HAL
open science

Résolution de la stéréophotométrie par apprentissage

Jean-Denis Durou, Yvain Quéau, Vincent Charvillat

To cite this version:

Jean-Denis Durou, Yvain Quéau, Vincent Charvillat. Résolution de la stéréophotométrie par apprentissage. RFIA 2012 (Reconnaissance des Formes et Intelligence Artificielle), Jan 2012, Lyon, France. hal-00656548

HAL Id: hal-00656548

https://hal.science/hal-00656548v1

Submitted on 17 Jan 2012

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 de la stéréophotométrie par apprentissage

Photometric Stereo Using Machine Learning Techniques

Jean-Denis DUROU Yvain QUEAU Vincent CHARVILLAT

IRIT, UMR CNRS 5505, Toulouse
durou@irit.fr yvain.queau@gmail.com
Vincent.Charvillat@enseeiht.fr

Résumé

Nous montrons dans cet article comment résoudre la stéréophotométrie par apprentissage. L'idée de base consiste à insérer une sphère dans la scène dont on souhaite retrouver le relief, et à recouvrir la sphère et la scène d'une même peinture. Les images de la sphère servant de données d'apprentissage, il n'est pas nécessaire de connaître la réflectance de la scène pour mettre en œuvre la stéréophotométrie. De plus, les ombres propres sont considérées comme des données utiles. Nous montrons que l'analyse en composantes principales (ACP) permet, d'une part, d'obtenir les mêmes résultats que Hertzmann et al. [8] tout en réduisant le temps de calcul, et d'autre part, de détecter les ombres portées de la scène, pour peu que la sphère en soit dépourvue. Nous montrons aussi comment généraliser notre approche à des scènes non préalablement peintes, mais de nature photométrique uniforme. Toutes ces contributions sont validées sur images de synthèse et images réelles.

Mots Clef

Reconstruction 3D, stéréophotométrie, apprentissage.

Abstract

In this paper, we solve photometric stereo using machine learning. The basic idea consists in inserting a sphere in the scene to be reconstructed, and in covering both the sphere and the scene with the same paint. The images of the sphere being considered as learning data, the knowledge of the scene reflectance is not required any more to implement photometric stereo. In addition, the self-shadows become meaningful. We show how the Principal Components Analysis (PCA) allows us, on one hand, to reach the same results as Hertzmann et al. [8] but faster, and on the other hand, to detect the cast-shadows of the scene in the case where no such shadow occurs on the sphere. We also show how to extend our approach to non-painted scenes, but of uniform photometric nature. All these contributions are validated on synthetic, as well as on real images.

Keywords

3D-reconstruction, photometric stereo, machine learning.

1 Introduction

Parmi les différentes techniques de reconstruction 3D, le shape-from-shading a la réputation d'être inutilement difficile à mettre en œuvre, en regard de ses performances décevantes [5]. Même si l'unicité de la solution peut être garantie sous certaines hypothèses, l'ambiguïté concave/convexe lui confère un caractère fondamentalement mal posé. Une solution pour y remédier consiste à utiliser plusieurs images du même objet vu sous le même angle, mais sous différents éclairages. L'idée est utilisée intuitivement en spéléologie : le déplacement d'une lampe permet d'explorer le relief d'une paroi sans se déplacer soi-même, juste en provoquant des variations de clarté. Cette technique, appelée stéréophotométrie (*photometric stereo*), est due à Woodham [15]. Elle connaît actuellement une certaine vogue, car elle permet de retrouver les détails fins des objets, contrairement aux autres techniques de reconstruction 3D [6].

Dans cet article, nous nous intéressons à une méthode de résolution de la stéréophotométrie initialement suggérée par Woodham dans [14], qui consiste non pas à utiliser un modèle théorique de réflectance, mais à apprendre la réflectance à partir d'images d'un objet de relief connu. Nous montrons comment améliorer les résultats obtenus par Hertzmann et al. dans [8].

Dans le paragraphe 2, nous rappelons les équations de la stéréophotométrie et décrivons les principales méthodes de résolution, en précisant leurs limites. Dans le paragraphe 3, nous donnons une illustration graphique de la résolution par apprentissage, et nous montrons qu'une analyse en composantes principales (ACP) permet d'accélérer la méthode de résolution proposée dans [8], sans perdre en précision. Dans le paragraphe 4, nous montrons que l'ACP permet en outre de tenir compte des ombres portées (les ombres propres étant déjà prises en compte lors de l'apprentissage) et de la texture. Toutes ces contributions sont validées sur images de synthèse et sur images réelles. Le paragraphe 5 résume nos contributions et suggère de futures voies à explorer.

2 Méthodes de résolution de la stéréophotométrie : bref état de l'art

2.1 Résolution classique

Supposons que n vues d'une même scène soient prises sous le même angle, mais sous n éclairages différents. Supposons également que la surface de la scène soit lambertienne. Sous ces hypothèses, le niveau de gris $I_i(x, y)$ au point (x, y) de la $i^{\text{ème}}$ image s'écrit :

$$I_i(x, y) = \rho(x, y)\, \mathbf{S}_i \cdot \mathbf{N}(x, y), \quad i \in [1, n] \tag{1}$$

où l'égalité est en fait une relation de proportionnalité. Dans (1), le vecteur $\mathbf{S}_i$ est le vecteur d'éclairage de la $i^{\text{ème}}$ image (en norme et en direction, de sens opposé au sens effectif de la lumière) et $\rho(x, y)$ et $\mathbf{N}(x, y)$ désignent, respectivement, l'albédo et la normale unitaire sortante au point de la scène conjugué de (x, y). En posant $\mathbf{M}(x, y) = \rho(x, y)\, \mathbf{N}(x, y)$, l'équation (1) se réécrit :

$$I_i(x, y) = \mathbf{S}_i \cdot \mathbf{M}(x, y), \quad i \in [1, n] \tag{2}$$

La résolution en $\mathbf{M}(x, y)$ du système linéaire (2) peut être menée en chaque point (x, y) indépendamment des points voisins. Elle est donc parallélisable. Le système (2) s'écrit matriciellement :

$$\mathbf{I}(x, y) = \mathbf{S}\, \mathbf{M}(x, y), \tag{3}$$

où $\mathbf{I}(x, y)$ désigne le vecteur des n niveaux de gris au point (x, y), et $\mathbf{S}$ la matrice $n \times 3$ dont la $i^{\text{ème}}$ ligne est formée par les composantes du $i^{\text{ème}}$ éclairage $\mathbf{S}_i$. Si cette matrice est de rang 3, c'est-à-dire si les vecteurs $\mathbf{S}_i$ ne sont pas tous coplanaires, alors le système (3) admet une solution approchée unique :

$$\mathbf{M}^*(x, y) = \mathbf{S}^+\, \mathbf{I}(x, y), \tag{4}$$

où $\mathbf{S}^+$ désigne la matrice pseudo-inverse de $\mathbf{S}$. Les estimations de $\rho^*(x, y)$ et $\mathbf{N}^*(x, y)$ se déduisent aisément de (4) :

$$\begin{cases} \rho^*(x, y) = \|\mathbf{M}^*(x, y)\| \\ \mathbf{N}^*(x, y) = \dfrac{\mathbf{M}^*(x, y)}{\|\mathbf{M}^*(x, y)\|} \end{cases} \tag{5}$$

Il reste à calculer le relief à partir des normales estimées, ce qui peut être plus compliqué qu'il n'y paraît [4].
Cette démarche est celle qui a été proposée originellement par Woodham pour résoudre le problème de la stéréophotométrie [15]. Elle permet d'obtenir des résultats de très bonne qualité [4]. En utilisant des sources lumineuses colorées non coplanaires, et en filtrant les images recueillies selon les fréquences utilisées (en l'occurrence : rouge, vert et bleu), Hernandez et al. ont même montré qu'il était possible d'appliquer la stéréophotométrie en temps réel à un flux vidéo d'objets déformables tels que des vêtements [6].

2.2 Variantes de la résolution classique

La simplicité de l'expression (4) montre que la résolution classique est facile à mettre en œuvre et rapide. En outre, aucune hypothèse particulière n'est faite sur l'albédo, dans la mesure où elle en fournit également une estimation. En ce sens, il s'agit vraiment d'une technique de scannage 3D, en comparaison des autres techniques de reconstruction 3D qui ne fournissent que le relief de la scène à reconstruire. Néanmoins, elle repose sur les hypothèses suivantes, qui sont plus ou moins réalistes : on doit disposer de $n \geq 3$ éclairages non coplanaires et connaître ces éclairages avec précision ; la surface de l'objet doit être lambertienne ; enfin, le modèle lambertien doit être valide en tout point, c'est-à-dire qu'il ne doit y avoir ni ombres ni réflexions mutuelles. Un certain nombre de méthodes de résolution ont donc été proposées afin de prendre en compte des situations plus difficiles où l'une ou l'autre de ces hypothèses ne serait pas vérifiée.

Entre le shape-from-shading ($n = 1$) et la stéréophotométrie usuelle ($n \geq 3$), le cas où l'on ne dispose que de $n = 2$ images fait ressortir deux différences majeures par rapport à l'approche classique [10, 11] : l'albédo doit être connu et l'unicité de la normale n'est pas garantie. Par ailleurs, l'utilisation d'harmoniques sphériques a permis de montrer que, sous l'hypothèse lambertienne, il n'est pas absolument indispensable de connaître les éclairages avec précision, ni même de supposer que chaque éclairage soit parallèle et uniforme [2]. D'autre part, deux stratégies ont été proposées pour remettre en question l'hypothèse lambertienne (1), qui n'est pas acceptable lorsqu'il s'agit de reconstruire des scènes réelles : utiliser un modèle de réflectance plus sophistiqué [1] ou retirer de chaque image les points où se traduit le comportement spéculaire de la surface [13]. Cette dernière solution a également été proposée pour traiter les ombres. Alors que l'ombre propre (*self-shadow*) est facile à caractériser localement, en remplaçant dans (1) le produit scalaire $\mathbf{S}_i \cdot \mathbf{N}(x, y)$ par $\max\{\mathbf{S}_i \cdot \mathbf{N}(x, y), 0\}$, il n'en va pas de même des ombres portées (*cast-shadows*). Néanmoins, comme il est généralement difficile de distinguer ces deux types d'ombres, l'idée de supprimer les niveaux de gris les plus faibles a été évoquée [3]. Bien sûr, la suppression d'un niveau de gris devient compliquée dans le cas où $n = 3$, puisque cela ramène le problème à la situation, déjà mentionnée ci-dessus, où $n = 2$ [7].

2.3 Résolution par apprentissage

La résolution classique et ses variantes fournissent des résultats de bonne qualité, mais aucune des méthodes citées ne permet de résoudre le problème dans un cadre vraiment général. La solution pourrait venir des techniques d'apprentissage, comme l'a d'ailleurs suggéré Woodham dès ses premiers travaux [14]. Cette idée a été reprise récemment [8, 9]. Sachant qu'elle constitue la base de notre travail, il nous faut maintenant la présenter plus en détail.

3 Résolution de la stéréophotométrie par apprentissage

3.1 Discussion des hypothèses

La résolution de la stéréophotométrie par apprentissage peut être illustrée à l'aide des images de synthèse de la figure 1. Les images du haut montrent une sphère vue sous $n = 5$ éclairages différents : elles constituent les *données d'apprentissage*. Les images du bas, qui montrent un vase (dont on souhaite prédire les normales) vu sous les mêmes éclairages, constituent les *données de test*. On suppose que chaque surface est vue exactement sous le même angle dans les cinq images. Cette hypothèse usuelle en stéréophotométrie n'est pas rigoureusement nécessaire, mais elle est requise en pratique afin d'éviter l'étape fastidieuse de mise en correspondance entre images.

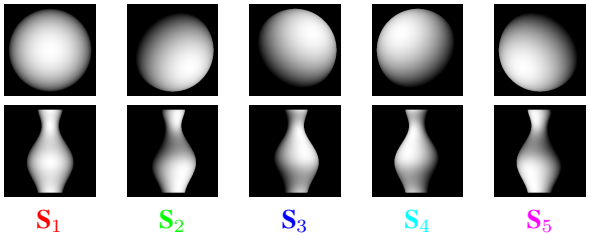

FIGURE 1 – Images de synthèse (de taille 256×256) d'une sphère et d'un vase vus sous $n = 5$ éclairages différents.

Les éclairages utilisés ici sont parallèles et uniformes, donc représentables par des vecteurs $\mathbf{S}_1$, $\mathbf{S}_2$, $\mathbf{S}_3$, $\mathbf{S}_4$ et $\mathbf{S}_5$ (les couleurs associées à ces vecteurs nous serviront plus loin), mais il suffirait que tous les points des deux surfaces possédant la même normale présentent la même clarté pour que la méthode fonctionne.

Les deux surfaces sont supposées lambertiennes, mais nous aurions pu tout aussi bien utiliser une autre réflectance sans que cela change la méthode. En revanche, elles doivent avoir le même albédo uniforme. Nous montrerons dans le paragraphe 4 comment relâcher cette contrainte.

Un autre avantage notable de la résolution par apprentissage est que les ombres propres ne posent pas de problème. Ces points sont ceux où la surface « tourne le dos » à la lumière, c'est-à-dire où $\mathbf{S}_i \cdot \mathbf{N}(x, y) \leq 0$. Or, ce critère caractérise localement la surface. La seule différence entre points éclairés et points d'une ombre propre est qu'une égalité comme (1) devient une inégalité. Mais, alors que la présence de contraintes telles que $\mathbf{S}_i \cdot \mathbf{N}(x, y) \leq 0$ compliquerait sérieusement la résolution classique (cf. paragraphe 2.1), nous allons voir que l'apprentissage s'en accommode très bien. En revanche, aucune image de la figure 1 ne comporte d'ombre portée, car ce deuxième type d'ombre ne peut pas être caractérisé localement (pas plus d'ailleurs que les réflexions mutuelles). En outre, il semble difficile de distinguer ombre propre et ombre portée. Nous verrons dans le paragraphe 4 comment détecter les ombres portées.

3.2 Principe de l'apprentissage

Nous allons illustrer la résolution de la stéréophotométrie par apprentissage à l'aide d'intersections entre « isophotes ». Une isophote est formée par les pixels d'une image qui ont un niveau de gris donné. Les isophotes sont généralement des courbes, sauf dans le cas des ombres où elles constituent des ensembles de surface non nulle.

Si nous connaissions la réflectance du vase, l'approche classique consisterait à résoudre, en chaque point du vase, un système de $n = 5$ équations du même type que (1), mais en plus compliqué : d'une part, les produits scalaires $\mathbf{S}_i \cdot \mathbf{N}(x, y)$ devraient être remplacés par $\max\{\mathbf{S}_i \cdot \mathbf{N}(x, y), 0\}$; d'autre part, nous devrions utiliser un modèle de réflectance plus réaliste que la loi de Lambert, donc plus complexe. L'idée de base de la résolution par apprentissage consiste à ne pas utiliser de modèle de réflectance, empirique ou pas, mais à *apprendre* cette réflectance à l'aide d'images d'une surface de relief connu, de même réflectance que l'objet à reconstruire, vue sous les mêmes éclairages.

Supposons par exemple que nous souhaitions estimer la normale au point P_1 du vase représenté par une croix noire sur l'image de gauche de la figure 2. Chaque niveau de gris de P_1 lu sur une des images du bas de la figure 1 est associée à une isophote dans l'image de la sphère correspondant au même éclairage. Chaque éclairage contraint donc le point P_1 à avoir la même normale qu'un des points de la sphère situé sur l'isophote correspondante. Si l'intersection de toutes les isophotes ainsi définies est unique, alors on aura bien trouvé un moyen de lever l'ambiguïté sur la normale. Or, la superposition des 5 isophotes ainsi recueillies montre qu'il en va bien ainsi (cf. l'image du centre de la figure 2, où les couleurs utilisées correspondent aux couleurs des vecteurs d'éclairage de la figure 1). La normale en ce point de la sphère nous donne donc une estimation de la normale en P_1.

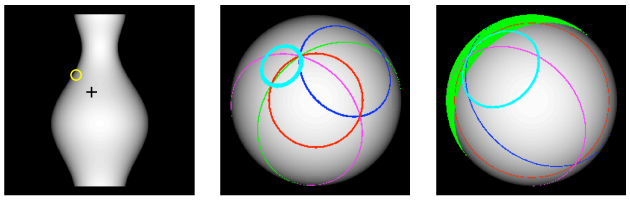

FIGURE 2 – À gauche : le point P_1 du vase repéré par une croix noire est éclairé dans les 5 images, alors que le point P_2 repéré par un cercle jaune est dans l'ombre sous l'éclairage $\mathbf{S}_2$, c'est-à-dire dans la deuxième image du bas de la figure 1. Au centre : les $n = 5$ isophotes des images de la sphère correspondant aux 5 niveaux de gris du point P_1 sont concourantes en un point unique, ce qui nous donne un moyen d'estimer la normale en P_1. À droite : ce résultat est encore vrai pour le point P_2, bien que le deuxième niveau de gris de P_2 soit nul, et que l'isophote correspondant (de couleur verte) soit un ensemble de surface non nulle.

L'estimation de la normale par intersection d'isophotes permet de comprendre pourquoi les ombres propres ne posent pas de problème particulier. Sur l'image de droite de la figure 2, on a reporté les $n = 5$ isophotes correspondant au point P_2 du vase repéré par un cercle jaune sur l'image de gauche de la figure 2. Comme ce point se situe dans l'ombre propre sous l'éclairage $\mathbf{S}_2$, l'isophote correspondant à $\mathbf{S}_2$ (de couleur verte) a une surface non nulle, mais les 5 isophotes ont encore une intersection unique.
L'intersection d'isophotes permet également de comprendre différentes situations connues pour être ambiguës (cf. figure 3). Le cas $n = 1$ illustre l'ambiguïté fondamentale du shape-from-shading. Le cas $n = 2$ montre qu'on passe d'une infinité de solutions à deux solutions, c'est-à-dire que l'ajout d'un deuxième éclairage suffit à rendre le problème « presque bien posé » [10, 11]. Enfin, le cas où l'on dispose de $n \geq 3$ éclairages coplanaires reste mal posé, ce qui s'illustre par le fait que les $n \geq 3$ isophotes n'ont pas une intersection unique (dans ce cas, sous l'hypothèse lambertienne, la pseudo-inverse de la matrice $\mathbf{S}$ du système (3) ne serait pas définie).

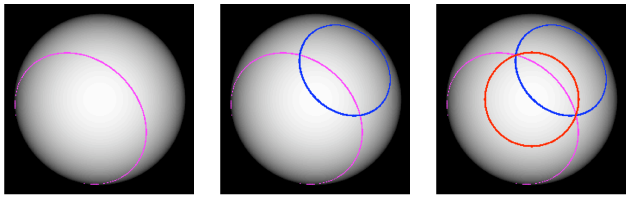

FIGURE 3 – À gauche : le cas $n = 1$ illustre l'ambiguïté du shape-from-shading. Au centre : si $n = 2$, il y a deux intersections donc deux normales possibles. À droite : avec $n = 3$ éclairages coplanaires, la même ambiguïté persiste.

3.3 Méthode de résolution de Hertzmann

L'idée de résoudre la stéréophotométrie par apprentissage est due à Woodham, qui la suggère dès ses premiers travaux [14]. Néanmoins, sa mise en œuvre est récente [8, 9]. En fait, l'estimation de la normale par intersection d'isophotes, telle nous l'avons présentée dans le paragraphe 3.2, est avant tout *pédagogique*. Comment faire, en pratique, pour trouver de telles intersections ?
La solution très intuitive proposée par Hertzmann et al. dans [8] consiste à rechercher, pour chaque point (x, y) de la surface de test (ici, le vase) caractérisé par les n niveaux de gris formant un « vecteur de test » $G_t \in \mathbb{R}^n$, le point de la surface d'apprentissage (ici, la sphère) dont le « vecteur d'apprentissage » $G_a \in \mathbb{R}^n$ est le plus proche (au sens de la distance euclidienne dans $\mathbb{R}^n$), et à attribuer la normale correspondant à G_a au point (x, y). Comme on dispose d'un nombre fini de points sur la surface d'apprentissage, puisque ces points sont en fait des pixels, il est envisageable d'effectuer une recherche exhaustive du point le plus proche. Hertzmann et al. utilisent toutefois l'algorithme rapide ANN (*Approximate Nearest Neighbor*). Dans le cas de $n = 8$ images d'apprentissage de taille 332×328 et de 8 images de test de taille 1176×398 (cf. figure 4), le temps de calcul sur un 2 GHz Intel Xeon passe ainsi d'un jour, pour une recherche exhaustive, à environ 5 minutes.
Dans [12], l'utilisation de la distance euclidienne dans un espace de dimension n potentiellement élevée est remise en question. Une distance géodésique sur la nappe passant par les vecteurs G_a lui est préférée. Néanmoins, si cela est peut-être vrai pour un nombre d'images très élevé, comme dans [12] et [9] où toutes les images d'une séquence vidéo sont utilisées, la méthode de Hertzmann semble valide, au vu des tests, pour un nombre d'images relativement faible.

3.4 Analyse en composantes principales

Notre première contribution consiste à réduire la dimension n de l'espace des vecteurs d'apprentissage et de test, pour accélérer la recherche du plus proche voisin qui constitue l'étape cruciale de la méthode de Hertzmann. Nous effectuons donc l'analyse en composantes principales (ACP) des vecteurs d'apprentissage G_a et de test G_t, afin de ne garder que les $p < n$ premières composantes, qui forment des « vecteurs réduits » $\bar{G}_a$ et $\bar{G}_t$. Il est bien difficile de prédire si cette réduction de dimension est appropriée : on sait que les données d'apprentissage doivent se répartir dans $\mathbb{R}^n$ sur une nappe, puisque la normale (normée) possède 2 degrés de liberté, mais la forme de cette nappe dépend beaucoup du matériau. C'est la validation par les tests qui nous dira si la démarche était justifiée.
Un premier test a été mené sur les images de synthèse de la figure 1. Le tableau 1 donne le temps de calcul (CPU, en secondes) et la racine carrée de l'écart quadratique moyen (RMSE, en pixels) des reconstructions 3D obtenues par la méthode de Hertzmann sans ACP, et après réduction de la dimension aux $p = 4$, 3, 2 ou 1 premières composantes. Le temps CPU ne prend en compte ni l'apprentissage sur les $n = 5$ images de la sphère, ni l'intégration des normales qui est effectuée à l'aide d'une méthode décrite dans [4].

Méthode	CPU (secondes)	RMSE (pixels)
Hertzmann	$0,66$	$1,58$
ACP ($p = 4$)	$0,57$	$1,58$
ACP ($p = 3$)	$0,47$	$1,59$
ACP ($p = 2$)	$0,37$	$4,98$
ACP ($p = 1$)	$0,28$	$6,92$

TABLE 1 – Reconstructions 3D du vase par stéréophotométrie à partir des images de la figure 1 : temps CPU et RMSE sans ACP (méthode de Hertzmann) ou avec ACP (en utilisant les $p = 4$, 3, 2 ou 1 premières composantes).

Ces résultats montrent que la réduction de la dimension de l'espace des configurations ne dégrade pas la précision du résultat tant que $p \geq 3$. Pour $p < 3$, la situation est similaire à celles de la figure 3 : l'estimation de la normale devient mal posée, ce qui se traduit par des valeurs très élevées de la RMSE. En revanche, la réduction de dimension permet bien de réduire significativement le temps CPU.

Un deuxième test a été mené sur les images réelles de la figure 4. Une sphère et une bouteille recouvertes de la même peinture sont photographiées sous $n = 8$ éclairages différents. La peinture utilisée possède une composante spéculaire non négligeable. Les images comportent des ombres propres, mais quasiment aucune ombre portée. Le tableau 2 confirme les enseignements que nous avons tirés du premier test. Les valeurs de la RMSE affichées dans ce tableau ne reflètent pas l'écart au relief réel de la bouteille, dont nous ne disposons malheureusement pas, mais l'écart au résultat obtenu par la méthode de Hertzmann, censée être la plus précise.
Les surfaces obtenues en intégrant les normales estimées sont montrées en perspective sur la figure 5. Ces résultats confirment la puissance de la résolution par apprentissage, alors que la surface est loin d'être lambertienne. Les surfaces montrées sur la figure 5 sont qualitativement satisfaisantes, à l'exception du cas $p = 2$ (en bas à droite) dont on savait par avance qu'il ne pourrait fournir de bons résultats. En particulier, les détails les plus fins sont correctement retrouvés. Comme nous l'avons déjà dit en introduction, c'est cette aptitude de la stéréophotométrie à retrouver les détails les plus fins qui explique sa vogue actuelle.

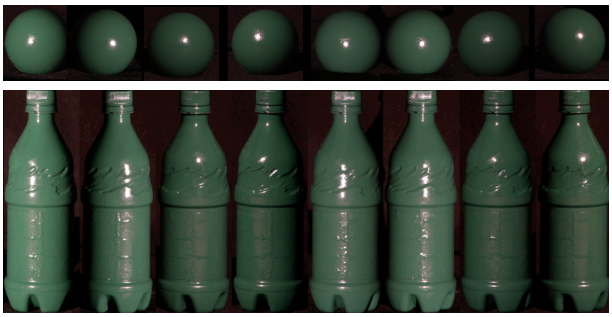

FIGURE 4 – Images réelles d'une sphère (taille 332×328) et d'une bouteille (taille 1176×398) vues sous $n = 8$ éclairages différents. Ces images sont accessibles à l'adresse : grail.cs.washington.edu/projects/sam/

Méthode	CPU (secondes)	RMSE (pixels)
Hertzmann	608	0.0
ACP ($p = 7$)	518	$0,116$
ACP ($p = 6$)	470	$0,150$
ACP ($p = 5$)	392	$0,164$
ACP ($p = 4$)	312	$0,214$
ACP ($p = 3$)	238	$0,265$
ACP ($p = 2$)	191	$0,400$
ACP ($p = 1$)	170	$0,911$

TABLE 2 – Reconstructions 3D de la bouteille par stéréophotométrie à partir des images de la figure 4 : temps CPU et RMSE sans ACP (méthode de Hertzmann) ou avec ACP (en utilisant les $p = 7, \ldots, 1$ premières composantes).

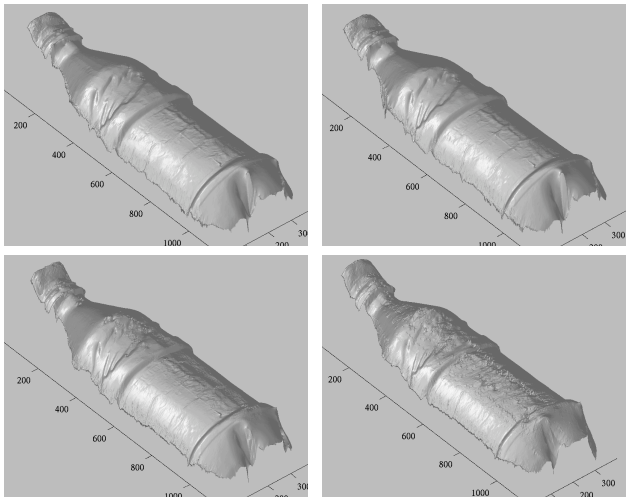

FIGURE 5 – Reconstructions 3D de la bouteille par stéréophotométrie à partir des images de la figure 4 : vues en perspective des surfaces reconstruites sans ACP (méthode de Hertzmann) ou avec ACP, à l'aide des $p = 7$, 3 ou 2 premières composantes (de gauche à droite, de haut en bas).

Nous disposons donc d'une méthode de résolution de la stéréophotométrie quasiment aussi précise que la méthode de Hertzmann décrite dans [8], mais plus rapide : d'après le tableau 2, les deux résultats montrés sur la colonne de gauche de la figure 5 requièrent 608 secondes pour la méthode de Hertzmann (en haut) et 238 secondes en utilisant les $p = 3$ premières composantes de l'ACP (en bas).
Deux obstacles limitent encore le potentiel de cette méthode. Primo, les ombres portées ne sont pas prises en compte. Plus généralement, l'estimation de la normale se dégrade dès lors qu'un niveau de gris est erroné. Secundo, et cet obstacle est le plus limitant, on ne peut pas envisager de systématiquement peindre les objets à reconstruire !

4 Ombres portées et texture

Reprenons l'exemple de la figure 1 pour illustrer l'effet des ombres portées. Supposons que le point P_1 du vase signalé par une croix noire (cf. l'image de gauche de la figure 6) se trouve dans une ombre portée sous l'éclairage $\mathbf{S}_4$, c'est-à-dire dans la quatrième image du bas de la figure 1. L'isophote (de couleur cyan) correspondant à ce niveau de gris devient erronée, et l'intersection des isophotes devient vide (cf. l'image du centre de la figure 6).
D'autres difficultés peuvent survenir si l'on ne peut plus peindre la surface de l'objet à reconstruire. Nous n'abordons pas ce problème difficile dans le cas général. Nous supposons pour simplifier que les deux objets ont la même réflectance, mais que l'objet d'apprentissage possède un albédo uniforme égal à 1, alors que l'objet de test est texturé. Si le point P_1 a un albédo égal à $0,8$, ses $n = 5$ niveaux de gris sont multipliés par un facteur $0,8$ et les cinq isophotes sont simultanément erronées (cf. l'image de droite de la figure 6).

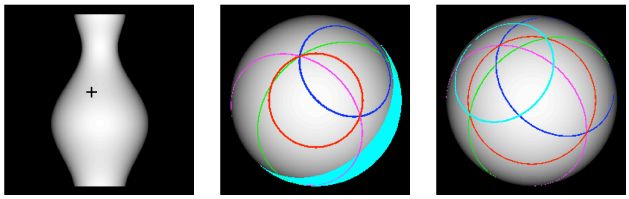

FIGURE 6 – À gauche : localisation d'un point P_1 sur le vase. Au centre : si le point P_1 se trouve dans une ombre portée sous l'éclairage $\mathbf{S}_4$, l'isophote correspondante (en cyan) est erronée et l'intersection des isophotes est vide. À droite : si l'albédo de P_1 est égal à $0,8$, toutes les isophotes sont erronées, et seules les isophotes des éclairages coplanaires se croisent encore.

Ces deux situations présentent plusieurs points communs. D'une part, bien que la perte de clarté ait des origines très différentes dans les deux cas, un certain nombre de niveaux de gris sont inférieurs à leurs valeurs théoriques. D'autre part, ces niveaux de gris erronés rendent erronées les isophotes correspondantes, en conséquence de quoi les isophotes ne sont généralement plus concourantes.

Heureusement, on peut quand même distinguer ces deux cas de figure et apporter à chacun le remède approprié. En présence d'ombres portées, on sait que les niveaux de gris erronés ont des valeurs faibles (mais pas forcément nulles, à cause de l'effet de pénombre). Il suffit donc de retirer les isophotes incriminées pour que les isophotes restantes soient à nouveau concourantes. En présence d'une texture, tous les niveaux de gris sont erronés, donc toutes les isophotes aussi, mais en rajoutant l'albédo ρ comme inconnue au problème, on peut y remédier : il suffit en effet de chercher la valeur de ρ telle que les isophotes correspondant aux n niveaux de gris divisés par ρ soient concourantes. Il est bien sûr impossible de procéder ainsi en pratique. Nous proposons de tirer à nouveau parti de l'ACP, en ne nous contentant plus de rechercher, dans l'espace de dimension réduite $\mathbb{R}^p$, le vecteur d'apprentissage réduit $\bar{G}_a$ le plus proche d'un vecteur de test réduit $\bar{G}_t$. Nous utilisons dorénavant la distance minimale entre ces deux vecteurs réduits comme mesure de l'écart au modèle de réflectance appris.

En guise de premier test, nous utilisons les images de synthèse d'une sphère et d'une théière vues sous $n = 8$ éclairages différents (cf. figure 7). En utilisant les $p = 5$ premières composantes de l'ACP, on peut effectivement calculer, pour chaque vecteur de test réduit $\bar{G}_t$ de la théière, la distance minimale à un vecteur d'apprentissage réduit $\bar{G}_a$ (à gauche sur la figure 8). Cette distance mesure la fréquence d'apparition d'une ombre portée. Si l'on applique la méthode d'estimation des normales par recherche du plus proche voisin dans $\mathbb{R}^5$, sans tenir compte de cette mesure, on obtient un champ de normales erroné dans les zones d'ombres portées, comme par exemple au voisinage de l'anse de la théière (à gauche sur la figure 9). En retirant les niveaux de gris inférieurs à un seuil, et en appliquant la méthode de Hertzmann sur les niveaux de gris restants, nous pouvons alors améliorer la cohérence du champ de normales près de la discontinuité (à droite sur la figure 9).

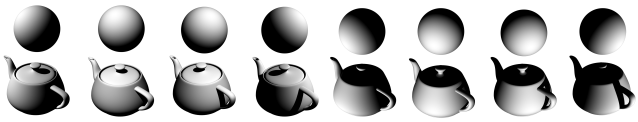

FIGURE 7 – Images de synthèse d'une sphère et d'une théière vues sous $n = 8$ éclairages différents. Les images de la théière comportent des ombres propres et des ombres portées. Images gracieusement fournies par FittingBox.

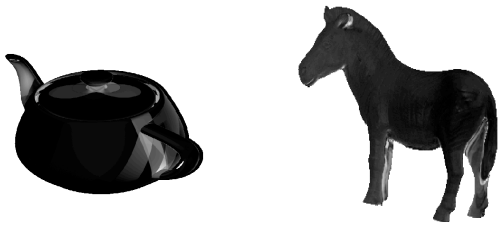

FIGURE 8 – À gauche : écarts au modèle appris à partir des images de la figure 7, en utilisant les $p = 5$ premières composantes de l'ACP. À droite : écarts au modèle appris à partir des images de la figure 10, en utilisant les $p = 5$ premières composantes de l'ACP. Le niveau de gris indique la fréquence d'apparition d'une ombre portée.

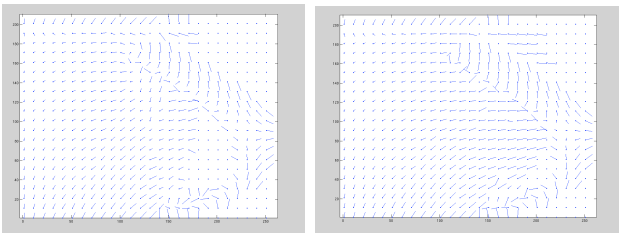

FIGURE 9 – Champs de normales au voisinage de l'anse de la théière estimés par stéréophotométrie à partir des images de la figure 7, en utilisant les $p = 5$ premières composantes de l'ACP : sans prise en compte (à gauche) et avec prise en compte (à droite) des écarts au modèle appris (cf. figure 8).

Effectuons un deuxième test sur les images réelles de la figure 10. Une sphère et un cheval miniature, peints avec la même peinture, sont vus sous $n = 12$ éclairages différents. En utilisant les $p = 5$ premières composantes de l'ACP, on obtient la carte des écarts au modèle appris (à droite sur la figure 8). Les reconstructions 3D obtenues, avec ou sans prise en compte de ces écarts, sont montrées sur la figure 11. L'amélioration du résultat dans une zone sujette aux ombres portées se voit très nettement en faisant un zoom sur les pattes avant du cheval (cf. figure 12).

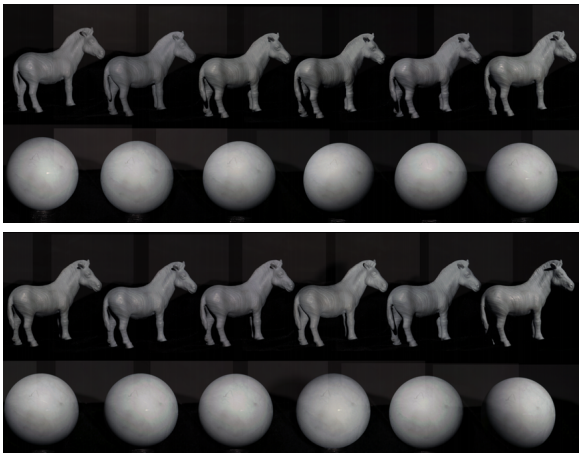

FIGURE 10 – Images réelles d'une sphère et d'un cheval ($n = 12$). Les images du cheval comportent des ombres propres et des ombres portées (près des pattes). Ces images sont accessibles à l'adresse suivante : www.cs.washington.edu/education/courses/csep576/05wi/projects/project3/project3.htm

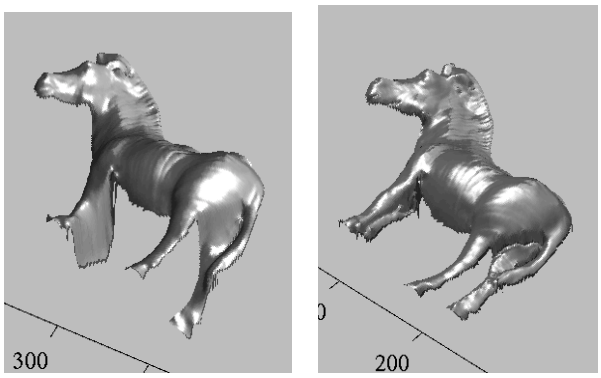

FIGURE 11 – Reconstructions 3D du cheval par stéréophotométrie à partir des images de la figure 10, en utilisant les $p = 5$ premières composantes de l'ACP : sans prise en compte (à gauche) et avec prise en compte (à droite) des écarts au modèle appris (cf. figure 8).

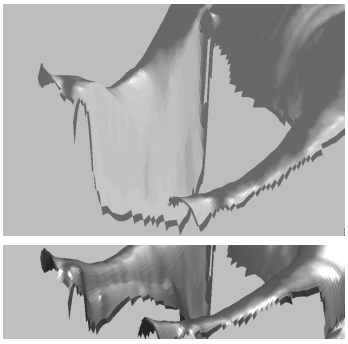

FIGURE 12 – Zoom sur les pattes avant du cheval des reconstructions 3D de la figure 11.

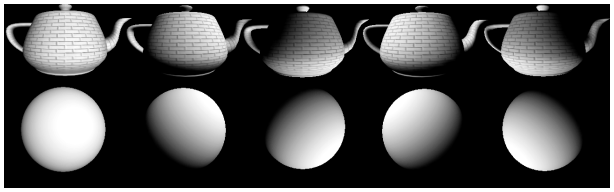

FIGURE 13 – Images de synthèse d'une sphère et d'une théière texturée ($n = 5$). Les deux surfaces sont supposées lambertiennes, mais la théière est texturée.

FIGURE 14 – Albédos estimés à partir des images de la figure 13, en utilisant $p = 3$ composantes principales.

Effectuons un dernier test sur les images de synthèse d'une sphère et d'une théière vues sous $n = 5$ éclairages différents (cf. figure 13). La théière est revêtue d'une texture de type « mur de briques ». Elle n'a donc pas un albédo uniforme. Bien que les théières soient rarement en briques, cet exemple donne une idée d'une version « non intrusive » de notre méthode, dans laquelle la surface à reconstruire est photographiée telle quelle.
Pour chaque point de la théière de vecteur de test réduit $\bar{G}_t$, on cherche l'albédo ρ tel que $\bar{G}_t$ divisé par ρ se trouve « au plus près » d'un vecteur d'apprentissage réduit $\bar{G}_a$. La figure 14 montre la carte des albédos ainsi estimés. La figure 15 montre les reconstructions 3D obtenues par la méthode de Hertzmann et par notre méthode. Comme la méthode de Hertzmann ne tient pas compte de la texture, elle provoque l'apparition de nervures le long des contours des briques, là où l'albédo est le plus faible.

5 Conclusion et perspectives

Dans cet article, nous reprenons une idée de Woodham récemment mise en œuvre par Hertzmann et al. [8], qui consiste à résoudre la stéréophotométrie par apprentissage. Cela permet de se libérer d'un certain nombre de contraintes de la résolution classique du problème. En particulier, il n'est plus nécessaire de connaître le modèle de réflectance de la surface à reconstruire, ni même d'estimer les paramètres d'un modèle empirique, puisque le modèle est appris sur les images d'un objet de relief connu. Un autre avantage est que les éclairages peuvent être quelconques (mais identiques pour les deux objets). Enfin, les ombres propres ne posent pas de problème particulier, car elles donnent une information sur la normale recherchée.

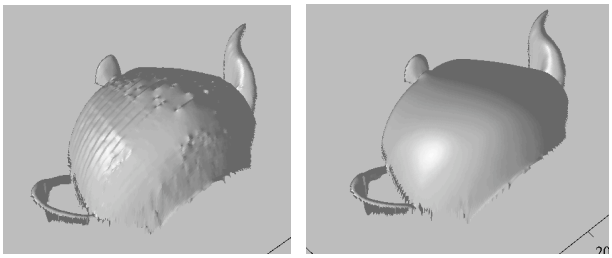

FIGURE 15 – Reconstructions 3D de la théière par stéréophotométrie à partir des images de la figure 13, en utilisant la méthode de Hertzmann (à gauche) et notre méthode avec estimation de l'albédo (à droite), en utilisant les $p = 3$ premières composantes de l'ACP.

Nous avons montré que l'analyse en composantes principales des vecteurs d'apprentissage et de test permet d'accélérer la méthode de Hertzmann sans perdre en précision, et avons proposé deux généralisations de cette méthode. D'une part, nous avons montré que l'utilisation de la distance au plus proche vecteur d'apprentissage réduit permet de détecter les ombres portées. D'autre part, l'introduction de l'albédo comme inconnue nous a permis de détecter les zones d'albédo faible et de ne pas les interpréter, à tort, comme des zones fortement inclinées par rapport à la lumière. Ces deux modifications nous ont permis d'améliorer significativement la qualité des résultats.
Ce travail ouvre autant de perspectives qu'il ne résout de problèmes. Primo, nous n'avons pas traité le cas d'une surface texturée avec ombres portées. Comme ces deux types d'écarts au modèle appris sont détectés par une mesure dans l'espace de dimension réduite $\mathbb{R}^p$, il ne sera pas forcément facile de les distinguer. Secundo, il arrive souvent que les ombres portées se concentrent près des discontinuités. Dans l'exemple de la figure 7, certains points situés sous l'anse de la théière ne sont éclairés que dans deux images, voire dans une seule image (cf. figure 8). Il nous faudrait proposer une estimation de la normale dans de tels cas de figure. Enfin, nous nous sommes restreints dans cet article à des images en niveaux de gris, contrairement au travail de Hertzmann et al. [8] qui tire parti de la couleur.

Références

[1] S. BARSKY et M. PETROU. « The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows ». *IEEE Trans. Patt. Anal. Mach. Intell.*, 25(10) :1239–1252, octobre 2003. 2

[2] R. BASRI, D. W. JACOBS et I. KEMELMACHER. « Photometric Stereo with General, Unknown Lighting ». *Int. J. Comp. Vis.*, 72(3) :239–257, mai 2007. 2

[3] M. K. CHANDRAKER, S. AGARWAL et D. J. KRIEGMAN. « ShadowCuts : Photometric Stereo with Shadows ». In *Proc. IEEE Conf. Comp. Vis. and Patt. Recog.*, Minneapolis, États-Unis, juin 2007. 2

[4] J.-D. DUROU, J.-F. AUJOL et F. COURTEILLE. « Integration of a Normal Field in the Presence of Discontinuities ». In *Proc. 7th Int. Worksh. En. Min. Meth. Comp. Vis. and Patt. Recog.*, volume 5681 de *LNCS*, pages 261–273, Bonn, Allemagne, août 2009. 2, 4

[5] J.-D. DUROU, M. FALCONE et M. SAGONA. « Numerical Methods for Shape-from-shading : A New Survey with Benchmarks ». *Comp. Vis. and Image Underst.*, 109(1) :22–43, janvier 2008. 1

[6] C. HERNÁNDEZ, G. VOGIATZIS, G. J. BROSTOW, B. STENGER et R. CIPOLLA. « Non-rigid Photometric Stereo with Colored Lights ». In *Proc. 11th IEEE Int. Conf. Comp. Vis.*, Rio de Janeiro, Brésil, octobre 2007. 1, 2

[7] C. HERNÁNDEZ, G. VOGIATZIS et R. CIPOLLA. « Overcoming shadows in 3-source photometric stereo ». *IEEE Trans. Patt. Anal. Mach. Intell.*, 33(2) :419–426, février 2011. 2

[8] A. HERTZMANN et S. M. SEITZ. « Example-Based Photometric Stereo : Shape Reconstruction with General, Varying BRDFs ». *IEEE Trans. Patt. Anal. Mach. Intell.*, 27(8) :1254–1264, août 2005. 1, 2, 4, 5, 7, 8

[9] S. J. KOPPAL et S. G. NARASIMHAN. « Appearance Derivatives for Isonormal Clustering of Scenes ». *IEEE Trans. Patt. Anal. Mach. Intell.*, 31(8) :1375–1385, août 2009. 2, 4

[10] R. KOZERA. « Existence and Uniqueness in Photometric Stereo ». *Appl. Math. Comput.*, 44(1) :1–103, 1991. 2, 4

[11] R. MECCA et J.-D. DUROU. « Unambiguous Photometric Stereo Using Two Images ». In *Proc. 16th Int. Conf. Im. Anal. and Proc.*, Ravenne, Italie, septembre 2011. 2, 4

[12] I. SATO, T. OKABE, Q. YU et Y. SATO. « Shape Reconstruction Based on Similarity in Radiance Changes under Varying Illumination ». In *Proc. 11th IEEE Int. Conf. Comp. Vis.*, Rio de Janeiro, Brésil, octobre 2007. 4

[13] J. SUN, M. SMITH, L. SMITH, S. MIDHA et J. BAMBER. « Object surface recovery using a multi-light photometric stereo technique for non-Lambertian surfaces subject to shadows and specularities ». *Im. and Vis. Comp.*, 25(7) :1050–1057, juillet 2007. 2

[14] R. J. WOODHAM. « Reflectance Map Techniques for Analyzing Surface Defects in Metal Castings ». Technical Report AITR-457, Artificial Intelligence Laboratory, MIT, Cambridge, Massachusetts, États-Unis, juin 1978. 1, 2, 4

[15] R. J. WOODHAM. « Photometric Method for Determining Surface Orientation from Multiple Images ». *Opt. Engin.*, 19(1) :139–144, 1980. 1, 2