



**HAL**  
open science

# **Stéréophotométrie avec estimation locale de l'éclairage - Application à la reconstruction 3D du patrimoine archéologique**

Benjamin Coupry, Jean Mélou, Antoine Laurent, Baptiste Brument, Pierre Gurdjos, Yvain Quéau, Jean-Denis Durou

## ► To cite this version:

Benjamin Coupry, Jean Mélou, Antoine Laurent, Baptiste Brument, Pierre Gurdjos, et al.. Stéréophotométrie avec estimation locale de l'éclairage - Application à la reconstruction 3D du patrimoine archéologique. Congrès Reconnaissance des Formes, Image, Apprentissage et Perception (RFIAP 2024), Jul 2024, Lille, France. <hal-04594418v2>

**HAL Id: hal-04594418**

**<https://hal.science/hal-04594418v2>**

Submitted on 25 Oct 2024

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.



HAL Authorization

# Stéréophotométrie avec estimation locale de l'éclairage – Application à la reconstruction 3D du patrimoine archéologique

Benjamin COUPRY<sup>1</sup> Jean MÉLOU<sup>1</sup> Antoine LAURENT<sup>1</sup> Baptiste BRUMENT<sup>1</sup>  
Pierre GURDJOS<sup>1</sup> Yvain QUÉAU<sup>2</sup> Jean-Denis DUROU<sup>1</sup>

<sup>1</sup> IRIT, UMR CNRS 5505, Université de Toulouse, France

<sup>2</sup> Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, France

benjamin.coupry@irit.fr

## Résumé

Les techniques de photogrammétrie multi-vues restituent bien la forme générale d'une scène 3D, tandis que la stéréophotométrie permet au contraire de reproduire les oscillations rapides du relief. Ces deux approches fournissant des informations complémentaires, il reste à savoir comment les faire coopérer. Nous utilisons le relief obtenu par photogrammétrie multi-vues pour estimer l'éclairage localement, ce qui nous permet de reconstruire le relief par stéréophotométrie, plus précisément et sans biais.

## Mots-clés

Stéréophotométrie, estimation de l'éclairage, reconstruction 3D, patrimoine archéologique.

## Abstract

Multi-view photogrammetry techniques reproduce the general shape of a 3D scene, while photometric stereo reproduces the rapid oscillations of relief. As these two approaches provide complementary information, the question is how to make them work together. We use the relief obtained by multi-view photogrammetry to estimate lighting locally, enabling us to reconstruct the relief by photometric stereo, more accurately and without bias.

## Keywords

Photometric stereo, lighting estimation, 3D reconstruction, archaeological heritage.

## 1 Introduction

La reconstruction 3D est essentielle dans de nombreux domaines, comme l'aide au diagnostic médical ou la création de jumeaux numériques, mais c'est dans le cadre de la numérisation du patrimoine archéologique que ce travail a pris naissance. La figure 1 représente le « Panneau des chevaux ponctués » de la grotte ornée du Pech Merle, à Cabrerets (Lot). La figure 2 montre trois champs de normales obtenus par *photogrammétrie*, terme qui regroupe différentes techniques photographiques de reconstruction 3D, d'une zone de ce panneau haute de 15 cm environ.



FIGURE 1 – « Panneau des chevaux ponctués » de la grotte du Pech Merle (hauteur  $\approx$  2 m), à Cabrerets (Lot). Source : site web du Centre de préhistoire et grotte du Pech Merle.

Les trois techniques utilisées sont classiques : *structure-from-motion*, *multi-view stereo* et stéréophotométrie. Conformément à un résultat connu de longue date [23], le résultat obtenu par stéréophotométrie est plus « piqué » que les deux autres, ce qui signifie qu'il restitue mieux les hautes fréquences du relief. Mais le résultat montré ici est également exempt de biais dans les basses fréquences du relief, grâce à une méthode d'estimation locale de l'éclairage à partir d'une connaissance partielle du relief, qui constitue notre principale contribution.

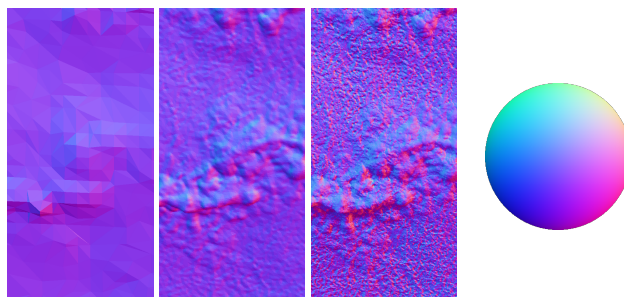


FIGURE 2 – À gauche : champs de normales d'une zone du « Panneau des chevaux ponctués » estimés par trois techniques de photogrammétrie (*structure-from-motion*, *multi-view stereo* et stéréophotométrie). À droite : représentation RVB des normales d'une sphère.

Après un bref état de l’art sur les techniques de photogrammétrie dans le paragraphe 2, où nous mettons l’accent sur la stéréophotométrie, nous présentons une méthode originale d’estimation de l’éclairage dans le paragraphe 3, que nous appliquons dans le paragraphe 4 à la stéréophotométrie, dans le contexte du patrimoine archéologique. Le paragraphe 5 améliore les résultats grâce à une estimation locale de l’éclairage. Enfin, nous résumons nos contributions et évoquons de nouvelles pistes dans le paragraphe 6.

## 2 État de l’art

Les techniques de photogrammétrie *géométriques* sont les plus connues. Le pipeline classique proposé par de nombreux logiciels [10, 25] comporte deux étapes. À partir d’une collection d’images de la scène 3D à reconstruire, obtenues sous différents angles, la technique de *structure-from-motion* estime les poses des caméras ainsi qu’un nuage de points 3D épars. La technique de *multi-view stereo* [8] utilise les poses estimées pour effectuer une reconstruction dense de la scène par maximisation de la cohérence photométrique entre les différentes vues.

Conjuguée aux approches neuronales de rendu volumique [27], cette deuxième technique peut traiter efficacement les géométries complexes et les auto-occultations. Les résultats restent malgré tout imprécis dans certaines configurations, notamment lorsque la scène 3D est faiblement texturée [30]. En outre, malgré les efforts récents en ce sens [19], la restitution des détails du relief reste limitée. Enfin, l’estimation de la réflectance de la scène constitue un verrou pour l’ensemble des techniques géométriques.

A contrario, les techniques de photogrammétrie *photométriques*, dont l’usage est moins répandu, excellent dans la récupération des hautes fréquences du champ de normales, ce qui signifie qu’elles restituent bien les détails du relief. Ces techniques reposent sur l’analyse des quantités lumineuses reçues par le récepteur photosensible de l’appareil photographique. Sous certaines précautions, le niveau de gris est proportionnel à la luminance émise par la scène, qui est caractéristique, du moins en partie, de son relief [6]. Parmi ces techniques, la stéréophotométrie (PS, pour *photometric stereo*) estime la géométrie et la réflectance d’une surface à partir d’images prises sous le même angle, mais sous différents éclairages [28]. Cette technique de reconstruction 3D est d’ailleurs la seule qui permette d’estimer la réflectance de la scène. Elle consiste à inverser le modèle de formation des images, qui lie le relief de la scène 3D, sa réflectance (réduite à l’albédo dans le cas lambertien) et l’éclairage. Si  $p$  éclairages, supposés directionnels, sont caractérisés par les vecteurs  $\mathbf{s}^i \in \mathbb{R}^3$ ,  $i \in \{1, \dots, p\}$ , alors le modèle lambertien s’écrit :

$$I^i(u, v) = \underbrace{\rho(u, v) \mathbf{n}(u, v)}_{\mathbf{m}(u, v)} \cdot \mathbf{s}^i \quad (1)$$

où  $I^i(u, v)$  désigne le niveau de gris au point  $(u, v)$ ,  $\rho(u, v)$  l’albédo et  $\mathbf{n}(u, v)$  le vecteur normal à la surface en ce point, sortant et de norme unitaire.

La *PS calibrée* suppose les éclairages connus. Sous cette hypothèse, avec  $p \geq 3$  éclairages non tous coplanaires, le système (1) admet une solution unique  $\mathbf{m}(u, v)$ , dont il est facile de déduire l’albédo et la normale au point  $(u, v)$ .

Un certain nombre de travaux ont étendu la PS calibrée aux matériaux non lambertiens [9, 22]. Plus récemment, de nombreuses solutions ont eu recours aux réseaux de neurones [14, 15]. L’hypothèse d’éclairages directionnels a également été assouplie [20]. L’ensemble de ces approches nécessitent toutefois que les images soient acquises dans un environnement parfaitement contrôlé.

Une situation qui se produit souvent dans le cadre de la numérisation du patrimoine archéologique est celle de la *PS non calibrée*, où les éclairages sont inconnus. Dans le contexte des grottes ornées, qui a motivé cette étude, il arrive en effet que l’accès aux parois soit limité pour des raisons de conservation des sols, auquel cas la source d’éclairage est dirigée « à la volée » vers la zone d’intérêt.

Il est connu que la PS non calibrée constitue un problème mal posé [2, 12], pour lequel la parade usuelle consiste à introduire un a priori [11, 26, 29]. Une autre démarche, plus proche de celle qui sera présentée dans le paragraphe 3, consiste à utiliser la scène 3D comme mire d’étalonnage [13], à ceci près qu’un certain nombre de points de la scène sont supposés avoir le même albédo.

Les approches neuronales de PS non calibrée font généralement l’hypothèse que les éclairages sont directionnels. Un premier réseau détermine les éclairages [5], qui rendent le problème calibré, mais cette approche échoue si l’hypothèse d’éclairages directionnels est mise en défaut.

La *PS universelle*, qui a été proposée très récemment pour limiter ce risque [16, 17], s’accommode de conditions d’éclairage très variées, apprises par *deep learning*. Il s’agit sans doute aujourd’hui de l’approche la plus performante pour résoudre la PS non calibrée. Néanmoins, le résultat de gauche de la figure 3 montre que, si la scène 3D est trop différente des scènes ayant servi à l’apprentissage, l’inférence des éclairages peut être erronée, ce qui se traduit par une inversion concave/convexe du sol, au bas de la paroi. Le résultat de droite bénéficie d’une meilleure estimation de l’éclairage, qui sera décrite dans le paragraphe 3.

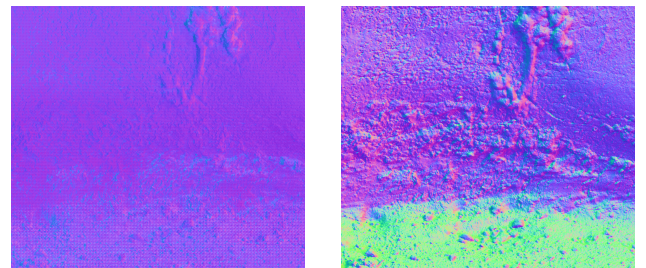


FIGURE 3 – Champs de normales d’une même zone du « Panneau des chevaux ponctués » située au bas de la paroi, estimés : par PS « universelle » [17]; par PS calibrée, les éclairages étant estimés par la méthode décrite ci-après.

L'estimation de l'albédo [17], qui est une prérogative de la PS, est également moins précise lorsqu'elle est effectuée par inférence, comme le montrent les résultats de la figure 4 : l'inscription sur le trépied est illisible, contrairement au résultat obtenu par PS calibrée.



FIGURE 4 – Albédos estimés par PS « universelle » [17] et par PS calibrée. L'albédo est nettement mieux estimé avec la PS calibrée : l'inscription sur le trépied devient lisible.

### 3 Estimation de l'éclairage

**Utilisation d'une mire d'étalonnage** – La mise en œuvre de la PS calibrée dans un contexte non contrôlé, c'est-à-dire hors laboratoire, nécessite de positionner une mire d'étalonnage dans la scène au moment de la prise de vues, cette mire étant le plus souvent une sphère. Si celle-ci est *brillante*, la direction de l'éclairage peut être estimée à partir de la position du point brillant, en utilisant la loi de réflexion de la lumière, mais il est alors impossible d'estimer l'intensité lumineuse. C'est pourquoi le protocole le plus classique d'estimation de l'éclairage consiste à placer une sphère blanche *mate* dans la scène [7]. En effet, la normale  $\mathbf{n}(u, v)$  et l'albédo  $\rho(u, v)$  étant alors connus en tout point  $(u, v)$  de la sphère, il est facile d'inverser le modèle (1) pour en déduire les  $p$  vecteurs d'éclairage  $\mathbf{s}^i$ .

Cependant, à moins que le flux lumineux incident soit parfaitement parallèle et uniforme, ce qui est difficile à garantir, une première limitation de cette méthode d'étalonnage est d'étendre à l'ensemble de la scène l'éclairage estimé à l'endroit où se trouve la sphère. Cela se traduit par un biais dans les basses fréquences du champ de normales estimé ensuite par PS calibrée. Ce phénomène peut être vérifié qualitativement en comparant le champ de normales de la figure 11, déduit du maillage 3D obtenu par MVS, donc réputé fiable dans les basses fréquences, avec le champ de normales de la figure 6, obtenu par PS calibrée à partir de 23 photographies telles que celle de la figure 5, en utilisant les éclairages estimés sur la sphère d'étalonnage positionnée en bas à droite de la paroi.

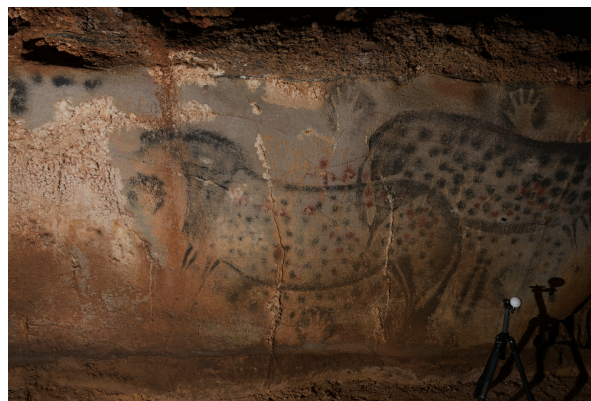


FIGURE 5 – Une image de stéréophotométrie (parmi 23) du « Panneau des chevaux ponctués ». Une sphère d'étalonnage a été positionnée en bas à droite de la paroi. Un appareil photographique Nikon Z7 est utilisé « en tout manuel » (résolution des images :  $8256 \times 5504$  pixels).

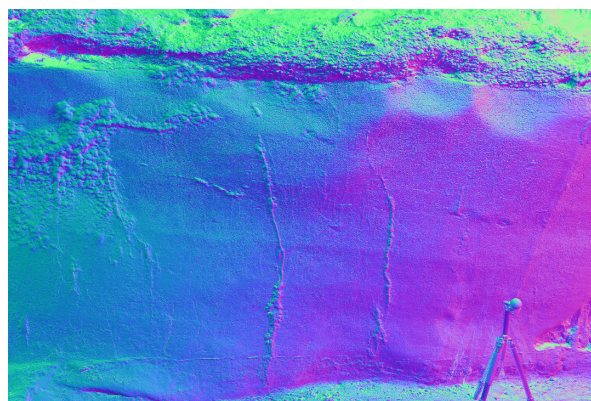


FIGURE 6 – Champ de normales estimé par PS calibrée, à partir de 23 images telles que celles de la figure 5. L'éclairage estimé sur la sphère située en bas à droite est étendu à l'ensemble de la scène, ce qui se traduit par un biais dans les basses fréquences du champ de normales.

L'utilisation d'une sphère d'étalonnage peut également se révéler problématique pour d'autres raisons, puisque celle-ci doit être placée au plus près de la scène, sans l'occulter ni provoquer d'ombre. En outre, dans le contexte du patrimoine archéologique, il est inconcevable qu'elle puisse se trouver en contact avec l'objet ou la scène à numériser. Pour toutes ces raisons, l'étalonnage de l'éclairage doit, si possible, être effectué autrement.

**Utilisation de la scène 3D comme mire d'étalonnage** – Il est également possible d'estimer l'éclairage en utilisant la scène 3D elle-même comme mire d'étalonnage, pour peu que l'on connaisse la normale en un certain nombre de points 3D. Nous verrons plus loin que ce nombre peut être faible, et que la connaissance de la normale en ces points tolère une imprécision relativement importante. Quant à savoir comment obtenir cette information, différentes possibilités seront décrites dans le paragraphe 4.

Comme cela a déjà dit dans le paragraphe 2, l'utilisation de la scène 3D comme mire d'étalonnage n'est pas nouvelle. Il semble à première vue insuffisant de connaître la normale  $\mathbf{n}(u, v)$  en  $q$  points de la scène pour estimer les  $p$  vecteurs d'éclairage  $\mathbf{s}^i$  à partir des équations (1), si les albédos  $\rho(u, v)$  sont inconnus. Or, si l'albédo d'une sphère d'étalonnage de couleur uniforme est connu, certes à un facteur près, il n'en va pas de même d'une scène 3D quelconque. Pour cette raison, il est généralement demandé de savoir identifier des points de même albédo pour utiliser la scène 3D comme mire d'étalonnage [11, 13].

Notre principale contribution consiste à reformuler le problème de façon à éviter cette contrainte sur l'albédo. Si la normale est connue en  $q$  points  $(u_j, v_j)$ ,  $j \in \{1, \dots, q\}$ , les  $q$  albédos  $\rho(u_j, v_j)$  peuvent être considérés comme des inconnues supplémentaires. En introduisant les  $q$  nouvelles inconnues  $\alpha_j$  définies par :

$$\alpha_j = \frac{1}{\rho(u_j, v_j)} \quad (2)$$

il est facile de reformuler les équations (1),  $i \in \{1, \dots, p\}$ ,  $j \in \{1, \dots, q\}$ , sous la forme suivante :

$$I^i(u_j, v_j) \alpha_j - \mathbf{n}(u_j, v_j) \cdot \mathbf{s}^i = 0 \quad (3)$$

La définition (2) suppose implicitement que  $\rho(u_j, v_j) \neq 0$ , c'est-à-dire que les points 3D utilisés pour l'estimation de l'éclairage ne sont pas noirs. Il est de toute façon préférable de disqualifier de tels points, ne serait-ce que pour éviter d'utiliser des niveaux de gris proches de 0, qui sont sujets au « bruit thermique » du capteur. Fort heureusement, ces points sont faciles à identifier, leurs niveaux de gris étant quasiment nuls sous n'importe quel éclairage.

Le système formé par les  $pq$  équations linéaires (3) comporte  $3p + q$  inconnues. Il est donc sur-déterminé dès lors que  $pq \geq 3p + q$ . Sachant que  $p \geq 3$ , cette condition est vérifiée si  $q \geq 5$ , une condition facile à garantir en pratique. Cependant, les équations (3) ne sont pas seulement linéaires, elles sont également homogènes. Pour éviter la solution triviale, à savoir  $\alpha_j = 0, \forall j \in \{1, \dots, q\}$ , et  $\mathbf{s}^i = [0, 0, 0]^T, \forall i \in \{1, \dots, p\}$ , il est nécessaire d'ajouter des contraintes. Or, un albédo représente une proportion d'énergie lumineuse. Il doit donc être positif et inférieur à 1. Par conséquent, le système d'équations (3) doit être complété par les  $q$  contraintes de type inégalité suivantes :

$$\alpha_j \geq 1, \quad j \in \{1, \dots, q\} \quad (4)$$

Afin d'assurer une certaine robustesse, l'estimation couplée des albédos et des éclairages peut alors être réécrite comme un problème d'optimisation en norme 1 :

$$\begin{aligned} \min_{\{\alpha_j\}, \{\mathbf{s}^i\}} & \sum_{i=1}^p \sum_{j=1}^q |I^i(u_j, v_j) \alpha_j - \mathbf{n}(u_j, v_j) \cdot \mathbf{s}^i| \\ \text{s. c.} & \alpha_j \geq 1, \quad j \in \{1, \dots, q\} \end{aligned} \quad (5)$$

qui peut être résolu, par exemple, par la méthode des moindres carrés repondérés.

**Validation sur le banc d'essai DiLiGenT-MV** – Pour valider cette méthode d'estimation de l'éclairage, nous utilisons le jeu de données réelles DiLiGenT-MV [18], constitué de cinq objets possédant des réflectances et des géométries variées. Chaque objet est placé sur un plateau tournant, de manière à le photographier sous 20 angles. Notons que l'aspect multi-vues de ce jeu de données n'est pas mis à profit par notre méthode, si ce n'est qu'il procure un plus grand nombre de données de test. Pour chacune de ces 20 vues, l'objet est photographié sous 96 éclairages parfaitement directionnels.

Ce jeu de données est idéalement adapté à l'évaluation de notre méthode d'estimation de l'éclairage, dans la mesure où la vérité terrain des normales et des éclairages est également fournie. Sur la figure 7, chacun des 96 éclairages est représenté par un point caractéristique de sa direction et une couleur caractéristique de son intensité, ce qui permet de représenter l'ensemble des éclairages sur un même disque unité.

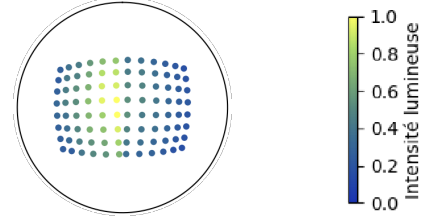


FIGURE 7 – À gauche : vérité terrain des 96 éclairages de DiLiGenT-MV. Chaque éclairage est représenté par un point caractéristique de sa direction et une couleur caractéristique de son intensité. À droite : échelle des intensités.

Pour commencer, nous utilisons la vérité terrain de toutes les normales de chaque objet pour estimer l'éclairage, à ceci près que pour les objets les plus brillants (Cow et Reading), nous masquons préalablement les taches spéculaires pour rendre l'estimation plus robuste.

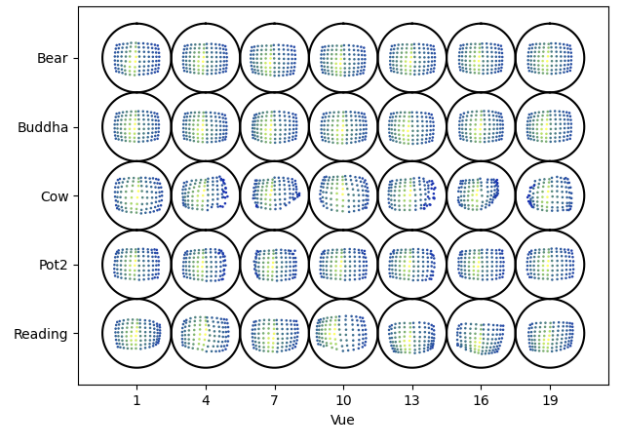


FIGURE 8 – Éclairages estimés avec les normales vérité terrain des objets de DiLiGenT-MV, pour différentes vues.

Les éclairages de la figure 8 peuvent être comparés qualitativement à ceux de la figure 7, qui constituent la vérité terrain. Quant à la table 1, elle recense l’erreur angulaire moyenne de l’estimation robuste de l’éclairage, qui résulte de la résolution du problème (5) en norme 1, et celle de la résolution en norme 2, moins robuste mais également moins coûteuse en temps de calcul. Dans les deux cas, les estimations sont très proches de la vérité terrain.

	Bear	Buddha	Cow	Pot2	Reading
Norme 2	3.9	3.6	6.6	3.1	16.4
Norme 1	3.3	2.8	6.9	3.1	7.1

TABLE 1 – Erreur angulaire moyenne (en degrés) calculée sur l’ensemble des vues et l’ensemble des éclairages, en utilisant la quasi-totalité des normales de la vérité terrain.

La figure 9 montre l’influence sur la direction de l’éclairage estimé du nombre de normales utilisées, qui sont choisies aléatoirement parmi les normales de la vérité terrain. Un nombre relativement faible de normales ( $\approx 50$ ) est suffisant pour les matériaux lambertiens (Bear, Buddha, Pot2), tandis qu’un nombre de normales plus élevé est nécessaire pour les matériaux spéculaires (Cow, Reading).

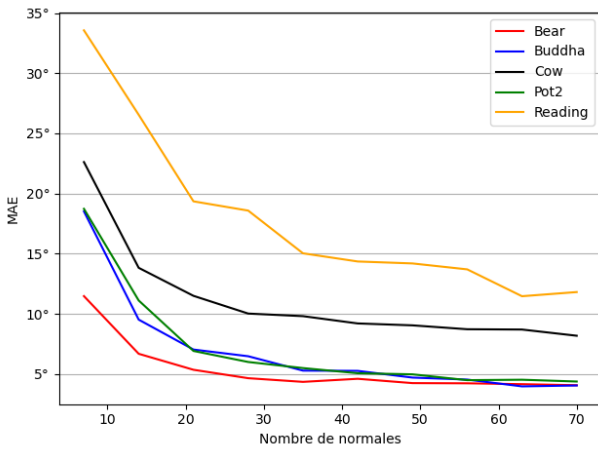


FIGURE 9 – Erreur angulaire moyenne de la direction d’éclairage, en fonction du nombre de normales utilisées, pour les différents objets de DiLiGenT-MV (moyennes calculées sur l’ensemble des vues et des éclairages).

Enfin, nous testons la tolérance au bruit de notre méthode d’estimation. Pour ce faire, 1000 normales de la vérité terrain, tirées aléatoirement, sont bruitées par une loi normale appliquée à leurs angles d’Euler. La figure 10 montre l’influence du bruit sur la précision des éclairages estimés, pour différents écarts-types de la loi normale. Au vu de ces résultats, notre méthode semble stable tant que l’écart-type de bruit angulaire sur les normales est inférieur à  $10^\circ$ .

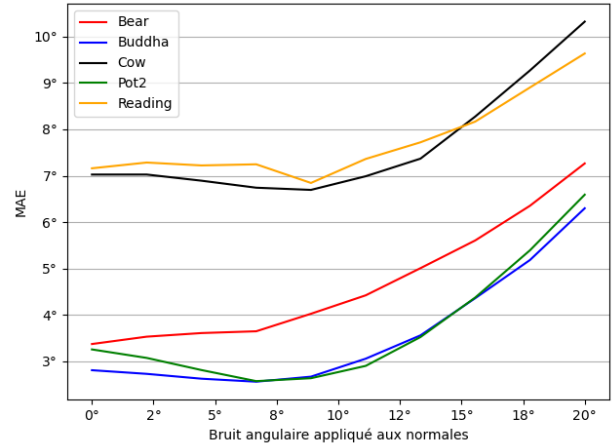


FIGURE 10 – Erreur angulaire moyenne de la direction d’éclairage, en fonction du bruit angulaire sur les normales, pour les différents objets de DiLiGenT-MV (moyennes calculées sur l’ensemble des vues et des éclairages).

## 4 Stéréophotométrie

La méthode d’estimation de l’éclairage décrite dans le paragraphe 3 répond à notre souhait de ne plus utiliser de sphère d’étalonnage, mais elle nécessite de connaître la normale en  $q \geq 5$  points de la scène.

**Obtention des normales** – Les normales nécessaires à la mise en œuvre de notre méthode d’estimation de l’éclairage peuvent être obtenues de différentes façons. Si l’on dispose du matériel nécessaire, l’utilisation de données RGB-D offre une bonne approximation des basses fréquences du relief de la scène [21]. D’autres méthodes mono-vue permettent d’inférer un champ de normales [1]. Il semble cependant plus judicieux d’utiliser les techniques classiques de photogrammétrie géométriques, *structure-from-motion* (SfM) et *multi-view stereo* (MVS), dont nous avons déjà dit que les biais d’estimation du relief dans les basses fréquences sont faibles. Mais s’il paraît naturel d’utiliser les normales déduites du MVS (cf. figure 11), le temps de calcul nécessaire à l’obtention de ces normales peut s’avérer dissuasif.

Pour réduire le temps de calcul, il est nécessaire d’éviter l’étape de MVS du processus de reconstruction 3D, en estimant les normales soit directement après la mise en correspondance [4], soit en lissant le nuage de points 3D épars obtenu par SfM. Les normales estimées de cette façon, quoiqu’imprécises dans les hautes fréquences, sont suffisamment précises dans les basses fréquences pour répondre à nos attentes. La figure 12 montre que ces normales sont effectivement très proches de celles de la figure 11. En outre, cette deuxième approche est très simple à mettre en œuvre, puisqu’il suffit de compléter le jeu d’images de stéréophotométrie, qui sont toutes prises sous le même angle, par une photographie prise sous un angle différent.

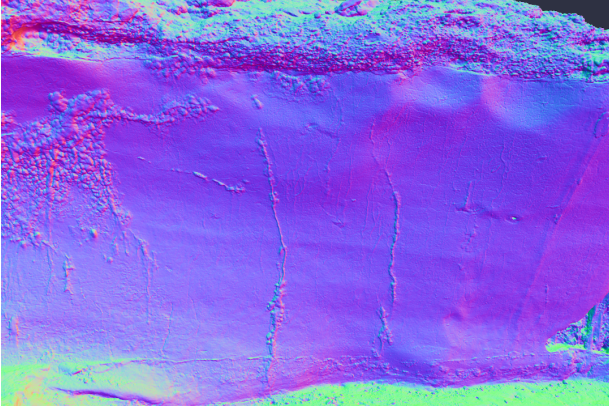


FIGURE 11 – Champ de normales déduit du maillage 3D obtenu par MVS. Les basses fréquences du relief sont bien restituées, au prix d’un temps de calcul très élevé.

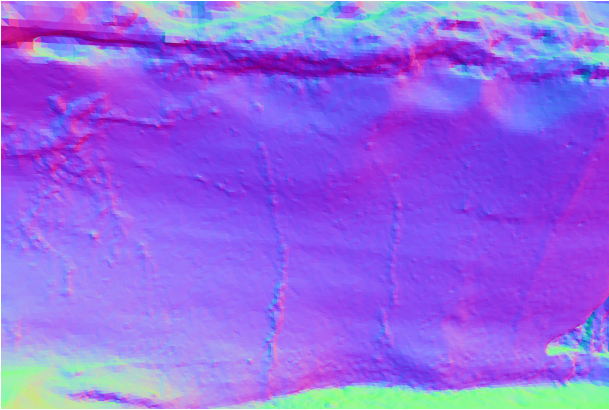


FIGURE 12 – Champ de normales déduit du maillage 3D obtenu en lissant le nuage de points 3D obtenu par SfM. Les basses fréquences du relief sont aussi bien restituées que sur la figure 11, mais bien plus rapidement.

**Nécessité d’estimer l’éclairage localement** – Si l’hypothèse d’un éclairage parallèle et uniforme est garanti pour les données de DiLiGenT-MV, il n’en va pas toujours de même lors d’une prise de vues in situ, en particulier dans le contexte de la numérisation du patrimoine archéologique. Par exemple, dans le cas du « Panneau des chevaux ponctués », l’éclairage de la paroi par un spot est loin de vérifier cette hypothèse, ce qui se traduit par un biais important dans les basses fréquences du champ de normales estimé par PS calibrée, lorsque les éclairages sont estimés par la méthode décrite dans le paragraphe 3 (cf. figure 13). La sphère d’étalonnage présente dans la scène n’a pas été utilisée pour obtenir ce résultat, ce qui correspond bien à notre objectif initial, mais il est décevant d’utiliser les normales obtenues par SfM pour estimer les éclairages, normales qui sont censées restituer fidèlement les basses fréquences du relief, si cette propriété ne se répercute pas sur le résultat final. Une parade consiste à généraliser la méthode présentée dans le paragraphe 3 à l’estimation locale de l’éclairage.

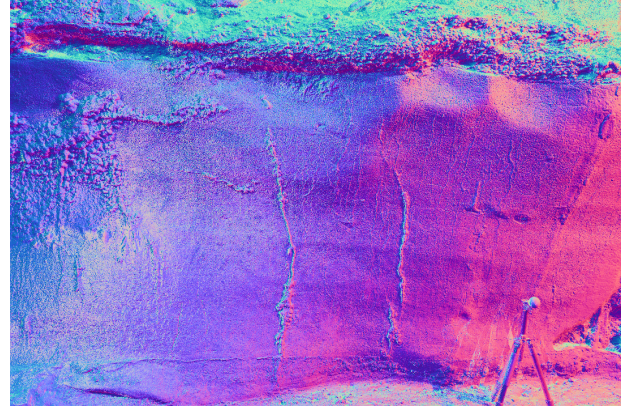


FIGURE 13 – Champ de normales estimé par PS calibrée, en estimant les éclairages à partir des normales obtenues par SfM. Le résultat est comparable à celui de la figure 6, certes sans utiliser la sphère d’étalonnage.

## 5 Estimation locale de l’éclairage

Dans ce paragraphe, nous nous proposons de généraliser notre méthode à l’estimation locale de l’éclairage. Cela suppose d’introduire une dépendance spatiale dans les  $p$  vecteurs  $\mathbf{s}^i \in \mathbb{R}^3$ ,  $i \in \{1, \dots, p\}$ , c’est-à-dire de modifier légèrement le modèle lambertien (1), comme cela a déjà été proposé, par exemple dans [11, 24] :

$$I^i(u, v) = \rho(u, v) \mathbf{n}(u, v) \cdot \mathbf{s}^i(u, v) \quad (6)$$

En chaque point de la surface, la direction et l’intensité de l’éclairage sont inconnues. Une estimation dense de l’éclairage, c’est-à-dire en chaque point de la scène, est impossible à effectuer, puisque chaque estimation nécessite de connaître les normales de  $q \geq 5$  points. Nous proposons donc, dans la veine des travaux présentés dans [11], de définir  $r$  voisinages  $\mathcal{V}_k$ ,  $k \in \{1, \dots, r\}$ , supposés infinitésimaux au regard de l’échelle de variation de l’éclairage dans la scène, de telle sorte que l’éclairage  $\mathbf{s}^i(u, v)$ ,  $i \in \{1, \dots, p\}$ , puisse être assimilé à un éclairage directionnel  $\mathbf{s}_k^i$  sur chaque voisinage  $\mathcal{V}_k$ ,  $k \in \{1, \dots, r\}$  :

$$\mathbf{s}^i(u, v) \equiv \mathbf{s}_k^i, \quad \forall (u, v) \in \mathcal{V}_k \quad (7)$$

De (6) + (7), nous déduisons le modèle lambertien local :

$$I^i(u, v) = \rho(u, v) \mathbf{n}(u, v) \cdot \mathbf{s}_k^i, \quad \forall (u, v) \in \mathcal{V}_k \quad (8)$$

Il devient alors possible d’estimer chaque éclairage local  $\mathbf{s}_k^i$  en utilisant les normales contenues dans le voisinage correspondant. Néanmoins, ces estimations locales ne doivent pas être effectuées indépendamment, à cause d’une ambiguïté dans le modèle local (8), qui peut être réécrit, en introduisant un coefficient réel  $\mu_k > 0$  :

$$I^i(u, v) = \frac{\rho(u, v)}{\mu_k} \mathbf{n}(u, v) \cdot (\mu_k \mathbf{s}_k^i), \quad \forall (u, v) \in \mathcal{V}_k \quad (9)$$

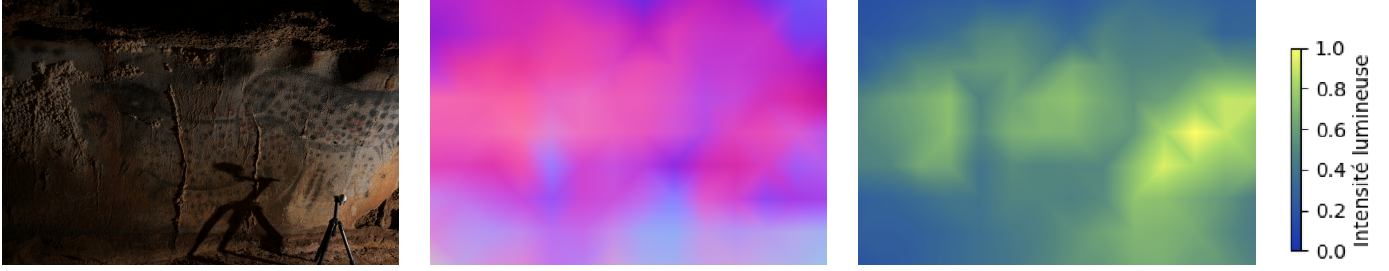


FIGURE 14 – De gauche à droite : exemple d’image du « Panneau des chevaux ponctués » ; direction de l’éclairage estimé localement, affichée en représentation RVB ; intensité de l’éclairage estimé localement, affichée en fausses couleurs.

Par conséquent, l’estimation locale de l’éclairage, voisinage par voisinage, risque de provoquer un effet de « tuilage » des éclairages  $\mathbf{s}^i$ , lequel se répercutera inévitablement dans le champ de normales estimé par PS calibrée. Il est donc nécessaire de mener l’estimation de tous les vecteurs  $\mathbf{s}_k^i$  simultanément ? Cela revient à résoudre le problème suivant, qui est une généralisation de (5) :

$$\begin{aligned} \min_{\{\alpha_j\}, \{\mathbf{s}_k^i\}} & \sum_{i=1}^p \sum_{k=1}^r \sum_{(u_j, v_j) \in \mathcal{V}_k} |I^i(u_j, v_j) \alpha_j - \mathbf{n}(u_j, v_j) \cdot \mathbf{s}_k^i| \\ \text{s. c.} & \quad \alpha_j \geq 1, \quad j \in \{1, \dots, q\} \end{aligned} \quad (10)$$

Les voisinages  $\mathcal{V}_k$  doivent se chevaucher. En effet, si un point  $(u_j, v_j)$  appartient à deux voisinages  $\mathcal{V}_{k_1}$  et  $\mathcal{V}_{k_2}$ , la cohérence entre les estimations des éclairages  $\mathbf{s}_{k_1}^i$  et  $\mathbf{s}_{k_2}^i$ ,  $i \in \{1, \dots, m\}$ , découle de la présence de  $\alpha_j$  dans les inconnues de (10). Il est également notable que les estimations des inconnues  $\alpha_j$  ne sont pas utilisées, non pas que ces estimations soient peu fiables, mais parce qu’elles sont éparpillées alors que l’albédo ré-estimé par PS calibrée prendra la forme d’une fonction dense (cf. figure 18).

L’estimation locale de l’éclairage permet de restituer les variations de direction et d’intensité de l’éclairage, qui peuvent être très importantes sur des scènes telles que le « Panneau des chevaux ponctués » de la grotte du Pech Merle, comme le montrent les résultats de la figure 14.

Pour les objets du banc d’essai DiLiGenT-MV, de petites fluctuations de l’intensité des éclairages peuvent être mises en évidence, comme le montre la figure 15. Même si toutes les précautions ont été prises pour que le flux lumineux *primaire* soit parfaitement parallèle et uniforme, le flux lumineux incident au niveau de la surface perd manifestement cette qualité, sans doute à cause de flux *secondaires* dus aux rebonds sur les parties concaves de l’objet.

Les figures 17 et 18 montrent les résultats obtenus par une version robuste de PS calibrée utilisant cette méthode d’estimation locale de l’éclairage. Le champ de normales estimé (cf. figure 17) semble à la fois fidèle dans les basses fréquences au champ de normales de la figure 11, qui a été déduit du MVS, et fidèle dans les hautes fréquences au champ de normales de la figure 13, qui a été obtenu par PS calibrée à partir d’une estimation globale des éclairages.

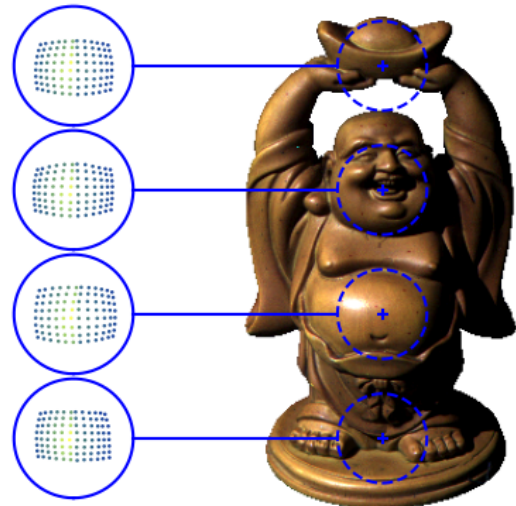


FIGURE 15 – Éclairages estimés en différents voisinages de l’objet Buddha de DiLiGenT-MV. Les directions des éclairages estimés varient très peu d’un point à l’autre de la surface, mais leurs intensités présentent de petites fluctuations. Ces éclairages sont donc parallèles, mais pas parfaitement uniformes au niveau de la surface.

Les normales obtenues par MVS, réputées fiables dans les basses fréquences, peuvent servir de référence pour les mesures d’écart angulaire. La figure 16 montre les cartes d’écart angulaire de trois champs de normales estimés par PS calibrée, pour lesquels les éclairages ont été estimés de trois manières différentes. C’est clairement la méthode d’estimation locale de l’éclairage qui restitue le mieux les basses fréquences du relief. Avec les deux autres méthodes, une déformation importante affecte les normales estimées.

Malgré tout, le résultat de droite de la figure 16 révèle un écart angulaire élevé en certains points de la scène 3D, en particulier au niveau de la sphère d’étalonnage, car l’éclairage n’y a pas été estimé localement. De plus, si le champ de normales déduit du MVS est fiable dans les basses fréquences, il n’en va pas de même des hautes fréquences. Les détails du relief de la paroi peuvent donc se traduire par des écarts angulaires élevés, mais par la faute d’un relief de référence biaisé.

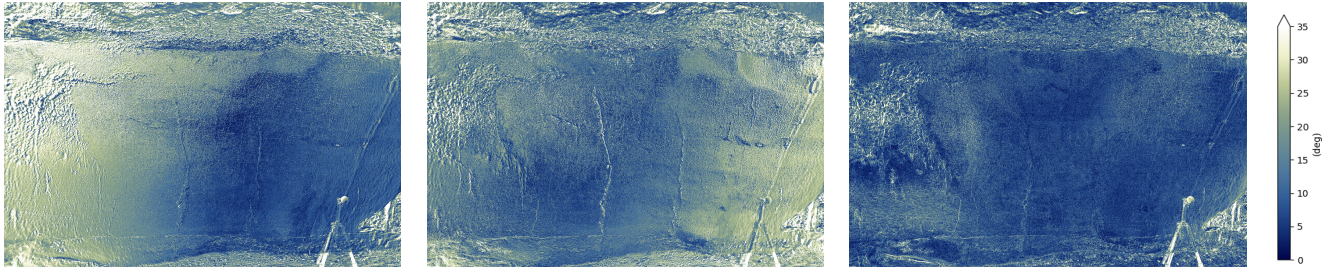


FIGURE 16 – Écart angulaire entre les normales déduites du maillage 3D obtenu par MVS (cf. figure 17) et trois champs de normales estimés par PS calibrée, où les éclairages sont estimés, de gauche à droite : avec une sphère d'étalonnage (cf. figure 6) ; avec notre méthode d'estimation globale (cf. figure 13) ; avec notre méthode d'estimation locale (cf. figure 17). Ce dernier résultat est clairement celui pour lequel les basses fréquences du relief sont les mieux restituées.

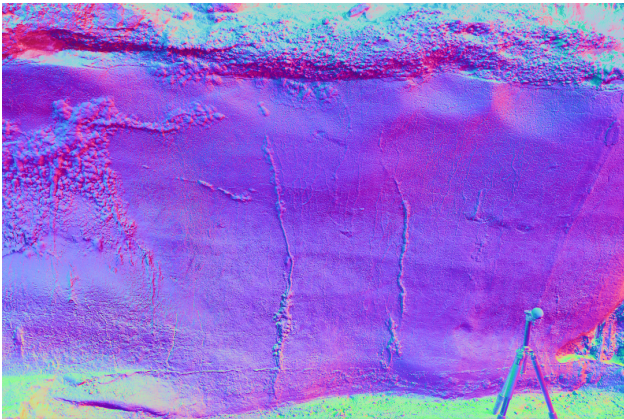


FIGURE 17 – Champ de normales estimé par PS calibrée, en utilisant l'estimation locale de l'éclairage à partir des normales déduites du SfM. En comparaison des résultats des figures 6 et 13, celui-ci est bien moins biaisé dans les basses fréquences.



FIGURE 18 – Albédo associé aux normales de la figure 17, obtenu par PS calibrée en utilisant l'estimation locale de l'éclairage à partir des normales déduites du SfM.

## 6 Conclusion et perspectives

Dans cet article, nous présentons une nouvelle méthode d'estimation locale de l'éclairage, qui améliore considérablement les résultats de stéréophotométrie calibrée. Cette étude préliminaire s'appuie sur un certain nombre d'évaluations qualitatives et quantitatives, mais il reste à effectuer ces comparaisons d'une façon plus systématique.

D'autre part, une étude comparative doit être menée avec différentes méthodes de reconstruction 3D, notamment les méthodes de stéréophotométrie non calibrée qui effectuent un pré-traitement d'estimation de l'éclairage [5].

À plus long terme, notre méthode pourra être étendue aux données de stéréophotométrie multi-vues, une technique qui fait actuellement l'objet de très nombreux travaux, grâce aux avancées spectaculaires des approches neuronales [3, 31].

## Remerciements

Le relevé photographique in situ a été réalisé en mars 2024 dans le cadre d'une opération archéologique programmée. Nous remercions la mairie de Cabrerets, le Centre de préhistoire du Pech Merle et le Service régional de l'archéologie pour leur confiance.

## Références

- [1] Bae, G., Davison, A.J.: Rethinking Inductive Biases for Surface Normal Estimation. In: CVPR (2024)
- [2] Belhumeur, P.N., Kriegman, D.J.: What is the set of images of an object under all possible illumination conditions? IJCV **28**, 245–260 (1998)
- [3] Brument, B., Bruneau, R., Quéau, Y., Mélou, J., Lauze, F., Durou, J.D., Calvet, L.: RNb-NeuS: Reflectance and Normal-Based Multi-View 3D Reconstruction. In: CVPR (2024)
- [4] Calvet, L., Maignan, N., Brument, B., Mélou, J., Tozza, S., Durou, J.D., Quéau, Y.: Multi-view Normal Estimation—Application to Slanted Plane-Sweeping. In: SSVM. Springer (2023)

- [5] Chen, G., Han, K., Shi, B., Matsushita, Y., Wong, K.Y.K.: Self-calibrating deep photometric stereo networks. In: CVPR (2019)
- [6] Durou, J.D., Falcone, M., Quéau, Y., Tozza, S.: A Comprehensive Introduction to Photometric 3D-reconstruction. In: *Advances in Photometric 3D-Reconstruction*, pp. 1–29. Springer (2020)
- [7] Fainsin, L., Mélou, J., Calvet, L., Carlier, A., Durou, J.D.: Neural detection of spheres in images for lighting calibration. In: *Sixteenth International Conference on Quality Control by Artificial Vision* (2023)
- [8] Furukawa, Y., Hernández, C., et al.: Multi-view stereo: A tutorial. *Foundations and Trends in Computer Graphics and Vision* **9**(1-2), 1–148 (2015)
- [9] Goldman, D.B., Curless, B., Hertzmann, A., Seitz, S.M.: Shape and Spatially-Varying BRDFs from Photometric Stereo. *PAMI* **32**(6), 1060–1071 (2009)
- [10] Griwodz, C., Gasparini, S., Calvet, L., Gurdjos, P., Castan, F., Maujean, B., Lillo, G.D., Lanthony, Y.: AliceVision Meshroom: An open-source 3D reconstruction pipeline. In: *MMSys* (2021)
- [11] Guo, H., Mo, Z., Shi, B., Lu, F., Yeung, S.K., Tan, P., Matsushita, Y.: Patch-based uncalibrated photometric stereo under natural illumination. *PAMI* **44**(11), 7809–7823 (2021)
- [12] Hayakawa, H.: Photometric stereo under a light source with arbitrary motion. *JOSA A* **11**(11), 3079–3089 (1994)
- [13] Hernandez, C., Vogiatzis, G., Cipolla, R.: Multiview photometric stereo. *PAMI* **30**(3), 548–554 (2008)
- [14] Ikehata, S.: CNN-PS: CNN-based photometric stereo for general non-convex surfaces. In: *ECCV* (2018)
- [15] Ikehata, S.: Ps-transformer: Learning sparse photometric stereo network using self-attention mechanism. *arXiv preprint arXiv:2211.11386* (2022)
- [16] Ikehata, S.: Universal photometric stereo network using global lighting contexts. In: *CVPR* (2022)
- [17] Ikehata, S.: Scalable, detailed and mask-free universal photometric stereo. In: *CVPR* (2023)
- [18] Li, M., Zhou, Z., Wu, Z., Shi, B., Diao, C., Tan, P.: Multi-view photometric stereo: A robust solution and benchmark dataset for spatially varying isotropic materials. *TIP* **29**, 4159–4173 (2020)
- [19] Li, Z., Müller, T., Evans, A., Taylor, R.H., Unberath, M., Liu, M.Y., Lin, C.H.: Neuralangelo: High-Fidelity Neural Surface Reconstruction. In: *CVPR* (2023)
- [20] Lichy, D., Sengupta, S., Jacobs, D.W.: Fast lightweight near-field photometric stereo. In: *CVPR* (2022)
- [21] Maier, R., Kim, K., Cremers, D., Kautz, J., Nießner, M.: Intrinsic3D: High-quality 3D reconstruction by joint appearance and geometry optimization with spatially-varying lighting. In: *ICCV* (2017)
- [22] Mecca, R., Quéau, Y., Logothetis, F., Cipolla, R.: A single-lobe photometric stereo approach for heterogeneous material. *Journal on Imaging Sciences* **9**(4), 1858–1888 (2016)
- [23] Nehab, D., Rusinkiewicz, S., Davis, J., Ramamoorthi, R.: Efficiently combining positions and normals for precise 3D geometry. *TOG* **24**(3), 536–543 (2005)
- [24] Quéau, Y., Lauze, F., Durou, J.D.: A L1-TV Algorithm for Robust Perspective Photometric Stereo with Spatially-Varying Lightings. In: *SSVM* (2015)
- [25] Schönberger, J.L., Frahm, J.M.: Structure-from-Motion Revisited. In: *CVPR* (2016)
- [26] Shi, B., Matsushita, Y., Wei, Y., Xu, C., Tan, P.: Self-calibrating photometric stereo. In: *CVPR* (2010)
- [27] Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W.: NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. In: *NeurIPS* (2021)
- [28] Woodham, R.J.: Photometric method for determining surface orientation from multiple images. *Optical engineering* **19**(1), 139–144 (1980)
- [29] Wu, Z., Tan, P.: Calibrating photometric stereo by holistic reflectance symmetry analysis. In: *CVPR* (2013)
- [30] Xu, Q., Kong, W., Tao, W., Pollefeys, M.: Multi-scale geometric consistency guided and planar prior assisted multi-view stereo. *PAMI* **45**(4), 4945–4963 (2022)
- [31] Zhao, D., Lichy, D., Perrin, P.N., Frahm, J.M., Sengupta, S.: Mvpsnet: Fast generalizable multi-view photometric stereo. In: *ICCV* (2023)