

Stéréophotométrie par *Gaussian Splatting* et rendu inverse

Matéo DUCASTEL Yvain QUÉAU David TSCHUMPERLÉ

Université Caen Normandie, ENSICAEN, CNRS, Normandie Univ, GREYC UMR 6072, F-14000 Caen, France

Résumé – Les algorithmes récents de l’état de l’art en stéréophotométrie recourent à des réseaux de neurones et fonctionnent soit par apprentissage à priori, soit par optimisation par rendu inverse. Ici, nous revisitons le problème de la stéréophotométrie calibrée en exploitant les avancées récentes en rendu inverse 3D, avec le formalisme des *Gaussian Splatting*. Ceci nous permet de paramétriser la scène 3D à reconstruire, et de l’optimiser de manière plus explicite. Notre approche intègre un modèle simplifié de représentation de la lumière, et montre le potentiel du moteur de rendu des *Gaussian Splatting* pour le problème de la stéréophotométrie.

Abstract – Recent state-of-the-art algorithms in photometric stereo rely on neural networks and operate either through prior learning or inverse rendering optimization. Here, we revisit the problem of calibrated photometric stereo by leveraging recent advances in 3D inverse rendering using the *Gaussian Splatting* formalism. This allows us to parameterize the 3D scene to be reconstructed and optimize it in a more interpretable manner. Our approach incorporates a simplified model for light representation and demonstrates the potential of the *Gaussian Splatting* rendering engine for the photometric stereo problem.

1 Introduction

L’un des objectifs de la stéréophotométrie est d’estimer une carte de normales 3D décrivant la surface d’un objet à partir de plusieurs photographies prises avec une caméra fixe, mais avec des positions d’éclairage différentes. Ici, nous proposons d’explorer une nouvelle approche pour résoudre le problème de la stéréophotométrie, approche fondée sur le formalisme récent des *Gaussian Splatting*. Ce formalisme consiste à modéliser une scène comme un champ de gaussiennes dans un espace 3D, auxquelles sont attachés différents attributs (orientation, taille et couleur notamment). Un moteur de rendu dédié permet par la suite de régénérer rapidement un rendu de la scène depuis n’importe quel point de vue. Ce qui est remarquable, c’est que ce moteur est différentiable, et qu’il peut donc intervenir dans des méthodes d’optimisation classiques (de type descente de gradient) afin d’estimer les paramètres des gaussiennes de manière à ce que les rendus des scènes resynthétisées soient les plus proches possibles des images d’entrée (résolution d’un problème inverse). Nous proposons donc de revisiter le problème de la stéréophotométrie à l’aune de ce formalisme.

État de l’art. L’un des atouts majeurs de la stéréophotométrie est qu’elle permet de reconstruire des détails fins des scènes imagées, avec par exemple des applications en numérisation 3D d’œuvres pour le patrimoine [6] ou encore la détection de défauts sur des pièces usinées [10]. L’approche historique en stéréophotométrie est celle de Woodham [12] qui définit un modèle pour estimer la normale en chaque point d’un objet à partir de plusieurs observations sous différentes illuminations. Ce modèle suppose que l’objet possède une réflectance lambertienne, son apparence ne dépendant donc que de l’angle entre normale et la direction de la lumière :

$$I_{(x,y)} \propto l^T \cdot n_{(x,y)}, \quad (1)$$

où I est l’image en niveaux de gris, $l \in \mathbb{S}^2$ la direction de la lumière et $n_{(x,y)} \in \mathbb{S}^2$ la normale de l’objet observé en chaque point (x, y) de l’image. Ce modèle peut être inversé grâce

à une optimisation par moindres carrés, avec cependant plusieurs limitations, notamment l’incapacité à traiter les ombres portées et les réflexions spéculaires, ce qui la rend en pratique peu robuste pour les matériaux et les scènes complexes. Suite à cette première approche pionnière, la majorité des méthodes qui ont suivi ont considéré les ombres portées et les réflexions spéculaires comme des données aberrantes, tout en développant des approches d’optimisation plus complexes et robustes afin de stabiliser la convergence.

Plus récemment, des approches exploitant l’apprentissage automatique profond ont vu le jour pour surmonter ces limitations. Hardy et al. [2] ont proposé une méthode d’apprentissage supervisé multi-échelle dans laquelle un réseau de neurones est entraîné sur une base de données contenant une grande diversité de formes et de matériaux afin d’estimer directement la carte des normales. Yamaguchi et al. [13] ont utilisé une estimation de carte de profondeur obtenue via *Depth-Anything* pour initialiser grossièrement la carte de normales avant de l’affiner grâce à un réseau de neurones dédié. Contrairement aux méthodes historiques, ces nouvelles approches ne cherchent pas à respecter un modèle physique explicite, mais s’adaptent automatiquement aux surfaces aux propriétés de réflexion complexes. Toutefois, leur performance dépend fortement de la qualité et de la diversité du jeu de données d’entraînement, ce qui peut limiter leur capacité de généralisation aux matériaux non vus lors de l’apprentissage. De plus, elles nécessitent des machines puissantes pour entraîner et utiliser des modèles neuronaux profonds et induisent une perte d’explicabilité.

D’autres travaux récents adoptent une approche de rendu inverse plus poussée, en intégrant des estimations de réflectance spéculaire. Par exemple, Li et al. [7] utilisent plusieurs petits réseaux de neurones spécialisés pour estimer des paramètres intermédiaires, tels que la position de la lumière et la carte de profondeur. Leur approche repose sur un modèle de rendu inverse différentiable, permettant l’utilisation de schémas d’optimisation itératifs pour ajuster les poids du réseau et minimiser l’écart entre l’image reconstruite et l’image observée. Une grande partie de l’optimisation est cependant réservée à l’estimation des ombres portées via un lancer de rayon effectué

depuis chaque pixel, ce qui augmente grandement le temps d'optimisation (en pratique plusieurs dizaines de minutes). Yang et al. [14] proposent d'utiliser la carte de profondeur pour estimer les ombres projetées et s'en servir pour mieux estimer les zones non éclairées de l'objet. Cependant, l'utilisation de réseaux pour estimer une première carte de profondeur requiert toujours des capacités de calcul importantes et des bases d'entraînement conséquentes.

La formulation d'un problème de reconstruction via un rendu inverse est intéressante, car elle rend le modèle plus flexible et adaptable à une grande variété de matériaux, à condition que celui-ci soit suffisamment généraliste. De plus, elle améliore l'explicabilité des résultats en prenant en compte un modèle physique de formation de l'image pour reconstruire l'image finale. Dans cette perspective, explorer de nouvelles approches de paramétrisation de surface est une piste prometteuse pour assurer à la fois l'explicabilité du processus de reconstruction et obtenir de bonnes performances, comme nous l'aborderons dans la suite de cet article, en nous limitant à la réflectance lambertienne afin d'établir une preuve de concept.

2 Formulation proposée

Gaussian Splatting. L'objectif est donc de pouvoir retrouver des cartes de normales d'objets avec des matériaux lambertiens en inversant le modèle de formation de l'image. Nous nous intéressons à la méthode du *Gaussian Splatting* développée par Kerbl et al. [5] pour utiliser les avantages des méthodes de rendus inverse différentiables associés. Cette technique a été initialement développée pour la reconstruction 3D multi-vues. Elle repose sur une représentation compacte et efficace de la géométrie et de l'apparence de la scène à l'aide de champs de gaussiennes placées dans un espace 3D.

Le *Gaussian Splatting* utilise un ensemble de gaussiennes localisées dans l'espace, chacune étant définie par plusieurs paramètres fondamentaux (orientation, taille, couleur). Le rendu d'une scène modélisée de cette façon s'effectue par la suite en projetant ces gaussiennes en 2D sur l'image finale, selon un angle de vue défini par l'utilisateur. Chaque gaussienne est définie par un centre $p_k \in \mathbb{R}^3$ ainsi qu'une matrice de covariance Σ de telle sorte que pour tout point $p \in \mathbb{R}^3$:

$$\mathcal{G}_k(p) = \exp\left(-\frac{1}{2}(p - p_k)^T \Sigma_k^{-1}(p - p_k)\right). \quad (2)$$

La matrice de covariance $\Sigma = RSS^T RT$ est définie par une matrice de rotation R et une matrice de mise à l'échelle S afin d'assurer qu'elle soit semi-définie positive.

Pour effectuer le *splatting* des gaussiennes en chaque pixel x de l'image, il faut tout d'abord obtenir le point 3D à évaluer sur chaque gaussienne. Pour cela, on récupère les coordonnées 3D de la projection du centre p_k vers le rayon $r_x(t) = o + td_x$ lancé depuis la caméra de position o selon la direction de visée d_x . La position sur le rayon retenue est celle qui minimise la distance orthogonale au centre p_k , notée t_{min} :

$$\hat{\mathcal{G}}(x) = \mathcal{G}(r_x(t_{min})) \quad (3)$$

Une couleur est également attribuée à chaque gaussienne et est définie à l'aide d'harmoniques sphériques, ce qui permet à la couleur perçue de varier dynamiquement en fonction de l'angle d'observation. Enfin, une opacité est associée à chaque

gaussienne afin de moduler son influence dans le processus de rendu, en contrôlant son importance relative dans le mélange des contributions via un processus d'*alpha blending* en prenant les k gaussiennes d'avant en arrière jusqu'à atteindre une saturation de l'opacité à $\alpha = 1$. On obtient la couleur $c(x)$ du pixel x par l'équation suivante :

$$c(x) = \sum_{k=1}^K c_k \alpha_k \hat{\mathcal{G}}_k(x) \prod_{j=1}^{k-1} (1 - \alpha_j \hat{\mathcal{G}}_j(x)) \quad (4)$$

où α_k est le paramètre d'opacité et c_k la couleur selon le point de vue concerné. Un des principaux avantages du *Gaussian Splatting* est que l'ensemble du processus est différentiable presque partout, ce qui permet d'optimiser les paramètres des modèles gaussiens par descente de gradient. L'optimisation repose sur une fonction de perte L_1 mesurant la différence entre l'image rendue et une image de référence (vérité terrain). Cette propriété rend donc la méthode particulièrement adaptée aux approches fondées sur l'apprentissage automatique et l'optimisation des paramètres de la scène.

En plus du *splatting* réalisé pour les couleurs, on peut reproduire ces étapes pour obtenir une carte de profondeur. À partir de cette carte de profondeur, il est naturellement possible d'estimer une carte de normales en calculant son gradient. Cependant, celle-ci ne sera pas forcément liée à la géométrie de l'objet car la nature volumétrique des gaussiennes 3D est en conflit avec la représentation de surfaces.

Pour résoudre ce problème, Huang et al. [3] proposent une extension du *Gaussian Splatting* en contraignant la gaussienne sur un espace tangent paramétré tel que le point P sur le plan de la gaussienne 2D de coordonnées (u, v) a pour équation :

$$P(u, v) = p_k + s_u t_u u + s_v t_v v = H(u, v, 1, 1)^T \quad (5)$$

$$\text{avec } H = \begin{bmatrix} RS & p_k \\ 0 & 1 \end{bmatrix}, \quad (6)$$

où $p_k \in \mathbb{R}^3$ est toujours le centre de la gaussienne dans l'espace 3D, $R = [t_u, t_v, t_w]$ est la matrice de rotation définie par les deux vecteurs tangents t_u et t_v ainsi que le vecteur orthogonal $t_w = t_u \times t_v$ au plan tangent et $S = (s_u, s_v)$ est la matrice de mise à l'échelle où la dernière composante est 0 pour former une gaussienne 2D. Le point pris en compte est celui sur le plan de la gaussienne. Grâce à cette contrainte, il devient possible d'obtenir une carte de normales en réalisant un *splatting* des normales de la même manière que pour la couleur. Nous avons donc choisi de reprendre l'idée générale de ces travaux afin de l'adapter pour résoudre le problème de la stéréophotométrie, en modélisant et exploitant la carte de normales pour obtenir un rendu final.

En temps normal, l'initialisation du *Gaussian Splatting* se fait par *Structure From Motion* ce qui permet d'obtenir un nuage de points épars grâce à plusieurs points de vue d'une même scène. Cependant, pour la stéréophotométrie un seul point de vue est disponible. Pour pallier à ce problème, nous plaçons donc les points sur un plan de coordonnées $z = 0$ afin que le rendu soit similaire à celui voulu pour faciliter le début de l'optimisation. Une gaussienne est placée pour chaque pixel et la couleur associée à chaque gaussienne correspond au maximum de la couleur observée sur l'ensemble des images I_i avec différents éclairages :

$$c(x, y) = \max\{I_1(x, y), I_2(x, y), \dots, I_n(x, y)\}. \quad (7)$$

Il est désormais possible de générer un rendu de la carte des normales ainsi que de l'albédo de l'objet. Dans le modèle de *Gaussian Splatting*, les normales ne sont pas directement utilisées pour le rendu car la lumière n'est pas prise en compte, ce qui implique qu'un rendu depuis un point de vue fixe sera toujours identique. Pour prendre en compte les différences de rendu en fonction de la position de la lumière, nous utilisons donc le modèle de l'équation (1) en analysant si les surfaces sont alignées ou non avec la direction de la lumière pour calculer les ombres propres.

Optimisation. Une fois que le modèle a été choisi, il faut définir la méthode d'optimisation pour pouvoir modifier les paramètres des gaussiennes (position, matrice de covariance, couleur et opacité). Comme mentionné précédemment, le moteur de rendu utilise exclusivement des gaussiennes afin de représenter la scène, ce qui permet de calculer efficacement les gradients pour les optimiser par une descente de gradient. La fonction de perte $\mathcal{L}$ que nous avons utilisée se décompose en deux termes. Tout d'abord, un terme de perte photométrique $\mathcal{L}_c$ qui va mesurer l'écart entre le rendu obtenu et le rendu réel par une fonction de perte L_1 . Ensuite, une régularisation de la normale $\mathcal{L}_n$ est ajoutée afin de s'assurer que les gaussiennes sont alignées avec la vraie surface de l'objet. Pour cela, on calcule une carte de normales avec le gradient de la carte de profondeur et on la compare avec la carte de normales obtenue via *splatting*. On obtient la fonction de perte suivante :

$$\mathcal{L} = \mathcal{L}_c + \lambda \mathcal{L}_n. \quad (8)$$

Le nombre de gaussiennes va aussi évoluer lors de l'optimisation. En particulier, les gaussiennes trop grandes ou dont l'opacité est trop faible vont être supprimées. Elles pourront être aussi clonées ou découpées lorsque plus de détails sont nécessaires. Les détails de ce procédé sont décrits dans [5]. L'ensemble de cette méthode va donc nous permettre d'utiliser le *Gaussian Splatting* pour la stéréophométrie calibrée.

3 Résultats

Notre procédé d'estimation de champs de normales par *Gaussian Splatting* est testé sur différents objets. Dans un premier temps, nous effectuons des tests sur des images synthétiques comprenant des formes géométriques simples, dont le rendu est effectué à l'aide de *Blender* [1], un logiciel libre de modélisation et de rendu 3D. Plusieurs images synthétiques sont ainsi générées, avec des positions de lumière différentes. Sur ces cas simples, notre algorithme retrouve des cartes de normales 3D cohérentes (Figure 1).

Dans un second temps, nous nous comparons aux résultats des méthodes de l'état de l'art, grâce aux images du jeu de données *DiliGenT*, crée par Shi et al. [9], en particulier en estimant travaillant sur les objets *Bear*, *Cat* et *Pot1*, qui sont des objets de la base avec des matériaux lambertiens (Figure 2). Remarquons que les zones où l'erreur angulaire est la plus élevée sont principalement les régions où une ombre portée est présente sur un grand nombre d'images d'entrée. Ceci s'explique par le fait que notre modèle ne prend pas encore en compte ce type d'ombre, qui sont les plus complexes à estimer

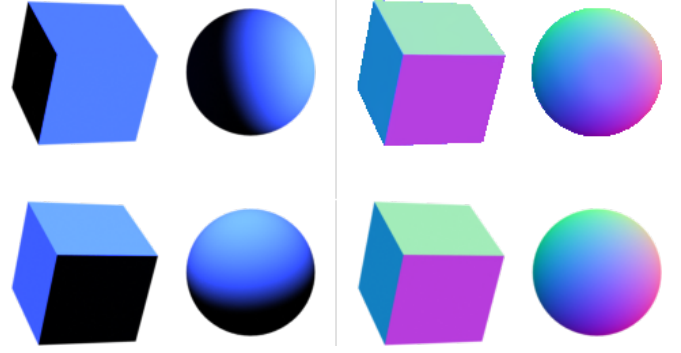


FIGURE 1 : Exemples de rendus via Blender (à gauche haut/bas avec différentes positions de lumière). Normales estimées (après optimisation en haut à droite, et vérité terrain en bas à droite).

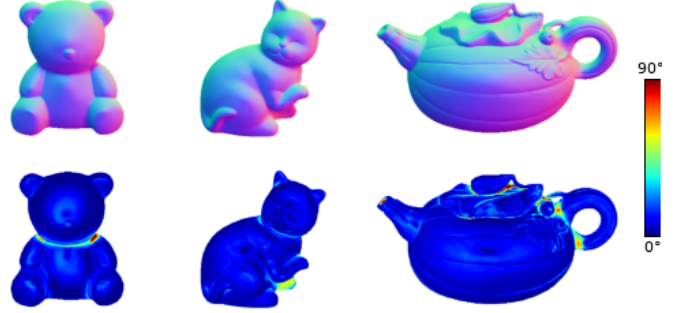


FIGURE 2 : Normales estimées et carte d'erreur angulaire sur les objets *Bear*, *Cat*, et *Pot1* de la base *DiliGenT*. Les orientations des normales sont correctement estimées. Les zones d'erreurs d'estimation sont concentrées sur les zones d'ombres portées.

à cause de leur nature globale. Nous illustrons ce phénomène en figure 3, où la patte arrière gauche du chat est cachée par l'ombre portée de la patte avant droite.

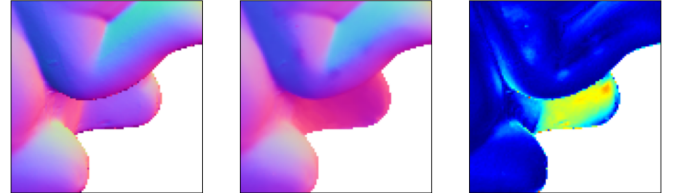


FIGURE 3 : Vérité terrain des normales à gauche. Normales estimées au centre et carte d'erreur angulaire à droite. On voit que la partie cachée par des ombres portées a été mal estimée.

La métrique de qualité classique pour évaluer les modèles est la moyenne des erreurs angulaires (en degrés). Des comparaisons sur les trois objets sont présentées en Table 1, en les comparant à la méthode historique ainsi qu'à des méthodes plus récentes. Qualitativement, nos premiers résultats sont proches de ceux obtenus par la méthode de Woodham [12], ce qui est attendu, car nous considérons un modèle de formation d'image identique. Néanmoins, revisiter ce problème d'estimation des champs de normales à l'aune des *Gaussian Splatting* ouvre des perspectives prometteuses, notamment une possible gestion des matériaux spéculaires ainsi qu'une détection et une prise en compte possible des ombres portées dans les images d'entrées (voir section suivante).

Il faut cependant noter que d'autres paramètres peuvent impacter l'intérêt d'une méthode, par exemple l'explicabilité du

	Cat	Bear	Pot1
Baseline [12]	8.41	8.39	8.89
DANINet [7]	4.73	4.11	6.41
Uni MS-PS [2]	3.45	3.14	4.12
GS-PS (Ours)	8.26	8.13	10.11

TABLE 1 : Performances sur le jeu de test DiliGenT, mesuré via l’erreur angulaire (en degrés).

modèle ou les ressources nécessaires pour le faire fonctionner. En particulier, notre méthode est plus explicable que les méthodes basées sur des réseaux de neurones profonds car elle optimise un modèle physique explicite de formation de l’image pour obtenir la carte de normales. Elle est également plus explicable que les méthodes de rendus inverse utilisant des petits réseaux car les paramètres que nous optimisons ont une interprétation géométrique, contrairement aux poids des réseaux de neurones.

4 Conclusions et travaux futurs

Nos premiers résultats d’utilisation du *Gaussian Splatting* pour résoudre le problème de l’estimation de champs de normales en stéréophotométrie sont très encourageants. Dans le futur, nous chercherons d’abord à affiner notre modèle afin de pouvoir gérer des matériaux non-lambertiens. Ceci est envisageable car il existe des moteurs de rendu spécifiques aux *Gaussian Splatting* permettant de générer des réflexions spéculaires [4, 8]. De même, notre méthode n’estime actuellement que les ombres propres sur les surfaces des objets. Or, il existe également des ombres portées qui apparaissent lorsqu’un point de la surface se trouve occulté par une autre zone de l’objet se trouvant sur la trajectoire de la lumière. Pour calculer la localisation de ces ombres portées, une technique classique est d’utiliser la *Shadow Mapping*, introduit par Williams [11].

Le *Shadow Mapping* consiste à effectuer un rendu de profondeur depuis le point de vue de la lumière, puis à comparer ce rendu avec les coordonnées des points 3D obtenus via la carte de profondeur générée depuis le point de vue de la caméra. Ce principe est illustré en Figure 4 sur un objet synthétique simple (en utilisant le moteur de rendu du *Gaussian Splatting*).

Références

- [1] BLENDER-FOUNDATION : *Blender - a 3D modelling and rendering package*.
- [2] Clément HARDY, Yvain QUÉAU et David TSCHUMPERLÉ : Uni MS-PS : A multi-scale encoder-decoder transformer for universal photometric stereo. *CVIU*, 248:104093, 2024.
- [3] Binbin HUANG, Zehao YU, Anpei CHEN, Andreas GEIGER et Shenghua GAO : 2D Gaussian Splatting for geometrically accurate radiance fields. *In SIGGRAPH*, 2024.
- [4] Yingwenqi JIANG, Jiadong TU, Yuan LIU, Xifeng GAO, Xiaoxiao LONG, Wenping WANG et Yuxin MA : Gaussianshader : 3d gaussian splatting with shading functions for reflective surfaces. *In CVPR*, 2024.
- [5] Bernhard KERBL, Georgios KOPANAS, Thomas LEIMKÜHLER et George DRETTAKIS : 3D Gaussian Splatting for real-time radiance field rendering. *ToG*, 42(4), 2023.

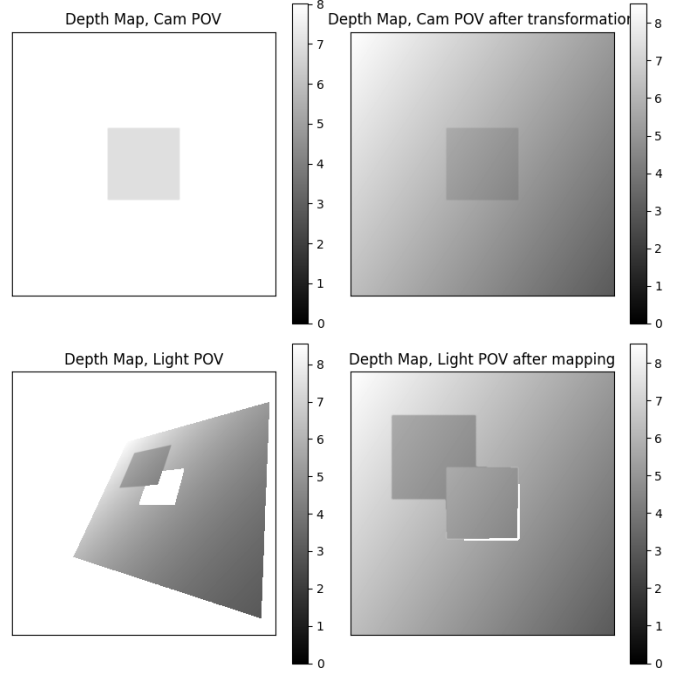


FIGURE 4 : Étapes du *Shadow Mapping* pour détecter les ombres portées. En haut à gauche, la carte de profondeur vue depuis la caméra. En haut à droite, la même carte après application d’une transformation rigide vers l’emplacement de la source lumière. En bas à gauche, la carte de profondeur du point de vue de la lumière. En bas à droite, la même carte après transformation vers l’espace de la caméra. Les deux images de droite sont comparées pour détecter les points correspondant aux ombres portées.

- [6] Antoine LAURENT, Benjamin COUPRY, Baptiste BRUMENT, Jean MÉLOU, Yvain QUÉAU, Carole FRITZ et Jean-Denis DUROU : Combining geometric and photometric 3d reconstruction techniques for cultural heritage. *JCH*, 73, 2025.
- [7] Zongrui LI, Qian ZHENG, Boxin SHI, Gang PAN et Xudong JIANG : DANI-Net : Uncalibrated photometric stereo by differentiable shadow handling, anisotropic reflectance modeling, and neural inverse rendering. *In CVPR*, 2023.
- [8] Zhihao LIANG, Qi ZHANG, Ying FENG, Ying SHAN et Kui JIA : GS-IR : 3d gaussian splatting for inverse rendering. *In CVPR*, 2024.
- [9] Boxin SHI, Zhe WU, Zhipeng MO, Dinglong DUAN, Sai-Kit YEUNG et Ping TAN : A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. *In CVPR*, 2016.
- [10] Shun WANG, Ke XU, Baohua LI et Xiangyu CAO : Online micro defects detection for ductile cast iron pipes based on twin light photometric stereo. *Case Studies in Construction Materials*, 19, 2023.
- [11] Lance WILLIAMS : Casting curved shadows on curved surfaces. *SIGGRAPH Comput. Graph.*, 12(3):270–274, 1978.
- [12] Robert J. WOODHAM : Photometric Method For Determining Surface Orientation From Multiple Images. *OptEng*, 19(1): 191139, 1980.
- [13] Masahiro YAMAGUCHI, Takashi SHIBATA, Shoji YACHIDA, Keiko YOKOYAMA et Toshinori HOSOI : MDCN-PS : Monocular-depth-guided coarse normal attention for robust photometric stereo. *In WACV*, 2025.
- [14] Yuze YANG, Jiahang LIU, Yangyu FU et Yue NI : Non-lambertian photometric stereo via neural inverse rendering and interreflection model. *OLEN*, 178:108138, 2024.