



HAL
open science

On Photometric Stereo in the Presence of a Refractive Interface

Yvain Quéau, Robin Bruneau, Jean Mélou, Jean-Denis Durou, François Lauze

► **To cite this version:**

Yvain Quéau, Robin Bruneau, Jean Mélou, Jean-Denis Durou, François Lauze. On Photometric Stereo in the Presence of a Refractive Interface. 9th International Conference on Scale Space and Variational Methods in Computer Vision (SSVM 2023), May 2023, Santa Margherita di Pula, Sardinia, Italy. pp.691-703, 10.1007/978-3-031-31975-4_53 . hal-04038233

HAL Id: hal-04038233

<https://hal.science/hal-04038233v1>

Submitted on 20 Mar 2023

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.

On Photometric Stereo in the Presence of a Refractive Interface

Yvain Quéau¹✉, Robin Bruneau^{2,3}, Jean Mélou²,
Jean-Denis Durou², and François Lauze³

¹ Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France

² DIKU, Copenhagen University, Denmark

³ IRIT, UMR CNRS 5505, Toulouse, France

✉yvain.queau@ensicaen.fr

Abstract. We conduct a discussion on the problem of 3D-reconstruction by calibrated photometric stereo, when the surface of interest is embedded in a refractive medium. We explore the changes refraction induces on the problem geometry (surface and normal parameterization), and we put forward a complete image formation model accounting for refracted lighting directions, change of light density and Fresnel coefficients. We further show that as long as the camera is orthographic, lighting is directional and the interface is planar, it is easy to adapt classic methods to take into account the geometric and photometric changes induced by refraction. Moreover, we show on both simulated and real-world experiments that incorporating these modifications of PS methods drastically improves the accuracy of the 3D-reconstruction.

1 Introduction

Photometric stereo (PS) is a 3D computer vision technique which was pioneered by Woodham in the late 70s [27]. It aims at inferring the shape of an opaque surface from a series of images captured under the same viewing angle, but varying illumination. Compared to other 3D-reconstruction techniques, PS excels at recovering the thinnest geometric variations (high-frequency information given by surface normals), and it is the only photographic 3D-reconstruction method which is also able to infer the reflectance of the surface. Such properties are essential in applications such as relighting or cultural heritage artifacts digitization.

However, a fundamental assumption in PS is that the light sources, the camera and the pictured surface all lie in the same homogeneous medium – usually the air. In the present paper, we revisit PS in the presence of a refractive interface i.e., when the camera and the light sources both lie in one homogeneous medium, while the surface is immersed in another homogeneous medium with a different index of refraction (pure water, glass, alcohol, etc.). This particular setting finds applications, for instance, in underwater imaging (Fig. 1a) or in the digitization of natural historic museal objects preserved in amber or alcohol.

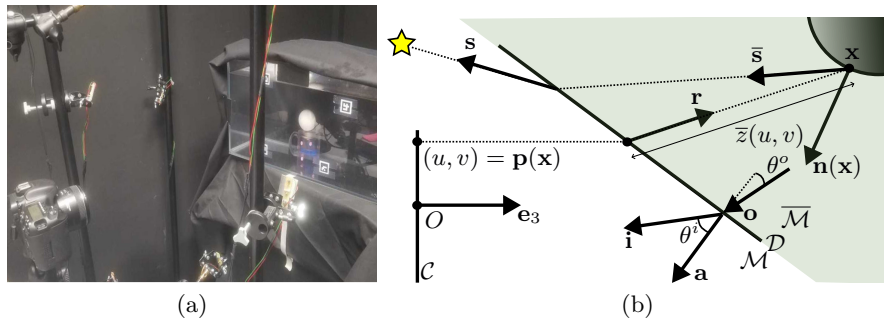


Fig. 1. We discuss the problem of recovering through photometric stereo the 3D-shape and the albedo of a surface immersed in a refractive medium, as in (a) where a white sphere is immersed in an aquarium filled with pure water. In particular, we show how to adapt classic PS methods when the lighting is directional, the camera is orthographic and the interface is planar, as illustrated in the sketch (b) which summarizes our notations. Therein, given a plane \mathcal{D} with normal \mathbf{a} , Snell’s law (2.3) gives the relation between an incident ray \mathbf{i} in medium \mathcal{M} , and the refracted one \mathbf{o} in medium $\overline{\mathcal{M}}$. Even if the camera is orthographic, a point \mathbf{x} on the surface projects non-orthogonally onto the image plane \mathcal{C} : a pixel (u, v) first deprojects onto \mathcal{D} along the viewing direction \mathbf{e}_3 , and then travels the distance $\bar{z}(u, v)$ along the refracted viewing direction \mathbf{r} . Besides, the effective lighting direction $\bar{\mathbf{s}}$ differs from the direction \mathbf{s} which is calibrated outside the refractive medium.

The difference with classic PS lies in the presence of an interface between the two media, which have different indices of refraction. Refraction will have profound consequences in 3D shape recovery techniques, as it modifies the geometry of image acquisition, and light direction and density will be changed as well (this is, after all, the principle behind a lot of lensing effects). While these points are well-understood by designers of optical systems, either to use them or for limiting some of their undesirable consequences, they have seldom been investigated from the photometric shape recovery side.

Assumptions and contributions We address the PS problem, in the presence of a *Lambertian* surface (specularities are viewed as outliers) embedded in a *homogeneous* refractive medium with *known geometry*, imaged in the *visible spectrum*. After reviewing related works in Sect. 2, we explore the impact of a *planar* (but not necessarily fronto-parallel) refractive interface on the geometry of PS under *orthographic projection* in Sect. 3. In Sect. 4, we derive a complete image formation model for this case, under *directional lighting calibrated outside the refractive medium*. This model accounts for refraction of lighting directions, attenuation of lighting densities, and Fresnel coefficients. Then, we discuss in Sect. 5 the inversion of this model by adapting classic PS algorithms. Eventually, in Sect. 6 we draw our conclusions, and mention possible extensions of our work to more complicated setups (pinhole camera, non-directional lighting, non-planar interface, and light absorption).

2 Background

Photometric stereo In the traditional PS setup, the pictured surface \mathcal{S} is assumed Lambertian i.e., it reflects light diffusively, as the reflectance at $\mathbf{x} \in \mathcal{S}$ is characterized by the albedo $\rho(\mathbf{x}) \in [0, 1]$. Let us consider a surface lit by a single, known point light source at infinity (calibrated directional lighting) represented by unit direction $\mathbf{s} \in \mathbb{R}^3$ and density $\varphi > 0$, and denote $\mathbf{n}(\mathbf{x}) \in \mathbb{R}^3$ the unit outward normal to the surface at \mathbf{x} . Then, the measured brightness at pixel $(u, v) = \mathbf{p}(\mathbf{x})$, which is the projection of the surface point \mathbf{x} onto the camera image plane, is $I(\mathbf{p}) \propto \varphi \max\{0, \rho(\mathbf{x}) \mathbf{n}(\mathbf{x})^\top \mathbf{s}\}$, with the proportionality constant independent of \mathbf{x} . Omitting the $\max\{\}$ operator, which models self-shadows (they are usually dealt with robust estimators), integrating the proportionality coefficient into the albedo (which can be normalized a posteriori), and considering $k \geq 3$ light sources yields the following image formation model:

$$I_i(\mathbf{p}) = \varphi_i \rho(\mathbf{x}) \mathbf{n}(\mathbf{x})^\top \mathbf{s}_i, \quad i \in \{1, \dots, k\}. \quad (2.1)$$

This model can be inverted as long as the light directions \mathbf{s}_i are non-coplanar, so as to compute the Lambertian reflectance $\rho(\mathbf{x})$ and the surface normal $\mathbf{n}(\mathbf{x})$ for each \mathbf{p} . This approach can also be extended to non-Lambertian reflectance and uncalibrated lighting, for instance by resorting to deep neural networks [7].

Surface parameterization The surface \mathcal{S} is parameterized as $(u, v) \mapsto \mathbf{x}(u, v) = \mathcal{S}(u, v)$ and its normal at \mathbf{x} is written as

$$\mathbf{n}(\mathbf{x}) = \pm \frac{\mathcal{S}_u \times \mathcal{S}_v}{|\mathcal{S}_u \times \mathcal{S}_v|}, \quad (2.2)$$

where \mathcal{S}_u and \mathcal{S}_v are the partial derivatives of \mathcal{S} , and where the \pm ambiguity is resolved by taking arbitrarily the normal oriented towards the camera. Once the normal field $\mathbf{n}(\mathbf{x})$ is estimated, retrieving \mathcal{S} then comes down to a 2D integration problem, for which various solutions exist [20]. The parameterization \mathcal{S} is a right-inverse to the projection: $\mathbf{p}(\mathcal{S}(u, v)) = (u, v)$. It is constrained by the form that \mathbf{p} takes (orthographic projection, perspective projection, etc.), and this has of course important consequences on the integration process.

Refraction The index of refraction (IoR) n of a material is the ratio c/v of the speed of light in vacuum and the velocity in the medium. Snell's laws assert that 1) the normal \mathbf{a} to the interface, the incident light direction \mathbf{i} and the refracted light direction \mathbf{o} are coplanar; and 2) the refracted and incident angles satisfy the relation $n \sin \theta^i = \bar{n} \sin \theta^o$, with n the IoR of the first medium, \bar{n} the IoR of the second one, θ^i the angle between \mathbf{i} and \mathbf{a} , and θ^o the angle between \mathbf{a} and \mathbf{o} (see Fig. 1b). In vectorial form [14], defining $\mu = n/\bar{n}$:

$$\mathbf{o} = \text{Snell}_\mu^{\mathbf{a}}(\mathbf{i}) = \mu \mathbf{i} + \left(\sqrt{1 - \mu^2 (1 - (\mathbf{i}^\top \mathbf{a})^2)} - \mu (\mathbf{i}^\top \mathbf{a}) \right) \mathbf{a}. \quad (2.3)$$

Refractive 3D-vision Snell’s law (2.3) of refraction has been considered in few 3D-vision contexts. For instance, the epipolar geometry theory has been extended to the case where the camera and the surface are separated by a refractive plane [5]. This constitutes the basis for the development of refractive structure-from-motion algorithms [4,13]. Multi-view stereo in the presence of a refractive interface has also been recently explored [3,6,11]. In the photometric stereo context, underwater imaging has attracted some attention [25,17,26,10,16]. These works focus mostly on light absorption, which occurs when scattering is involved (inhomogeneous medium such as murky water) or in near-infrared imaging. Yet, other refraction effects (e.g., change of incident light direction and density, and of surface parameterization) are neglected. For instance, it is usually assumed that all the sources have the same relative intensity, and that their directions can be obtained using a calibration target immersed in the medium. Yet, even if all the sources outside the refractive medium have exactly the same intensity, the refractive interface will induce luminous fluxes with different densities (see Sect. 4). Therefore, it would be more convenient to calibrate light directions and densities outside the refractive medium, and account for refraction within the image formation model. This has been achieved in [18] but only for a fronto-parallel interface and with a somehow naive numerical solution, and in [9] but by relying on laser triangulation. Instead, the present paper aims at modeling and evaluating the effects of a refractive interface with arbitrary orientation on shape recovery by pure PS, and at providing an efficient numerical solution by adapting state-of-the-art algorithms.

3 Geometry of Refractive PS

Notations As illustrated in Fig. 1b, we work in \mathbb{R}^3 with its canonical frame $(O, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, where O is the camera’s principal point, \mathbf{e}_3 is the optical axis direction and $\mathcal{C} := \mathbf{e}_3^\perp$ is the image plane. A generic point in \mathbb{R}^3 is denoted by \mathbf{x} , while $\mathbf{p}(\mathbf{x}) = (u, v)^\top$ denote the 2D coordinates of its conjugate pixel in the frame $(O, \mathbf{e}_1, \mathbf{e}_2)$. The projection from $\mathbb{R}^3 \rightarrow \mathbb{R}^2$ keeping the first two coordinates is represented by the matrix $\mathbf{\Pi} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$, whose transpose is the canonical injection $\mathbb{R}^2 \rightarrow \mathbb{R}^3$. The interface plane \mathcal{D} is given by the equation $\mathbf{a}^\top \mathbf{x} + \alpha = 0$, $\alpha \in \mathbb{R}$, where $\mathbf{a} = (a_1, a_2, a_3)^\top \in \mathbb{S}^2$ is a known unit normal vector to \mathcal{D} (\mathbb{S}^2 being the unit sphere of \mathbb{R}^3), oriented towards the camera ($\mathbf{a}^\top \mathbf{e}_3 \leq 0$). We assume that $\mathbf{a}^\top \mathbf{e}_3 \neq 0$. The medium containing the camera is located in $\mathcal{M} = \{\mathbf{x} \in \mathbb{R}^3, \mathbf{a}^\top \mathbf{x} + \alpha > 0\}$ and has IoR n , while the medium containing the object under scrutiny is located in $\overline{\mathcal{M}} = \{\mathbf{x} \in \mathbb{R}^3, \mathbf{a}^\top \mathbf{x} + \alpha \leq 0\}$ and has IoR \bar{n} , and we denote $\mu = n/\bar{n} < 1$. Lastly, for a plane \mathcal{P} of equation $\mathbf{v}^\top \mathbf{x} + \beta = 0$ and $\mathbf{w} \in \mathbb{R}^3$ with $\mathbf{v}^\top \mathbf{w} \neq 0$, we define the projection on plane \mathcal{P} along direction \mathbf{w} as

$$\mathbf{P}_{\mathcal{P}}^{\mathbf{w}}(\mathbf{x}) = \mathbf{x} - \frac{\mathbf{v}^\top \mathbf{x} + \beta}{\mathbf{v}^\top \mathbf{w}} \mathbf{w} = \left(\text{id} - \frac{\mathbf{w} \mathbf{v}^\top}{\mathbf{w}^\top \mathbf{v}} \right) \mathbf{x} - \frac{\beta \mathbf{w}}{\mathbf{v}^\top \mathbf{w}}. \quad (3.1)$$

The orthogonal projection on $\mathbf{v}^\top \mathbf{x} = 0$ is simply denoted by $\mathbf{P}_{\mathbf{v}^\perp}$.

Depth from the Interface In the orthographic case, all the light rays reaching the camera are orthogonal to the image plane, i.e., parallel to \mathbf{e}_3 . In the absence of a refractive interface (or when the interface is fronto-parallel as in [16,18]), the projection is simply $\mathbf{p}(\mathbf{x}) = \mathbf{\Pi}\mathbf{x}$ and the surface parameterization, as its right inverse, is $\mathcal{S}(u, v) = (u, v, z(u, v))^\top$ with z the depth map. However, when a non fronto-parallel refractive interface comes into play, the projection becomes non-orthogonal (see Fig. 1b). In this case, the rays reaching the camera are parallel to direction \mathbf{e}_3 , and come from parallel incident rays with common direction $\mathbf{r} \in \mathbb{S}^2$ within the refractive medium $\overline{\mathcal{M}}$ as the interface is planar. Vector \mathbf{r} is fully determined by Snell's law (2.3) as it must refract to viewing direction \mathbf{e}_3 :

$$\mathbf{r} = \text{Snell}_{\mu}^{-\mathbf{a}}(\mathbf{e}_3), \quad (3.2)$$

where the sign before \mathbf{a} comes from the fact that \mathbf{a} is oriented towards the camera, while \mathbf{e}_3 and \mathbf{r} are oriented towards the surface (see Fig. 1b).

Therefore, a point $\mathbf{x} \in \overline{\mathcal{M}}$ on the immersed surface first projects non-orthogonally onto \mathcal{D} along the refracted viewing direction \mathbf{r} , before being orthogonally projected onto the camera image plane along the viewing direction \mathbf{e}_3 :

$$\mathbf{p}(\mathbf{x}) = \mathbf{\Pi}P_{\mathbf{e}_3^\perp}(\mathbf{P}_{\mathcal{D}}^{\mathbf{r}}(\mathbf{x})). \quad (3.3)$$

This leads to a straightforward model where we deproject pixel $(u, v)^\top$ in the image plane to a point on the refractive interface \mathcal{D} , and then follow the incident ray up to the object: $\mathcal{S}(u, v) = \mathbf{P}_{\mathcal{D}}^{\mathbf{e}_3}(u, v, 0)^\top + \bar{z}(u, v)\mathbf{r}$, with \bar{z} the pseudo-depth (distance travelled along the refracted ray \mathbf{r}). One readily checks that $\mathbf{p}(\mathcal{S}(u, v)) = (u, v)^\top$. Using (3.1), we can write

$$\mathcal{S}(u, v) = \mathbf{P}_{\mathcal{D}}^{\mathbf{e}_3}(u, v, 0)^\top + \bar{z}(u, v)\mathbf{r} = \mathbf{A}(u, v, 0)^\top + \mathbf{t} + \bar{z}(u, v)\mathbf{r}, \quad (3.4)$$

with known quantities

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{a_1}{a_3} & -\frac{a_2}{a_3} & 0 \end{pmatrix}, \quad \mathbf{t} = -\frac{\alpha}{a_3}\mathbf{e}_3. \quad (3.5)$$

Surface Normals Now, let us establish the link between the pseudo-depth \bar{z} from the interface, and the normal \mathbf{n} to the surface. To do this, let us consider the partial derivatives of the parameterization. They are given by $\mathcal{S}_u = \mathbf{A}\mathbf{e}_1 + \bar{z}_u\mathbf{r}$ and $\mathcal{S}_v = \mathbf{A}\mathbf{e}_2 + \bar{z}_v\mathbf{r}$. An (unnormalized) normal to the surface $\mathcal{S}(u, v)$ is $\mathcal{S}_u \times \mathcal{S}_v = (\mathbf{A}\mathbf{e}_1 + \bar{z}_u\mathbf{r}) \times (\mathbf{A}\mathbf{e}_2 + \bar{z}_v\mathbf{r})$. Set $\mathbf{b}^1 = \mathbf{r} \times \mathbf{A}\mathbf{e}_2$, $\mathbf{b}^2 = \mathbf{A}\mathbf{e}_1 \times \mathbf{r}$ and $\mathbf{b}^3 = \mathbf{A}\mathbf{e}_2 \times \mathbf{A}\mathbf{e}_1$. Then $\mathcal{S}_u \times \mathcal{S}_v = \bar{z}_u\mathbf{b}^1 + \bar{z}_v\mathbf{b}^2 - \mathbf{b}^3$. By letting \mathbf{B} be the matrix $-(\mathbf{b}^1, \mathbf{b}^2, \mathbf{b}^3)$,

$$\mathbf{n}(u, v) = \mathbf{n}(\mathcal{S}(u, v)) \propto \mathbf{B} \begin{pmatrix} \nabla \bar{z}(u, v) \\ -1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \frac{a_2 r_2}{a_3} + r_3 & -\frac{a_1 r_2}{a_3} & \frac{a_1}{a_3} \\ -\frac{a_2 r_1}{a_3} & \frac{a_1 r_1}{a_3} + r_3 & \frac{a_2}{a_3} \\ -r_1 & -r_2 & 1 \end{pmatrix}. \quad (3.6)$$

Eq. (3.6) relates the surface normals to the underlying gradient of the pseudo-depth from the interface. When the interface is fronto-parallel, $\mathbf{a} = -\mathbf{e}_3$, $\mathbf{r} = \mathbf{e}_3$ and $\bar{z} = z - \beta$. Hence, $\mathbf{B} = \mathbf{I}_3$ and the formula matches the classic one obtained in the absence of refraction: $\mathbf{n}(u, v) \propto (\nabla z(u, v)^\top, -1)^\top$.

4 Image Formation Model under Directional Lighting

Now, we turn our attention to extending the image formation model (2.1) to the refractive case. We assume to have at hand a series of k images I_1, \dots, I_k , with lighting directions $\mathbf{s}_1, \dots, \mathbf{s}_k$ and densities $\varphi_1, \dots, \varphi_k$ calibrated inside \mathcal{M} . The effective lighting inside $\overline{\mathcal{M}}$ will however be different from the calibrated one, due to refraction.

Effective Lighting Directions and Densities We assume that light directions and densities are known in the camera medium \mathcal{M} thanks to calibration, and we denote these calibrated parameters by \mathbf{s}_i and φ_i . However, the directions $\overline{\mathbf{s}}_i$ and densities $\overline{\varphi}_i$ of the effective light beams reaching the surface differ from calibrated values, see Fig. 2.

After crossing the refractive interface, the incident light beams obviously remain parallel, yet their directions become, according to Snell's law (2.3):

$$\overline{\mathbf{s}}_i = -\text{Snell}_{\mu}^{-\mathbf{a}}(-\mathbf{s}_i), \quad i \in \{1, \dots, k\}. \quad (4.1)$$

Moreover, the size of the surface elements orthogonal to the rays also changes, according to $\frac{d\Sigma_i}{\mathbf{a}^\top \overline{\mathbf{s}}_i} = d\Sigma_{\mathcal{D}} = \frac{d\Sigma_i}{\mathbf{a}^\top \mathbf{s}_i}$. Then:

$$\overline{\varphi}_i = \frac{\mathbf{a}^\top \mathbf{s}_i}{\mathbf{a}^\top \overline{\mathbf{s}}_i} \varphi_i, \quad i \in \{1, \dots, k\}. \quad (4.2)$$

Let us emphasize that, even if all the sources have exactly the same intensity i.e., $\varphi_i = \varphi_j, \forall i \neq j$, the effective densities will be different. For instance, when $n = 1$, $\overline{n} = 1.5$, and $\varphi_1 = \varphi_2 = 1$, a lighting orthogonal to the interface yields $\overline{\varphi}_1 = 1$, while an incident angle of 30° yields $\overline{\varphi}_2 = 0.91$. This effect is thus far from negligible in a calibrated PS setup.

Fresnel Coefficients The interface may act partially as a mirror, with the amount of transmitted light being a function of the incident angle. This happens twice in the process: first when going from the light source in \mathcal{M} to the surface embedded in $\overline{\mathcal{M}}$, and then when going from the latter to the camera, back in \mathcal{M} .

The incident and outgoing angles when going from \mathcal{M} towards $\overline{\mathcal{M}}$ will vary depending on the incident direction \mathbf{s}_i , $i \in \{1, \dots, k\}$: each light source will thus induce a different transmission rate. This rate is however the same whatever the point \mathbf{x} , since lighting is assumed directional - this would not be the case for

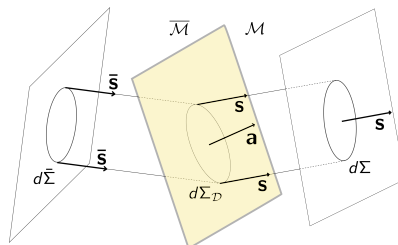


Fig. 2. Light refraction by a planar interface with normal \mathbf{a} , with \mathbf{s} the light direction calibrated outside the refractive medium, and $\overline{\mathbf{s}}$ the effective refracted direction (in this drawing, the light source is on the right). Light direction is changed according to Snell's law, while its density is multiplied by $\frac{\mathbf{a}^\top \mathbf{s}}{\mathbf{a}^\top \overline{\mathbf{s}}}$.

instance under a near point light source. Assuming that all the light beams are unpolarized, each transmission rate is given by the Fresnel coefficient

$$T_i^{\mathcal{M} \rightarrow \overline{\mathcal{M}}} = 1 - \frac{1}{2} \left(\frac{(\mu \mathbf{a}^\top \mathbf{s}_i - \mathbf{a}^\top \overline{\mathbf{s}}_i)^2}{(\mu \mathbf{a}^\top \mathbf{s}_i + \mathbf{a}^\top \overline{\mathbf{s}}_i)^2} + \frac{(\mu \mathbf{a}^\top \overline{\mathbf{s}}_i - \mathbf{a}^\top \mathbf{s}_i)^2}{(\mu \mathbf{a}^\top \overline{\mathbf{s}}_i + \mathbf{a}^\top \mathbf{s}_i)^2} \right), \quad i \in \{1, \dots, k\}. \quad (4.3)$$

Taking again as an example the case $n = 1$, $\bar{n} = 1.5$, an incident lighting orthogonal to the interface yields $T_1^{\mathcal{M} \rightarrow \overline{\mathcal{M}}} = 0.9600$, while an incident angle of 30° yields $T_2^{\mathcal{M} \rightarrow \overline{\mathcal{M}}} = 0.9585$. This shows that this Fresnel coefficient is non-negligible, although less dramatic than the change in the incident densities.

When going from $\overline{\mathcal{M}}$ to \mathcal{M} , the incident and outgoing angles are the same for all images I_1, \dots, I_k (viewing direction is independent from the incident lighting directions), therefore the transmission rate simply scales all the brightness values at pixel \mathbf{p} conjugate to \mathbf{x} by the same coefficient $T^{\overline{\mathcal{M}} \rightarrow \mathcal{M}}(\mathbf{x})$, $\forall i \in \{1, \dots, k\}$. Besides, since we assume orthographic viewing, these angles are the same for all pixels, hence $T^{\overline{\mathcal{M}} \rightarrow \mathcal{M}}$ is independent from \mathbf{x} as well - this would not be the case under pinhole projection. The Fresnel coefficient is then written as

$$T^{\overline{\mathcal{M}} \rightarrow \mathcal{M}} = 1 - \frac{1}{2} \left(\frac{(-\mathbf{a}^\top \mathbf{r} + \mu \mathbf{a}^\top \mathbf{e}_3)^2}{(-\mathbf{a}^\top \mathbf{r} - \mu \mathbf{a}^\top \mathbf{e}_3)^2} + \frac{(-\mathbf{a}^\top \mathbf{e}_3 + \mu \mathbf{a}^\top \mathbf{r})^2}{(-\mathbf{a}^\top \mathbf{e}_3 - \mu \mathbf{a}^\top \mathbf{r})^2} \right). \quad (4.4)$$

Note that this second Fresnel coefficient simply scales all the observations by the same constant, hence it can be taken into account by normalization.

Forward Model for Refractive PS To summarize the effects described above, in the presence of refraction the classic image formation model (2.1) becomes

$$I_i(\mathbf{p}) = \underbrace{(\overline{\varphi}_i T_i^{\mathcal{M} \rightarrow \overline{\mathcal{M}}})}_{:= \overline{\psi}_i} \underbrace{(T^{\overline{\mathcal{M}} \rightarrow \mathcal{M}} \rho(\mathbf{x}))}_{:= \varrho(\mathbf{x})} \mathbf{n}(\mathbf{x})^\top \underbrace{(-\text{Snell}_\mu^{-\mathbf{a}}(-\mathbf{s}_i))}_{:= \overline{\mathbf{s}}_i}, \quad i \in \{1, \dots, k\}, \quad (4.5)$$

where:

- the effective lighting directions $\overline{\mathbf{s}}_i$ must be deduced from the calibrated ones \mathbf{s}_i according to Snell's law (4.1);
- the effective lighting densities $\overline{\psi}_i$ must be deduced from the calibrated ones φ_i using (4.2) (density attenuation) and (4.3) (Fresnel coefficients);
- the Fresnel-scaled albedo $\varrho(\mathbf{x})$ (see (4.4)) and the surface normal $\mathbf{n}(\mathbf{x})$ (see (3.6)) constitute the unknowns of the PS problem.

To summarize, we have established the geometric parameterization of the surface, and shown how to deduce the effective lighting directions and densities from the ones calibrated outside the refractive medium. In the next section, we turn our attention to the numerical resolution of the system of equations (4.5), by adapting state-of-the-art strategies.

5 Solving Refractive PS

To invert the image formation model (4.5), it is possible to either sequentially estimate normals and the 3D-shape, or to directly compute the 3D-shape.

Normal and Albedo Estimation Estimating the surface normals and albedo comes down to solving the system of equations (4.5) with known effective lighting densities $\bar{\psi}_i$ and effective incident lighting directions $\bar{\mathbf{s}}_i$. This system of equations admits a unique approximate solution as long as $k \geq 3$ and the effective directions $\bar{\mathbf{s}}_i$ are non-coplanar (which is the case if the incident directions \mathbf{s}_i are themselves non-coplanar). Any calibrated PS method can be applied for this task, simply changing the light directions and densities so as to take refraction into account. For instance, defining $\mathbf{m} := \varrho \mathbf{n}$, one may consider the following pixelwise linear least-squares solution, $\forall \mathbf{p}$:

$$\mathbf{m}(\mathbf{p}) = \operatorname{argmin}_{\mathbf{m} \in \mathbb{R}^3} \sum_{i=1}^k (\bar{\psi}_i \bar{\mathbf{s}}_i^\top \mathbf{m} - I_i(\mathbf{p}))^2, \quad \varrho(\mathbf{x}) = |\mathbf{m}(\mathbf{p})|, \quad \mathbf{n}(\mathbf{x}) = \frac{\mathbf{m}(\mathbf{p})}{|\mathbf{m}(\mathbf{p})|}, \quad (5.1)$$

which can be computed in closed-form by using the pseudo-inverse. If robustness (e.g., to shadows or specularities) needs to be addressed, more evolved solutions based on deep neural networks [7] can be considered. Semi-calibrated algorithms [8] could also be employed for automatically inferring the coefficients $\bar{\psi}_i$. Provided that the integrability constraint [28] is adapted to the refractive case, uncalibrated algorithms [12] would even provide the $\bar{\mathbf{s}}_i$ up to a generalized bas-relief ambiguity [2], which could be resolved a posteriori using one of the methods discussed in [24].

Normal Integration The next stage consists in obtaining the surface from its normals. Eq. (3.6) tells us that once $\mathbf{n}(u, v)$ is estimated, computing $\mathbf{B}^{-1} \mathbf{n}(u, v)$ using the definition in (3.6) of \mathbf{B} , and then normalizing both its first components by the third one provides an estimate for $\nabla \bar{z}(u, v)$. Given these gradient estimates, the pseudo-depth map from the interface can be obtained by integration. Any approach designed for the classic case can be employed at this stage, just changing the input gradient estimates (see [20]). Once the pseudo-depth has been computed, one simply has to apply Eq. (3.4) to obtain the 3D-surface.

Direct Differential Approach To avoid bias accumulation due to the sequential estimation of normals and shape, it is also possible to follow a direct differential approach. Plugging (3.6) into (4.5), we get, $\forall (i, \mathbf{p})$:

$$I^i(\mathbf{p}) = \bar{\psi}_i \underbrace{\left[\mathbf{B} \begin{pmatrix} \varrho(\mathbf{x}) \\ \nabla \bar{z}(\mathbf{p}) \\ -1 \end{pmatrix} \right]}_{:= \bar{\varrho}(\mathbf{p})} \underbrace{(\mathbf{B}^\top \bar{\mathbf{s}}_i)^\top}_{:= \bar{\mathbf{s}}_i^\top} \begin{pmatrix} \nabla \bar{z}(\mathbf{p}) \\ -1 \end{pmatrix}, \quad (5.2)$$

which is a system of nonlinear PDEs. Therein, $\tilde{\varrho}$ will be considered as the unknown “pseudo-albedo” and vectors $\tilde{\mathbf{s}}_i$ as known “pseudo light vectors”. The direct joint estimation of the pseudo-albedo and the pseudo-depth from the interface can then be written as a variational problem:

$$\min_{\tilde{z}, \tilde{\varrho}} \sum_{\mathbf{p}} \sum_i \Phi \left(\overline{\psi}_i \tilde{\varrho}(\mathbf{p}) \tilde{\mathbf{s}}_i^\top \begin{pmatrix} \nabla \tilde{z}(\mathbf{p}) \\ -1 \end{pmatrix} - I_i(\mathbf{p}) \right), \quad (5.3)$$

using some robust estimator Φ and a finite differences approximation of the gradient operator. Once the depth from the interface and the pseudo-albedo have been estimated, it only remains to deduce the true Fresnel-scaled albedo ϱ from $\tilde{\varrho}$ and $\nabla \tilde{z}$ using the definition in Eq. (5.2), and eventually the 3D-surface by using Eq. (3.4). Again, such a differential approach can be extended to the semi-calibrated scenario [21], or even to refine the pseudo light vectors [22].

Validation on Synthetic and Real-world Data In order to empirically validate our forward model and its inversion, we first generated synthetic PS images using Blender [1]. The Lambertian surfaces were placed inside glass ($\bar{n} = 1.5$) while the light sources and the orthographic camera were placed inside air ($n = 1$). In each experiment, 12 images were rendered under varying parallel lighting, whose direction and relative density are provided by the engine.

We first considered images of a perfect sphere. We used the sequential approach (5.1) followed by DCT integration [20], as well as the differential approach (5.3) with Cauchy estimator [22]. In both cases, we carried out 3D-reconstruction first neglecting all refraction effects, and then with refraction considered. To quantitatively evaluate the results, we fit a sphere to the 3D-reconstruction using least-squares, and compute the normalized RMSE between the 3D-reconstruction and the spherical fit. Results are shown in Table 1. It can be seen that for both approaches, considering refraction drastically improves the 3D-reconstruction, even when the interface is not rotated. Indeed, as can be seen in Fig. 3, neglecting refraction causes the 3D-reconstruction to “flatten”.

	No interface	(0°, 0°)	(11.5°, 0°)	(11.5°, 22.5°)
Sequential w/o refraction	0.0035	0.0195	0.0232	0.0403
Sequential w/ refraction	0.0035	0.0116	0.0129	0.0261
Differential w/o refraction	0.0020	0.0202	0.0239	0.0396
Differential w/ refraction	0.0020	0.0114	0.0127	0.0254

Table 1. Normalized root mean square error between the estimated surface and a least-squares spherical fit, for a planar refractive interface (the angles stand for the rotations around the horizontal and vertical axes, respectively). Considering refraction systematically improves performance, for both the sequential and the differential approaches.

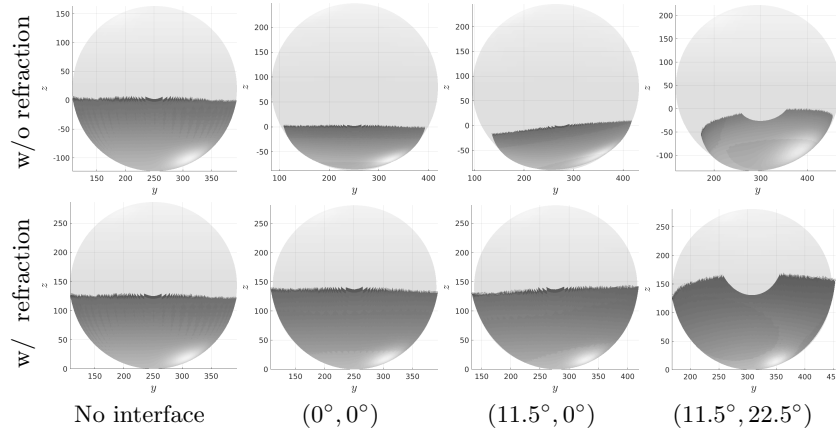


Fig. 3. 3D-reconstructions of the spheres from Table 1 using the differential approach, neglecting (top) or considering (bottom) refraction. The light grey spheres are the least-squares spherical fits to the estimated surfaces used for the quantitative evaluation in Table 1. Neglecting refraction induces a severe “flattening”.

Then, we replaced the sphere by two objects with a more complex shape: an insect (imaged with the interface rotated by 5° around the horizontal axis) and a skull (imaged with the interface rotated by 20° around the horizontal axis, and by 11.25° around the vertical one). The results in Fig. 4, obtained with the differential approach, show that it is possible to achieve a 3D-reconstruction which is indistinguishable from the one obtained in the absence of refraction. In particular, the “flattening” effect is corrected.

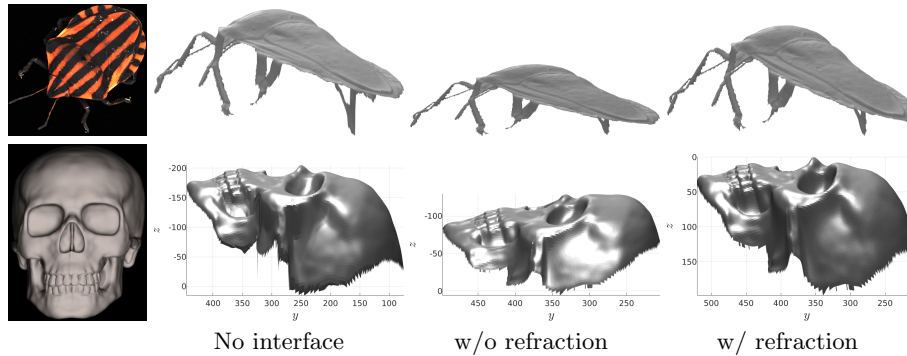


Fig. 4. 3D-reconstruction of an insect (top) and a skull (bottom). In each row, the first image represents one of the input images (out of 12); the second one shows the 3D-reconstruction obtained in the absence of the interface (for reference); and the other ones show the 3D-reconstruction in the presence of the interface, while neglecting or considering refraction.

Lastly, we conducted experiments on a real-world dataset. Our acquisition setup, illustrated in Fig. 1a, consists of 8 calibrated directional light sources. A diffuse sphere was imaged in the air, and then immersed in an aquarium filled with pure water (see Fig. 5). We performed the 3D-reconstruction using (5.3), and compared the results neglecting or considering refraction effects. For both a fronto-parallel and a rotated interface, considering refraction largely reduces the flattening and distortion effects, which empirically validates our method.

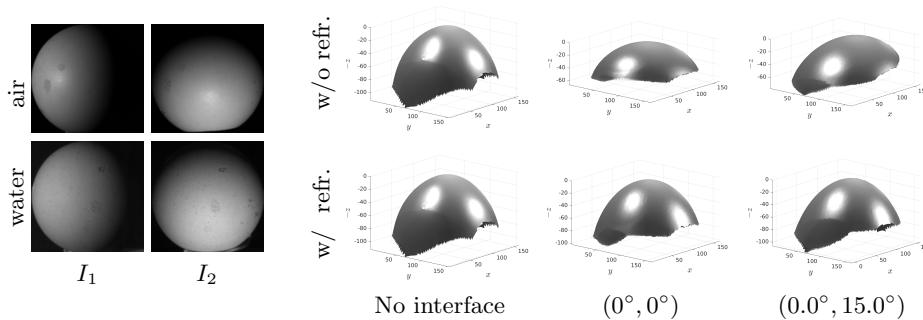


Fig. 5. 3D-reconstruction of a real-world sphere. On the left, we show two of the input images, in the absence (“air”) and in presence (“water”) of a refractive interface which is rotated by 15.0° around the vertical axis. On the right, we show the 3D-reconstruction of the sphere, taking into account (top) or not (bottom) refraction effects, in three cases: in the air, with a fronto-parallel interface, and with a rotated interface. Neglecting refraction leads to flattened and distorted 3D-reconstructions, while these effects are much attenuated with the proposed approach.

6 Conclusion and Future Work

In this paper, we have explored the impact of the presence of a refractive interface on the modeling of the photometric stereo problem, both in terms of geometry and of photometric image formation model. We further showed how to adapt existing solutions so as to take into account geometric deformation, refraction of incident directions, attenuation of densities and Fresnel coefficients. We showed that taking into account such phenomena largely improves the accuracy of the 3D-reconstruction. However, the explicit modeling of refraction effects was eased by a few simplifying assumptions: orthographic viewing, directional lighting, planar interface and absence of light absorption. In the future, we plan to explore the changes induced by the relaxation of these assumptions. This can partially be achieved by making the forward more realistic through the incorporation of, e.g., a refractive near-field illumination model [23] or distance-dependent light attenuation [26,16]. However, we believe that differentiable inverse rendering frameworks may constitute an even more promising track for solving nonstandard photometric 3D-reconstruction problems in a somehow generic manner.

Such approaches have recently been successfully employed for solving complex multi-view 3D-reconstruction problems [15], yet for now they remain limited to cases where the surface projection onto the camera comes down to a simple rasterization. To cope with evolved refractive effects, one could thus imagine combining differentiable inverse rendering with powerful renderers such as Mitsuba 2 [19].

Acknowledgements

This project was partially supported by the KU Data+ Project Phylorama, the ALICIA-Vision LabCom (ANR-19-LCV1-0002), and the Inclusive Museum Guide project (ANR-20-CE38-0007).

References

1. Blender - a 3D modelling and rendering package, <http://www.blender.org>
2. Belhumeur, P.N., Kriegman, D.J., Yuille, A.L.: The bas-relief ambiguity. *IJCV* **35**(1), 33–44 (1999)
3. Cassidy, M., Mélou, J., Quéau, Y., Lauze, F., Durou, J.D.: Refractive multi-view stereo. In: *3DV* (2020)
4. Chadebecq, F., Vasconcelos, F., Lacher, R., Maneas, E., Desjardins, A., Ourselin, S., Vercauteren, T., Stoyanov, D.: Refractive two-view reconstruction for underwater 3d vision. *IJCV* **128**(5), 1101–1117 (2020)
5. Chari, V., Sturm, P.: Multiple-view geometry of the refractive plane. In: *BMVC* (2009)
6. Chen, C., Wang, H., Zhang, Z., Gao, F.: Three-dimensional reconstruction from a fringe projection system through a planar transparent medium. *OptEx* **30**(19), 34824–34834 (2022)
7. Chen, G., Han, K., Shi, B., Matsushita, Y., Wong, K.Y.K.: Deep photometric stereo for non-Lambertian surfaces. *PAMI* **44**(1), 129–142 (2020)
8. Cho, D., Matsushita, Y., Tai, Y.W., Kweon, I.S.: Semi-calibrated photometric stereo. *PAMI* **42**(1), 232–245 (2018)
9. Fan, H., Qi, L., Chen, C., Rao, Y., Kong, L., Dong, J., Yu, H.: Underwater optical 3-D reconstruction of photometric stereo considering light refraction and attenuation. *IEEE J. Ocean. Eng.* **47**(1), 46–58 (2021)
10. Fujimura, Y., Iiyama, M., Hashimoto, A., Minoh, M.: Photometric stereo in participating media considering shape-dependent forward scatter. In: *CVPR* (2018)
11. Fujitomi, T., Sakurada, K., Hamaguchi, R., Shishido, H., Onishi, M., Kameda, Y.: LB-NERF: Light Bending Neural Radiance Fields for Transparent Medium. In: *ICIP* (2022)
12. Hayakawa, H.: Photometric stereo under a light source with arbitrary motion. *JOSA A* **11**(11), 3079–3089 (1994)
13. Hu, X., Lauze, F., Pedersen, K.S.: Refractive Pose Refinement. *IJCV* (2023). <https://doi.org/10.1007/s11263-023-01763-4>
14. Mikš, A., Novák, P.: Determination of unit normal vectors of aspherical surfaces given unit directional vectors of incoming and outgoing rays: comment. *JOSA A* **29**(7), 1356–1357 (2012)

15. Munkberg, J., Hasselgren, J., Shen, T., Gao, J., Chen, W., Evans, A., Müller, T., Fidler, S.: Extracting Triangular 3D Models, Materials, and Lighting From Images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8280–8290 (2022)
16. Murai, S., Kuo, M.Y.J., Kawahara, R., Nobuhara, S., Nishino, K.: Surface normals and shape from water. In: CVPR (2019)
17. Murez, Z., Treibitz, T., Ramamoorthi, R., Kriegman, D.: Photometric stereo in a scattering medium. In: CVPR (2015)
18. Narasimhan, S.G., Nayar, S.K.: Structured light methods for underwater imaging: light stripe scanning and photometric stereo. In: OCEANS (2005)
19. Nimier-David, M., Vicini, D., Zeltner, T., Jakob, W.: Mitsuba 2: A retargetable forward and inverse renderer. *ACM Transactions on Graphics (TOG)* **38**(6), 1–17 (2019)
20. Quéau, Y., Durou, J.D., Aujol, J.F.: Normal integration: a survey. *JMIV* **60**(4), 576–593 (2018)
21. Quéau, Y., Wu, T., Cremers, D.: Semi-calibrated near-light photometric stereo. In: *SSVM* (2017)
22. Quéau, Y., Wu, T., Lauze, F., Durou, J.D., Cremers, D.: A non-convex variational approach to photometric stereo under inaccurate lighting. In: CVPR (2017)
23. Sanao, H., Yingjie, S., Ming, L., Jingwei, Q., Ke, X.: Underwater 3D reconstruction using a photometric stereo with illuminance estimation. *Applied Optics* **62**(3), 612–619 (2023)
24. Shi, B., Wu, Z., Mo, Z., Duan, D., Yeung, S.K., Tan, P.: A benchmark dataset and evaluation for non-Lambertian and uncalibrated photometric stereo. In: CVPR (2016)
25. Tsiotsios, C., Angelopoulou, M.E., Kim, T.K., Davison, A.J.: Backscatter compensated photometric stereo with 3 sources. In: CVPR (2014)
26. Tsiotsios, C., Davison, A.J., Kim, T.K.: Near-lighting Photometric Stereo for unknown scene distance and medium attenuation. *IVC* **57**, 44–57 (2017)
27. Woodham, R.J.: Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. In: *Image understanding systems and industrial applications I*. vol. 155, pp. 136–143 (1979)
28. Yuille, A.L., Snow, D., Epstein, R., Belhumeur, P.N.: Determining generative models of objects under varying illumination: Shape and albedo from multiple images using SVD and integrability. *IJCV* **35**(3), 203–222 (1999)