'adapter à la stéréophotométrie, qui est une version multi-oculaire du shape-from-shading dans laquelle plusieurs photographies de la même scène sont prises sous le même angle, mais sous différents éclairages [21]. Cette technique, qui connaît actuellement une grande vogue, permet d'obtenir des reconstructions 3D précises [6], pour la bonne et simple raison que le problème est généralement bien posé. Or, la grande majorité des méthodes de résolution de la stéréophotométrie commencent par estimer le champ de normales de la scène à reconstruire. C'est pourquoi nous pensons que l'adaptation de notre estimateur des normales à la stéréophotométrie pourrait être pertinente. D'ailleurs, certaines techniques d'apprentissage ont déjà été utilisées avec succès pour résoudre la stéréophotométrie [11].

Références

- [1] A. BLAKE, A. ZISSERMAN et G. KNOWLES. « Surface descriptions from stereo and shading ». *Image and Vision Computing*, volume 3, numéro 4, pages 183–191, novembre 1985. 3
- [2] M. J. BROOKS, W. CHOJNACKI et R. KOZERA. « Shading Without Shape ». *Quarterly of Applied Mathematics*, volume 50, numéro 1, pages 27–38, mars 1992. 3
- [3] A. R. BRUSS. « The Eikonal Equation : Some Results Applicable to Computer Vision ». *Journal of Mathematical Physics*, volume 23, numéro 5, pages 890–896, mai 1982. 2, 3
- [4] F. COURTEILLE, A. CROUZIL, J.-D. DUROU et P. GURDJOS. « Shape from shading en conditions réalistes d'acquisition photographique ». In *Actes du 14^{ème} Congrès RFIA (volume II)*, pages 925–934, Toulouse, France, janvier 2004. 3
- [5] P. DANIEL. « Peut-on extraire le relief d'une seule image ? ». Thèse de doctorat, Université Paul Sabatier, Toulouse, France, janvier 2000. 3
- [6] J.-D. DUROU, J.-F. AUJOL et F. COURTEILLE. « Integration of a Normal Field in the Presence of Discontinuities ». In *Proceedings of the 7th EMMCVPR*, volume 5681 de *LNCS*, pages 261–273, Bonn, Allemagne, août 2009. 3, 5, 6, 8
- [7] J.-D. DUROU, M. FALCONE et M. SAGONA. « Numerical Methods for Shape-from-shading : A New Survey with Benchmarks ». *Computer Vision and Image Understanding*, volume 109, numéro 1, pages 22–43, janvier 2008. 5
- [8] J.-D. DUROU et D. PIAU. « Ambiguous Shape from Shading with Critical Points ». *Journal of Mathematical Imaging and Vision*, volume 12, numéro 2, pages 99–108, avril 2000. 3, 4
- [9] M. FALCONE et M. SAGONA. « An algorithm for the global solution of the Shape-from-Shading model ». In *Proceedings of the 9th ICIAP (volume I)*, volume 1310 de *LNCS*, pages 596–603, Florence, Italie, septembre 1997. 3
- [10] R. T. FRANKOT et R. CHELLAPPA. « A Method for Enforcing Integrability in Shape from Shading Algorithms ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 10, numéro 4, pages 439–451, juillet 1988. 3
- [11] A. HERTZMANN et S. M. SEITZ. « Shape and Materials by Example : A Photometric Stereo Approach ». In *Proceedings of the 9th ICCV (volume I)*, pages 533–540, Nice, France, octobre 2003. 8
- [12] B. K. P. HORN. « *Shape from Shading : a Method for Obtaining the Shape of a Smooth Opaque Object from One View* ». PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, États-Unis, novembre 1970. 1
- [13] B. K. P. HORN et M. J. BROOKS. « The Variational Approach to Shape From Shading ». *Computer Vision, Graphics, and Image Processing*, volume 33, numéro 2, pages 174–208, février 1986. 5
- [14] P.-L. LIONS, E. ROUY et A. TOURIN. « Shape-from-Shading, viscosity solutions and edges ». *Numerische Mathematik*, volume 64, numéro 3, pages 323–353, mars 1993. 3
- [15] J. OLIENSIS. « Existence and Uniqueness in Shape from Shading ». In *Proceedings of the 10th ICPR (volume I)*, pages 341–345, Atlantic City, New Jersey, États-Unis, juin 1990. 3
- [16] A. P. PENTLAND. « Local Shading Analysis ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 6, numéro 2, pages 170–187, mars 1984. 2, 3
- [17] E. PRADOS et O. FAUGERAS. « “Perspective Shape from Shading” and Viscosity Solutions ». In *Proceedings of the 9th ICCV (volume II)*, pages 826–831, Nice, France, octobre 2003. 3
- [18] E. PRADOS et O. FAUGERAS. « Rôle clé de la Modélisation en “Shape From Shading” ». In *Actes 15^{ème} Congrès RFIA*, Tours, France, janvier 2006. 3
- [19] E. PRADOS, O. FAUGERAS et F. CAMILLI. « Shape from Shading : a well-posed problem ? ». Rapport de Recherche 5297, INRIA, Sophia Antipolis, France, août 2004. 3
- [20] A. TANKUS, N. SOCHEN et Y. YESHURUN. « A New Perspective [on] Shape-from-Shading ». In *Proceedings of the 9th ICCV (volume II)*, pages 862–869, Nice, France, octobre 2003. 3
- [21] R. L. WILDEY. « Radarclinometry for the Venus Radar Mapper ». *Photogrammetric Engineering and Remote Sensing*, volume 52, numéro 1, pages 41–50, janvier 1986. 8