

Apprentissage par cartes de réflectance pour la stéréo-photométrie non lambertienne pixel par pixel

Clément JOUBERT¹, Nicolas PROUTEAU^{1,2}, Benjamin BRINGIER¹, Majdi KHOUDEIR¹

¹Laboratoire XLIM UMR CNRS 7252, Université de Poitiers

²Laboratoire PPRIME UPR CNRS 3346, Université de Poitiers

prenom.nom@univ-poitiers.fr

Résumé – Résoudre le problème de la stéréo-photométrie calibrée avec un nombre limité de lampes présente un grand intérêt pour des applications réelles. Aujourd’hui, les méthodes utilisant des réseaux de neurones sont devenues la norme. Dans ce cadre, nous proposons une nouvelle représentation basée sur les cartes de réflectance adaptée aux systèmes avec un nombre limité de lampes.

Abstract – Solving calibrated photometric stereo under a sparse set of lights is of great interest for real world applications. Since neural networks are showing advantages in dealing with specular and shadowed areas, this paper proposes a new representation based on reflectance maps to overcome the lack of information in observation maps for sparse set of lights.

1 Introduction

La stéréo-photométrie calibrée est une méthode de reconstruction 3D utilisant plusieurs images prises du même point de vue et différentes positions d’éclairage. Historiquement, en chaque point de la surface, la normale est calculée en utilisant les radiances venant de 3 positions d’éclairage. Dans [13], la méthode suppose le cas limite de la BRDF (Fonction de Distribution de Réflectance Bidirectionnelle) diffuse. Cette hypothèse n’est pas réaliste et la difficulté est de traiter le cas de BRDF générale avec un nombre limité de positions de lampe. Pour surmonter ces limites, l’approche classique est d’exclure les pixels acquis dans le lobe spéculaire ou dans l’ombre [1]. Celle-ci n’est pas performante lorsque les ombres et les lobes spéculaires se chevauchent. Une autre catégorie d’approches utilise des modèles de réflectance plus sophistiqués comme le modèle de Torrance et Sparrow [3]. Ces approches permettent d’obtenir de bons résultats mais nécessitent souvent un nombre de lampes et des temps de traitement importants.

Les méthodes récentes utilisent l’apprentissage profond pour résoudre la stéréo-photométrie. Elles présentent principalement des variations au niveau de la représentation de la donnée d’entrée mais aussi de l’architecture du réseau utilisé. Deux catégories d’approches existent : locale (pixel à pixel) et globale (niveau image). Parmi les approches globales, Chen et al. [2] proposent un CNN qui utilise l’ensemble des images normalisées et leurs positions d’éclairage. Cette méthode permet un meilleur traitement des ombres mais souffre d’un effet de lissage. Parmi les approches locales, Santo et al. [10] utilisent un réseau constitué de couches denses avec en entrée le vecteur normalisé de pixels pour une position donnée dans l’ensemble d’images. Ikehata [6] propose une représentation intermédiaire

de la donnée appelée carte d’observation. Diverses approches ont été développées autour de cette carte d’observation. Zheng et al. [14] appliquent cette méthode en ajoutant une première étape pour densifier la carte d’observation. Alors que les résultats des méthodes globales sont dégradés sur les surfaces rugueuses et lorsque la surface est composée d’une multitude de BRDF [7], la carte d’observation utilisée pour les méthodes locales n’est pas une représentation adaptée pour des systèmes avec un nombre limité de lampes.

Ainsi, nous proposons une méthode locale basée sur une nouvelle représentation de la donnée en stéréo-photométrie par cartes de réflectance. Pour des acquisitions réelles et des BRDF complexes, ces cartes décrivent correctement les phénomènes spécifiques des zones non-linéaires comme l’ombre portée et le spéculaire. Cette approche permet de représenter l’information de façon dense même si le nombre d’éclairages est faible. Un motif discriminant est produit pour chaque normale de la surface dans les zones d’ombre portée ou spéculaires. Nous montrons que notre descripteur permet une représentation prédictible de ces phénomènes sans discontinuités selon les variations des paramètres physiques.

2 Représentation de la radiance

La radiance est réfléchi par un élément de surface éclairé par son environnement. Deux représentations sont généralement utilisées pour présenter ces valeurs sous forme de carte dans le cadre de la stéréo-photométrie.

Si nous considérons un capteur parfait, l’intensité d’un pixel à la position (x, y) selon une seule direction d’éclairage \mathbf{i} est :

$$I(x, y, \mathbf{o}, \mathbf{n}) = L_i(x, y, \mathbf{i}) f_r(x, y, \mathbf{i}, \mathbf{o}, \mathbf{n}) \max(\mathbf{i} \cdot \mathbf{n}, 0) \quad (1)$$

avec I l’intensité acquise par le capteur, L_i l’intensité de la

source lumineuse, \mathbf{n} la normale à la surface, \mathbf{i} le vecteur de direction d'éclairage, \mathbf{o} la position du capteur et f_r la BRDF. Le produit scalaire entre le vecteur de direction d'éclairage \mathbf{i} et la normale à la surface \mathbf{n} est :

$$\mathbf{in} = \frac{-p \cos(\tau) \sin(\sigma) - q \sin(\tau) \sin(\sigma) + \cos(\sigma)}{\sqrt{p^2 + q^2 + 1}} \quad (2)$$

avec τ et σ l'azimut et le zénith de la lumière et $p = -\frac{\partial S(x,y)}{\partial x}$ et $q = -\frac{\partial S(x,y)}{\partial y}$ les dérivées partielles de la surface S .

La méthode la plus simple pour représenter les intensités $I_k(x, y)$ de k positions d'éclairage est la carte d'observation (Fig. 1). Elle utilise les coordonnées polaires τ_k et σ_k pour dessiner chaque intensité $I_k(x, y)$. Cette méthode est utilisée par Ikehata [6] comme donnée d'entrée pour une approche locale. Dans ce cadre, la fonction f_r est représentée pour une position \mathbf{o} et tous les \mathbf{i} possibles. Cette représentation est problématique avec la stéréo-photométrie peu dense [14].

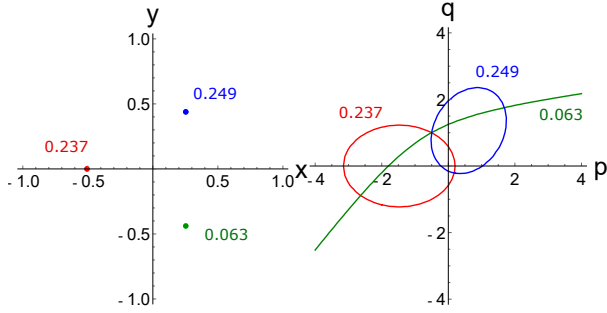


FIGURE 1 – Représentation des valeurs de pixel I_k : (a) une carte d'observation et (b) une carte de réflectance.

La seconde représentation possible est la carte de réflectance. Une valeur I_k est représentée par un iso-contour ou iso-réflectance pour tous les gradients compatibles p et q et un modèle de BRDF. Dans [13], Woodham suggère d'utiliser une BRDF lambertienne, soit un terme constant ρ lié à la couleur de l'objet. La Fig. 1 montre un exemple de ces deux représentations pour 3 positions d'éclairage. En comparaison avec une carte d'observation, la carte de réflectance est plus dense.

3 Méthode proposée

La Fig. 2 présente la chaîne de traitement pour estimer le vecteur normal de chaque élément de surface. Ce traitement pixel par pixel consiste en : 1) générer une carte de réflectance en fonction des intensités de pixel des images HDR et 2) l'analyser par un réseau de neurone convolutif. Ce dernier est similaire à celui mentionné par Ikehata [6].

3.1 Carte de réflectance à ratio

Un problème connu avec les cartes de réflectance est que les courbes d'iso-réflectances sont fortement dépendantes de la BRDF f_r . Dans le cas particulier du modèle lambertien, il y a un problème d'échelle. Dans le cas plus général, les courbes d'iso-réflectances dépendent de \mathbf{in} et f_r . La solution la plus

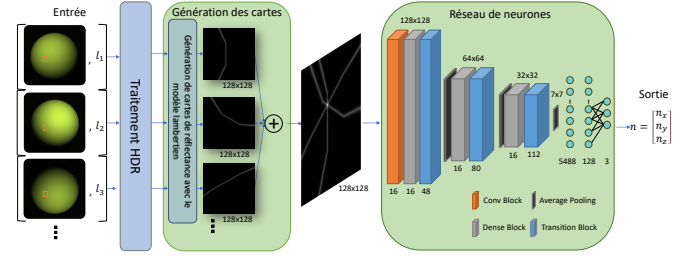


FIGURE 2 – Diagramme de notre méthode.

utilisée dans le cas lambertien est l'utilisation du ratio [9]. Si une direction d'éclairage \mathbf{i}_k est exprimée dans l'espace des gradients :

$$\mathbf{i}_k = \frac{1}{\sqrt{p_{sk}^2 + q_{sk}^2 + 1}} [p_{sk}, q_{sk}, 1] \quad (3)$$

L'iso-réflectance d'un ratio de deux valeurs I_k est représentée par une droite mais est très dépendante des I_k pour une faible variation de p et q . Une approche plus robuste est d'utiliser la moyenne des intensités. Pour chaque position d'éclairage une droite est définie par :

$$Ap + Bq + C = 0 \quad (4)$$

avec :

$$\begin{aligned} A &= I_1 \sum_k \frac{f_r(x, y, \mathbf{i}_k, \mathbf{o}, \mathbf{n}) p_{sk}}{\sqrt{p_{sk}^2 + q_{sk}^2 + 1}} - \frac{f_r(x, y, \mathbf{i}_1, \mathbf{o}, \mathbf{n}) p_{s1}}{\sqrt{p_{s1}^2 + q_{s1}^2 + 1}} \sum_k I_k \\ B &= I_1 \sum_k \frac{f_r(x, y, \mathbf{i}_k, \mathbf{o}, \mathbf{n}) q_{sk}}{\sqrt{p_{sk}^2 + q_{sk}^2 + 1}} - \frac{f_r(x, y, \mathbf{i}_1, \mathbf{o}, \mathbf{n}) q_{s1}}{\sqrt{p_{s1}^2 + q_{s1}^2 + 1}} \sum_k I_k \\ C &= I_1 \sum_k \frac{f_r(x, y, \mathbf{i}_k, \mathbf{o}, \mathbf{n})}{\sqrt{p_{sk}^2 + q_{sk}^2 + 1}} - \frac{f_r(x, y, \mathbf{i}_1, \mathbf{o}, \mathbf{n})}{\sqrt{p_{s1}^2 + q_{s1}^2 + 1}} \sum_k I_k \end{aligned} \quad (5)$$

Tracer la carte de réflectance n'est pas trivial à cause de la BRDF f_r dans les coefficients A , B et C .

Dans notre approche, les cartes de réflectances sont converties en petites images qui sont utilisées en entrée d'un CNN. Il est donc primordial d'utiliser un algorithme rapide pour discrétiser les lignes de la carte de réflectance. Cependant, des discontinuités apparaissent et correspondent à l'apparition d'ombres. De plus, l'espace des gradients n'est pas linéaire par rapport à l'orientation de la surface (angle σ). Le contrôle de la précision des faibles ou grandes valeurs de σ est donc réalisé grâce à une échelle logarithmique sur les axes p et q :

$$q_{log} = q_{max} \times \frac{e^{\frac{|q|}{R}} - e^0}{e^{\frac{q_{max}}{R}} - e^0} \times \text{sgn}(q) \quad (6)$$

où q_{max} est la limite du graphique et R une constante pour paramétrer la précision sur les faibles valeurs de σ . L'image générée étant petite, une approche par balayage est utilisée pour associer à chaque pixel une mesure de différence entre le ratio théorique et le ratio observé :

$$\varepsilon_j(p, q) = \frac{Lo_j(p, q)}{\sum_{i=1}^k Lo_i(p, q)} - \frac{k I_j}{\sum_{i=1}^k I_i} \quad (7)$$

Le premier membre est précalculé pour chaque éclairage sous la forme de k cartes. L'intensité du pixel est égale à :

$$X_j(p, q) = \frac{\left[\left(\frac{1}{a} - \min\left(\frac{1}{a}, |\varepsilon_j(p, q)|\right)\right) \times a\right]^b}{k} \quad (8)$$

où a contrôle l'épaisseur et b le dégradé de la ligne. L'intensité du pixel $X(p, q)$ de l'image d'entrée du CNN pour k éclairages est :

$$X(p, q) = \sum_{i=1}^k X_i(p, q) \quad (9)$$

La Fig. 3 montre un exemple de carte de réflectance d'entrée du CNN dans le cas d'une échelle linéaire et d'une échelle logarithmique pour différentes valeurs de a et b .

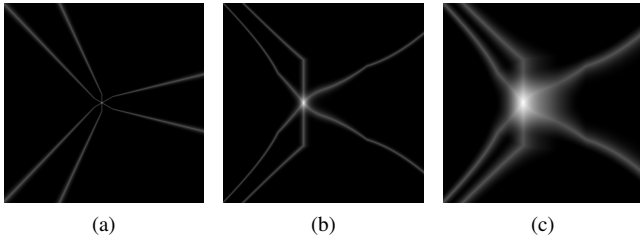


FIGURE 3 – Trois représentations de cartes de réflectance avec trois lampes pour une échelle (a) linéaire, (b) et (c) logarithmique et les paramètres (a) et (b) $a = 5, b = 4$ et (c) $a = 2, b = 2$.

3.2 Réseau de neurones et base d'entraînement

Nous utilisons un CNN basé sur l'architecture du DenseNet [5], utilisé dans Ikehata [6], pour l'analyse des cartes de réflectance (Fig 2). Pour l'apprentissage, nous générons une base de données composée de 1 000 000 de cartes de réflectance à partir de mesures de BRDF (MERL [8]). Chaque carte a une résolution de 128 pixels et une limite de gradient à 20. En utilisant l'échantillonnage d'importance [4], 40% des cartes possèdent au moins une valeur spéculaire. De la même manière, 40% des cartes contiennent au moins un pixel correspondant à une ombre [7].

4 Expériences

Dans cette section, une comparaison est réalisée avec les méthodes de l'état de l'art : résolution par les moindres carrés [13], CNN local (CNN-PS [6]) et CNN global (PS-FCN [2]). La métrique pour l'évaluation est l'erreur angulaire absolue entre la normale de la vérité terrain et celle estimée.

4.1 Base de données DiLiGenT

DiLiGenT est une base de données de test [11] composée de 10 objets illuminés depuis 96 directions différentes et associés à une carte de normales acquises par projection de motifs pour vérité terrain. Nous avons sélectionné 6 directions d'éclairage sur les 96 disponibles pour répliquer un système de stéréo-photométrie équipé de 6 lampes placées de manière

homogène. Cette configuration permet de comparer les performances entre l'état de l'art et notre méthode sous des conditions proches de celles de notre système.

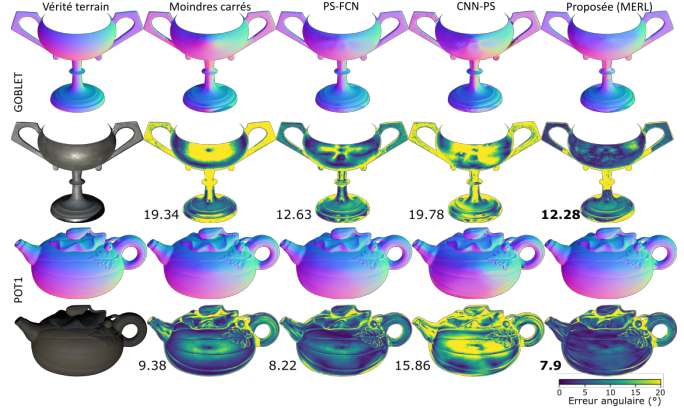


FIGURE 4 – Comparaison avec les méthodes de l'état de l'art pour 2 objets DiLiGenT. Chaque estimation est associée à son erreur angulaire moyenne.

Dans la Fig. 4, nous obtenons de meilleurs résultats par rapport aux méthodes de l'état de l'art. Cette conclusion est similaire pour toute la base DiLiGenT. L'estimation des normales en présence de pixels spéculaires est particulièrement précise comparée aux autres méthodes. De plus, la méthode locale CNN-PS [6] donne des résultats moins précis que ceux obtenus par la méthode des moindres aux carrés [13]. Notre méthode permet d'obtenir des résultats similaires ou meilleurs à la méthode globale PS-FCN [2]. Pour cette base, la vérité terrain est obtenue par projection de franges et présente moins de détails que par stéréo-photométrie [11]. Ainsi, elle ne permet d'évaluer que la robustesse globale d'une méthode et ne permet pas de conclure sur la supériorité d'une méthode par rapport à une autre lorsque les variations entre les méthodes sont proches.

4.2 Notre système d'acquisitions

Pour préciser les conclusions des tests précédents, notre système de stéréo-photométrie est utilisé. Il dispose de 11 directions d'éclairage distribuées sur 2 couronnes. En son centre, on retrouve une caméra et un objectif avec une distance focale de 75 mm. Ce système est conçu pour être utilisé sur site. Des acquisitions à différents temps d'exposition permettent un traitement HDR [12]. Cette méthode assure un traitement HDR linéaire avec un bruit limité.

Obtenir une vérité terrain aussi précise que les estimations en stéréo-photométrie est complexe. Nous utilisons donc des objets de même géométrie avec des modifications d'aspect. Alors que l'objet en plâtre diffus sera notre référence lambertienne sur laquelle l'estimation des normales est bonne avec toutes les méthodes, les objets avec des peintures différentes permettront d'évaluer la robustesse des méthodes aux changements de BRDF. Comme le montre la Fig. 5, notre méthode est robuste pour une grande diversité de BRDF, particulièrement en présence de valeurs spéculaires. Sur ces pixels, l'estimation de

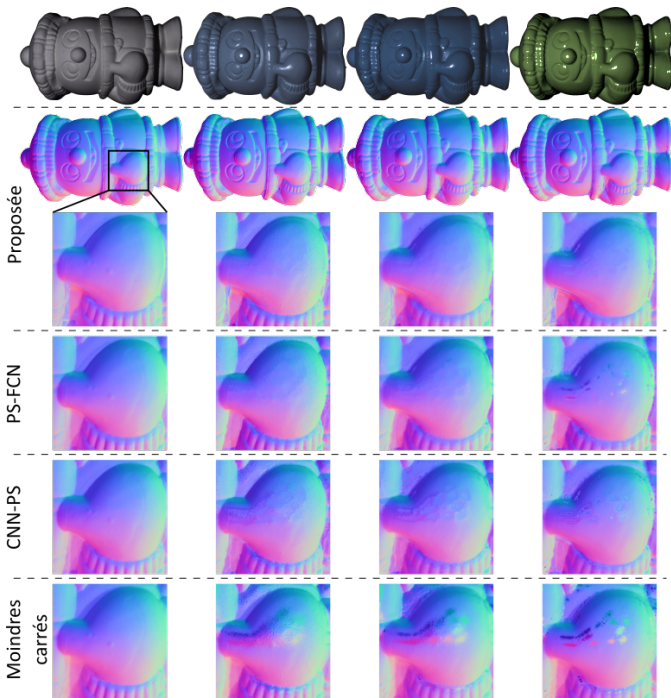


FIGURE 5 – Comparaison entre CNN-PS, PS-FCN, moindres carrés et notre méthode sur un objet de même géométrie avec des modifications d’aspect.

CNN-PS et PS-FCN n’est pas homogène. De plus, la méthode PS-FCN [2] tend à lisser ces détails et créer des artefacts sur la carte de normales comme le montre la Fig. 6. Ce phénomène permet d’obtenir de très bons résultats lorsque la vérité terrain est issue d’un autre dispositif d’acquisition. Ainsi, pour les deux expérimentations menées, nous montrons que notre approche est la plus robuste et la plus précise sur ce point.

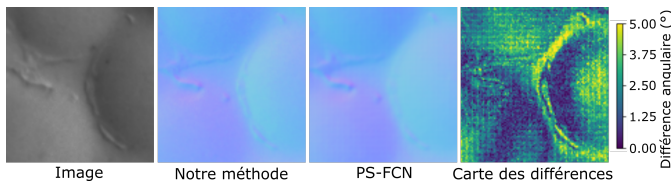


FIGURE 6 – Comparaison des méthodes locales et globales.

5 Conclusion et limitations

Dans ce travail, une nouvelle approche pour résoudre la stéréophotométrie locale calibrée sur les systèmes avec un nombre limité de lampes est présentée. Les résultats montrent que notre méthode est meilleure ou similaire à l’état de l’art. Cette conclusion est valable sur 2 bases de données différentes : DiLiGenT et acquisitions réelles avec notre système. Appliquer la stéréophotométrie pixel par pixel dans ces conditions nécessite une représentation adaptée. Pour les normales sur les ombres propres, les zones diffuses et spéculaires notre approche est robuste et précise. A l’image de toutes les méthodes de l’état de l’art, les

ombres portées sont parfois complexes à traiter. Il serait intéressant d’introduire une notion de voisinage dans notre approche.

Références

- [1] S. Barsky and M. Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10) :1239–1252, October 2003.
- [2] G. Chen, K. Han, and K. K. Wong. PS-FCN : a flexible learning framework for photometric stereo. *ECCV*, pages 3–19, 2018.
- [3] A. S. Georghiades. Incorporating the Torrance and Sparrow model of reflectance in uncalibrated photometric stereo. *IEEE ICCV*, pages 816–823 vol.2, October 2003.
- [4] E. Heitz and E. D’Eon. Importance Sampling Microfacet-Based BSDFs using the Distribution of Visible Normals. *Computer Graphics Forum*, 33(4) :103–112, July 2014.
- [5] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. *IEEE CVPR*, pages 2261–2269, jul 2017.
- [6] S. Ikehata. CNN-PS : cnn-based photometric stereo for general non-convex surfaces. *ECCV*, 11219 :3–19, 2018.
- [7] J. Li, A. Robles-Kelly, S. You, and Y. Matsushita. Learning to minify photometric stereo. *IEEE CVPR*, pages 7560–7568, 2019.
- [8] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM Transactions on Graphics*, 22(3) :759–769, July 2003.
- [9] Y. Quéau, R. Mecca, and J.-D. Durou. Unbiased photometric stereo for colored surfaces : A variational approach. *IEEE CVPR*, pages 4359–4368, 2016.
- [10] H. Santo, M. Samejima, Y. Sugano, B. Shi, and Y. Matsushita. Deep Photometric Stereo Network. *IEEE ICCVW*, pages 501–509, October 2017.
- [11] B. Shi, Z. Mo, Z. Wu, D. Duan, S.-K. Yeung, and P. Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2) :271–284, 2019.
- [12] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen. A versatile HDR video production system. *ACM Trans. Graph.*, 30(4), jul 2011.
- [13] R. J. Woodham. Photometric Method For Determining Surface Orientation From Multiple Images. *Optical Engineering*, 19(1), February 1980.
- [14] Q. Zheng, Y. Jia, B. Shi, X. Jiang, L. Duan, and A. Kot. SPLINE-Net : sparse photometric stereo through lighting interpolation and normal estimation networks. *IEEE ICCV*, pages 8548–8557, 2019.