Variational Uncalibrated Photometric Stereo under General Lighting

Bjoern Haefner, Zhenzhang Ye, Maolin Gao, Tao Wu, Yvain Quéau, Daniel Cremers

Variational Uncalibrated Photometric Stereo under General Lighting

Bjoern Haefner*
TU Munich
Munich, Germany
bjoern.haefner@in.tum.de

Zhenzhang Ye*
TU Munich
Munich, Germany
yez@in.tum.de

Maolin Gao
Artisense
Garching, Germany
maolin.gao@in.tum.de

Tao Wu
TU Munich
Munich, Germany
tao.wu@tum.de

Yvain Quéau
GREYC, UMR CNRS 6072
Caen, France
yvain.queau@ensicaen.fr

Daniel Cremers
TU Munich
Munich, Germany
cremers@tum.de

Abstract

Photometric stereo (PS) techniques nowadays remain constrained to an ideal laboratory setup where modeling and calibration of lighting is amenable. To eliminate such restrictions, we propose an efficient principled variational approach to uncalibrated PS under general illumination. To this end, the Lambertian reflectance model is approximated through a spherical harmonic expansion, which preserves the spatial invariance of the lighting. The joint recovery of shape, reflectance and illumination is then formulated as a single variational problem. There the shape estimation is carried out directly in terms of the underlying perspective depth map, thus implicitly ensuring integrability and bypassing the need for a subsequent normal integration. To tackle the resulting nonconvex problem numerically, we undertake a two-phase procedure to initialize a balloon-like perspective depth map, followed by a "lagged" block coordinate descent scheme. The experiments validate efficiency and robustness of this approach. Across a variety of evaluations, we are able to reduce the mean angular error consistently by a factor of 2–3 *compared to the state-of-the-art.*

1. Introduction

Photometric stereo techniques aim at acquiring both the shape and the reflectance of a scene. To this end, multiple images are acquired under the same viewing angle but varying lighting, and a physics-based image formation model is inverted. However, the classic way to solve this inverse problem requires lighting to be highly controlled, which restricts practical applications to laboratory setups where careful calibration of lighting must be carried out.

The objective of this research work is to simplify the overall photometric stereo pipeline, by providing an efficient solution to uncalibrated photometric stereo under general lighting, as illustrated in Figure 1 (the code is released[1]). In comparison with existing efforts in the same direction, the proposed one has the following advantages:

- The joint estimation of shape, reflectance and general lighting is formulated as an end-to-end, mathematically transparent variational problem;
- A real 3D-surface represented as a depth map is recovered, rather than possibly non-integrable normals;
- It is robust, due to the use of Cauchy's robust M-estimator and Huber-TV albedo regularization;
- It is computationally efficient, thanks to a tailored lagged block coordinate descent scheme initialized using a simple balloon-like shape.

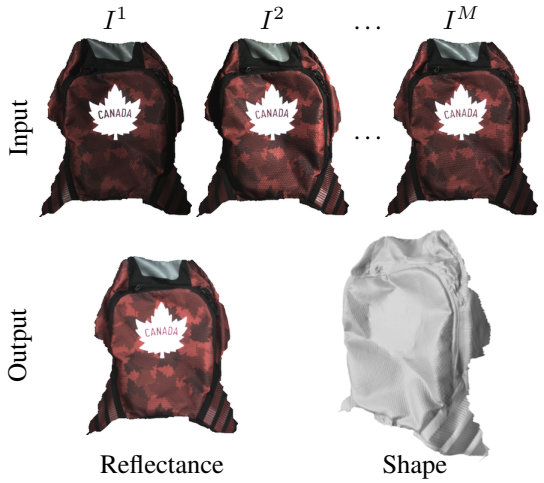


Figure 1. We present an efficient variational scheme to solve uncalibrated photometric stereo under general lighting. Given a set of input RGB images captured from the same viewing angle but under unknown, varying general illumination (top, $M = 20$ images were acquired in an office under daylight, while freely moving a hand-held LED light source), fine-detailed reflectance and shape (bottom, we show the estimated albedo and perspective depth maps) are recovered by an end-to-end variational approach.

After reviewing related works in Section 2, we discuss in Section 3 the image formation model considered in this work. It can be inverted using the variational approach in Section 4. A dedicated numerical solution is then introduced in Section 5 and empirically evaluated in Section 6. Section 7 eventually draws the conclusion of this research.

*Authors contributed equally.

[1] https://github.com/zhenzhangye/general_ups

2. Related Work

3D-models of scenes are essential in many applications such as visual inspection [14] or computer-aided surgery using augmented reality [12]. A 3D-model consists of geometric (position, orientation, etc.) and photometric (color, texture, etc.) properties. Given a set of photographies, the aim of 3D scanning is to invert the image formation process in order to recover these geometric and photometric properties of the observed scene. This notion thus includes both those of 3D-reconstruction (geometry) and of reflectance estimation (photometry).

Many approaches to the problem of 3D-reconstruction from photographies have been studied, and they are grouped under the generic naming "shape-from-X", where X stands for the clue which is being used (shadows [44], contours [10], texture [49], template [6], structured light [16], motion [35], focus [36], silhouettes [21], etc.). Geometric shape-from-X techniques are based on the identification and analysis of feature point or areas in the image. In contrast, photometric techniques build upon the analysis of the quantity of light received by each photosite of the camera's sensor. Among photometric techniques, *shape-from-shading* is probably the most famous one. This technique, developed in the 70s by Horn *et al.* [25], consists in 3D-reconstruction from a single image of a shaded scene. It is a classic ill-posed inverse problem whose numerical solving usually requires the surface's reflectance to be known [13]. In order both to limit the ambiguities of shape-from-shading and to allow for automatic reflectance estimation, it has been suggested to consider not just one image of the scene, but several ones acquired from the same viewing angle but under varying lighting. This variant, which was introduced in the late 70s by Woodham [50], is known as *photometric stereo*.

Among the various shape-from-X techniques mentioned above, photometric stereo is the only 3D-scanning technique i.e., the only one which is able to achieve both 3D-reconstruction and reflectance estimation. However, early photometric approaches strongly rely on the control of lighting. The latter is usually assumed for simplicity to be directional, although the case of nearby point light sources has recently regained some attention [31, 33]. More importantly, lighting is assumed to be calibrated. Indeed, the uncalibrated problem is ill-posed: the underlying normal map can be estimated only up to a linear ambiguity [20], which reduces to a generalized bas-relief one if integrability is enforced [9]. To resolve the latter ambiguity, some prior on the scene's surface or geometry must be introduced, see [48] for a recent survey. A natural way to enforce integrability consists in following a differential approach to photometric stereo [11, 32] i.e., directly estimate the 3D-surface as a depth map instead of first estimating the surface normals and then integrating them. Such a differential approach to photometric stereo can be coupled with variational methods in order to iteratively refine depth, reflectance and lighting in a robust manner [42]. In addition to the theoretical interest of enforcing integrability in order to limit ambiguities, differential approaches to photometric stereo have the advantages of easing combination with other 3D-reconstruction methods [17, 40], and of bypassing the problem of integrating the estimated normal field, which is by itself a non-trivial problem [41]. Besides, any error in the estimated normal field might propagate during integration, and thus robustness to specularities or shadows must be enforced during normal estimation, see again [48] for some discussion.

All the research works mentioned in the previous paragraph assume that lighting is induced by a single light source. Nevertheless, many studies rather considered the case of more general illumination conditions, which finds a natural application in outdoor conditions [43]. For instance, the apparent motion of the sun within a day induces changes in the illumination direction which, in theory, allow photometric stereo-based 3D-reconstruction. However, this apparent motion is close to being planar, and thus the rank of the set of illumination vectors is equal or close to 2 [45] (see also [23] for additional discussion on the stability of single-day photometric stereo). This situation is thus similar to the two-image case, which is known to be ill-posed since the early 90s [28, 37, 51], although it is still an active research area [29]. In order to limit the instabilities due to this issue, one possibility is to consider images acquired over many seasons as in [2, 3], or to resort to deep neural networks [22]. Another one is to consider a non-directional illumination model to represent natural illumination, as for instance in [26]. Modeling natural illumination is a promising track, since such a model would not be restricted to sunny days, and images acquired under cloudy days are known to yield more accurate 3D-reconstructions [23].

However, the previous approaches to photometric stereo under natural illumination assume calibrated lighting, where calibration is deduced from time and GPS coordinates or from a calibration target. The case of both general and uncalibrated lighting is much more challenging and has been fewly explored, apart from studies restricted to sparse 3D-reconstructions [46] or relying on the prior knowledge of a rough geometry [4, 27, 40, 47]. Uncalibrated photometric stereo under natural illumination has been revisited recently in [34], using a spatially-varying equivalent directional lighting model. However, results were limited to the recovery of possibly non-integrable surface normals. Instead, the method which we propose in the present paper directly recovers the underlying surface represented as a depth map. Following the seminal work of Basri and Jacobs [7], it considers the spherical harmonics representation of general lighting in lieu of the equivalent directional approximation, as discussed in the next section.

3. Image Formation Model

In photometric stereo (PS), we are given a number of observations $\{I^i\}_{i=1}^M$, each $I^i : \Omega \subset \mathbb{R}^2 \to \mathbb{R}^C$ representing a multi-channel image (i.e. $C \geq 1$) over a masked pixel domain Ω. Assuming that the object being pictured is Lambertian, the surface's reflectance is represented by the albedo ρ, and the general image formation model is as follows, for all $i \in \{1, ..., M\}$, $c \in \{1, ..., C\}$, and $\mathbf{p} \in \Omega$:

$$I_c^i(\mathbf{p}) = \int_{\mathbb{S}^2} \rho_c(\mathbf{p}) \ell_c^i(\boldsymbol{\omega}) \max\{\boldsymbol{\omega} \cdot \mathbf{n}(\mathbf{p}), 0\} \, \mathrm{d}\boldsymbol{\omega}. \tag{1}$$

Here $\mathbb{S}^2$ is the unit sphere in $\mathbb{R}^3$, $\ell_c^i : \mathbb{S}^2 \to \mathbb{R}_+$ represents the channel-wise intensity of the incident light, and $\rho_c(\mathbf{p}) \in \mathbb{R}_+$ and $\mathbf{n}(\mathbf{p}) \in \mathbb{S}^2$ are the channel-wise albedos and the unit-length surface normals, respectively, at the surface point conjugate to pixel $\mathbf{p} \in \Omega$. The max operation in (1) encodes self-shadows. The overall integral $\int_{\mathbb{S}^2}$ collects elementary luminance contributions arising from all incident lighting directions $\boldsymbol{\omega}$. In the setup of uncalibrated PS, the quantities $\{\ell_c^i\}$, $\{\rho_c\}$, in addition to $\mathbf{n}$, are unknown.

Equivalent directional lighting [24] approximates (1) via

$$\begin{aligned} I_c^i(\mathbf{p}) &= \rho_c(\mathbf{p}) \, \bar{\ell}_c^i(\mathbf{p}) \cdot \mathbf{n}(\mathbf{p}), \\ \bar{\ell}_c^i(\mathbf{p}) &:= \int_{\{\boldsymbol{\omega} \in \mathbb{S}^2 : \boldsymbol{\omega} \cdot \mathbf{n}(\mathbf{p}) \geq 0\}} \ell_c^i(\boldsymbol{\omega}) \boldsymbol{\omega} \, \mathrm{d}\boldsymbol{\omega}. \end{aligned} \tag{2}$$

where $\bar{\ell}_c^i(\mathbf{p})$ represents the mean lighting over the visible hemisphere at $\mathbf{p}$. The field $\bar{\ell}_i^c$ is *spatially variant* but can be approximated by directional lighting over small local patches. Over each patch, one is thus faced with the ambiguities of directional uncalibrated PS [20]. State-of-the-art patch-wise methods [34] first solve this problem over each patch, then connect the patches to form a complete normal field up to rotation, and eventually estimate the rotation which best satisfies the integrability constraint. Errors may however get propagated during the sequence, resulting in a possibly non-integrable normal field.

Instead of such an equivalent directional lighting model, we rather consider a *spherical harmonic approximation* (SHA) of general lighting [8, 7]. By defining the half-cosine kernel k as

$$k(\boldsymbol{\omega}, \mathbf{n}) := \max\{\boldsymbol{\omega} \cdot \mathbf{n}, 0\}, \tag{3}$$

we can view (1) as an analog of a convolution:

$$I_c^i(\mathbf{p}) = \rho_c(\mathbf{p}) \int_{\mathbb{S}^2} k(\boldsymbol{\omega}, \mathbf{n}(\mathbf{p})) \ell_c^i(\boldsymbol{\omega}) \, \mathrm{d}\boldsymbol{\omega}. \tag{4}$$

Invoking the Funk-Hecke theorem, we obtain the following harmonic expansion analogous to Fourier series:

$$\int_{\mathbb{S}^2} k(\boldsymbol{\omega}, \mathbf{n}(\mathbf{p})) \ell_c^i(\boldsymbol{\omega}) \, \mathrm{d}\boldsymbol{\omega} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} (k_n \ell_{n,m}^{i,c}) h_{n,m}(\mathbf{n}(\mathbf{p})). \tag{5}$$

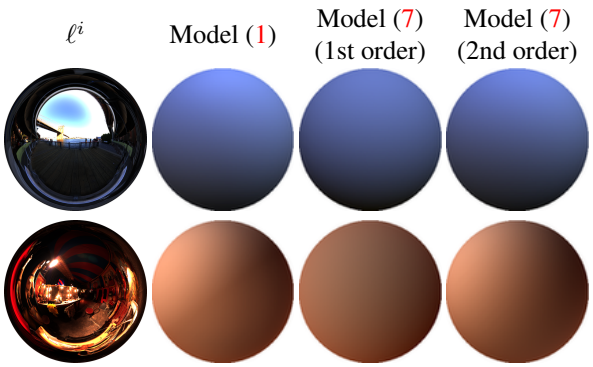


Figure 2. Illustration of RGB ($C = 3$) environment lighting $\ell^i = (\ell_1^i, \ell_2^i, \ell_3^i)$, the resulting images (assuming white albedos and a spherical shape) under the image formation model (1) and its approximation by spherical harmonics. The approximation by the second-order spherical harmonics is nearly perfect.

Here the spherical harmonics $\{h_{n,m}\}$ form an orthonormal basis of $L^2(\mathbb{S}^2)$, and $\{k_n\}$ and $\{\ell_{n,m}^{i,c}\}$ are the expansion coefficients of k and ℓ_c^i with respect to $\{h_{n,m}\}$. Since most energy in the expansion (5) concentrates on low-order terms [8], we obtain the *second-order SHA* by truncating the series up to the first nine terms (i.e., $0 \leq n \leq 2$):

$$\int_{\mathbb{S}^2} k(\boldsymbol{\omega}, \mathbf{n}(\mathbf{p})) \ell_c^i(\boldsymbol{\omega}) \, \mathrm{d}\boldsymbol{\omega} \approx \sum_{n=0}^{2} \sum_{m=-n}^{n} (k_n \ell_{n,m}^{i,c}) h_{n,m}(\mathbf{n}(\mathbf{p})). \tag{6}$$

The first-order SHA refers to the truncation up to the first four terms (i.e., $0 \leq n \leq 1$). It is shown in [8] that, for distant lighting, at least 75% of the resulting irradiance is captured by the first-order SHA, and 98% by the second-order SHA (cf. Figure 2 for a visualization).

Plugging (6) and specifics of spherical harmonics [8] into (4), we finalize our image formation model as:

$$I_c^i(\mathbf{p}) \approx \rho_c(\mathbf{p}) \, \mathbf{l}_c^i \cdot \mathbf{h}[\mathbf{n}](\mathbf{p}), \tag{7}$$

$$\mathbf{h}[\mathbf{n}] = \left[1, \mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3, \mathbf{n}_1\mathbf{n}_2, \mathbf{n}_1\mathbf{n}_3, \mathbf{n}_2\mathbf{n}_3, \mathbf{n}_1^2 - \mathbf{n}_2^2, 3\mathbf{n}_3^2 - 1\right]^\top. \tag{8}$$

Here $\mathbf{h}[\mathbf{n}] : \Omega \to \mathbb{R}^9$ represents the second-order harmonic images, and $\mathbf{l}_c^i \in \mathbb{R}^9$ represents the harmonic lighting vector whose entries have absorbed $\{k_n \ell_{n,m}^{i,c}\}$ and constant factors of $\{h_{n,m}\}$. A key advantage of the SHA (7) over the equivalent directional lighting model (2) lies in the *spatial invariance* of the lighting vectors $\{\mathbf{l}_c^i\}$, which yields a less ill-posed inverse problem [7]. The counterpart is the non-linear dependency upon the normal components, which we will handle in Section 5 using a tailored numerical solution. In the next section, we build upon the key observations that integrability [9] and perspective projection [39] both largely reduce the ambiguities of uncalibrated PS to derive a variational approach to inverting the SHA (7).

4. Variational Uncalibrated PS

In this section, we shall propose a joint variational model for uncalibrated PS. To this end, let a 3D-frame $(Oxyz)$ be attached to the camera, with O the optical center, the z-axis aligned with the optical axis such that $z > 0$ for any 3D point (x, y, z) in front of the camera. Further let a 2D-frame $(O'uv)$ be attached to the focal plane which is parallel to the xy-plane and contains the masked pixel domain Ω. Under perspective projection, the surface geometry is modeled as a map $\mathbf{x} : \mathbf{p} = (u, v) \in \Omega \mapsto \mathbf{x}(u, v) \in \mathbb{R}^3$ given by

$$\mathbf{x}(u, v) = z(u, v)K^{-1}[u, v, 1]^\top, \tag{9}$$

with $z : \Omega \to \mathbb{R}_+$ the *depth* map and

$$K := \begin{bmatrix} f_u & 0 & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{10}$$

the calibrated camera's intrinsics matrix. In the following we denote for convenience $(\widetilde{u}, \widetilde{v}) := (u - u_0, v - v_0)$.

Assuming that z is differentiable, the surface normal $\mathbf{n}$ at point $\mathbf{x}(u, v)$ is the unit vector oriented towards the camera such that $\mathbf{n}(u, v) \propto \partial_u \mathbf{x}(u, v) \times \partial_v \mathbf{x}(u, v)$, which yields the following parameterization of the normal by the depth:

$$\mathbf{n}[z](u, v) = \frac{\widetilde{\mathbf{n}}[z](u, v)}{|\widetilde{\mathbf{n}}[z](u, v)|}, \tag{11}$$

$$\widetilde{\mathbf{n}}[z](u, v) := \begin{bmatrix} f_u \partial_u z(u, v) \\ f_v \partial_v z(u, v) \\ -z(u, v) - \widetilde{u}\, \partial_u z(u, v) - \widetilde{v}\, \partial_v z(u, v) \end{bmatrix}. \tag{12}$$

Note that the dependence of $\widetilde{\mathbf{n}}[z]$ on z is linear.

Based on the forward model (7) and the parameterization (11) of normals, we formulate the joint recovery of reflectance, lighting and geometry as the following variational problem:

$$\min_{\{\rho_c\},\{\mathbf{l}_c^i\},z} \sum_{i=1}^{M} \sum_{c=1}^{C} \int_\Omega \phi_\lambda \Big(\rho_c(u, v)\, \mathbf{l}_c^i \cdot \mathbf{h}[\mathbf{n}[z]](u, v) - I_c^i(u, v) \Big)\, \mathrm{d}u\, \mathrm{d}v + \mu \sum_{c=1}^{C} \int_\Omega |\nabla \rho_c(u, v)|_\gamma \, \mathrm{d}u\, \mathrm{d}v. \tag{13}$$

In the first term above, we use *Cauchy's M-estimator* to penalize the data-fitting discrepancy:

$$\phi_\lambda(s) = \lambda^2 \log(1 + s^2/\lambda^2), \tag{14}$$

It is indeed well-known that Cauchy's estimator, being non-convex, is robust against outliers; see for instance [42] in the context of PS. The scaling parameter $\lambda = 0.15$ is used in all experiments.

The second term in (13) represents a Huber total-variation (TV) regularization on each albedo map ρ_c, with the Huber loss defined by

$$|s|_\gamma := \begin{cases} |s|^2/(2\gamma) & \text{if } |s| \le \gamma, \\ |s| - \gamma/2 & \text{if } |s| > \gamma, \end{cases} \tag{15}$$

and $\gamma = 0.1$ being fixed in the experiments. It turns out that the Huber TV imposes desirable smoothness on the albedo maps $\{\rho_c\}$ and in turn improves the joint estimation overall. Eventually, $\mu > 0$ is a weight parameter which balances the data-fitting term and the Huber TV one. Its value was empirically set to $2 \cdot 10^{-6}$ (see Section 6 for some discussion).

In (13), geometry is directly optimized in terms of the depth z (rather than indirectly in terms of the normal $\mathbf{n}$). This both ensures integrability and avoids integration of normals into depths as a post-processing step.

5. Solver and Implementation

To solve the variational problem (13) numerically, we follow a "discretize-then-optimize" approach. There, $\Omega \subset \mathbb{R}^2$ is replaced by $\mathbb{R}^N$, N being the number of pixels inside Ω, which yields discretized vectors $z, \{\rho_c\}_{c=1}^C \in \mathbb{R}^N$. To alleviate notational burden, we sometimes refer to a pixel by its index $j \in \{1, \dots, N\}$ and sometimes by its position $\mathbf{p} = (u, v) \in \Omega$. The spatial gradient ∇ is discretized using a forward difference stencil.

We shall apply a lagged block coordinate descent (LBCD) method to find a local minimum of the objective function in (23). Due to the (highly) non-convex nature of (23), initialization of optimization variables has a strong influence on the final solution. In our implementation, we initialize $\rho_{c,j} = \text{median}(\{I_{c,j}^i\}_{i=1}^M)$ for all c, j and $\mathbf{l}_c^i = [0.2, 0, 0, -1, 0, 0, 0, 0, 0]^\top$ for all c, i. Moreover, during the first eight iterations we freeze the second-order spherical harmonics coefficients $(\mathbf{l}_c^i)_5 = (\mathbf{l}_c^i)_6 = \dots = (\mathbf{l}_c^i)_9 = 0$ i.e., we reconstruct using only first-order spherical harmonic approximation as a warm start. Most real-world scenes being convex, we initialize the depth z as a balloon-like surface, as discussed in the following.

5.1. Depth Initialization

It is readily seen that a trivial constant initialization of the depth z yields uniform vertically aligned normals $\mathbf{n}[z]$ and, hence, zero entries in the initial harmonic images $\mathbf{h}[\mathbf{n}[z]]$. This would cause non-meaningful updates on albedos $\{\rho_c\}$ and lighting vectors $\{\mathbf{l}_c^i\}$; cf. Figure 3 for an illustration.

To solve this issue, we specialize the depth initialization which undergoes two phases:

1. Following [38], we generate a balloon-like depth map z_o under orthographic projection.

2. We then convert the orthographic depth z_o to a perspective depth z_p via normal integration [41].

Phase 1 is pursued via seeking a depth map z_o which has minimal surface area subject to a constant volume V:

$$\begin{aligned} &\min_{z_o} \int_\Omega \sqrt{1+|\nabla z_o|^2}\,\mathrm{d}u\,\mathrm{d}v \\ &\text{s.t.} \int_\Omega z_o\,\mathrm{d}u\,\mathrm{d}v = V. \end{aligned} \tag{16}$$

A global minimizer of this model can be efficiently computed by simple projected gradient iterations:

$$z_o^{(k+1/2)} = z_o^{(k)} - \tau\nabla^\top\left(\frac{1}{\sqrt{1+|\nabla z_o^{(k)}|^2}}\nabla z_o^{(k)}\right), \tag{17}$$

$$z_o^{(k+1)} = z_o^{(k+1/2)} + \left(\frac{V - \int_\Omega z_o^{(k+1/2)}\,\mathrm{d}u\,\mathrm{d}v}{\int_\Omega \mathrm{d}u\,\mathrm{d}v}\right)\cdot \mathbf{1}_\Omega, \tag{18}$$

where $\mathbf{1}_\Omega(u,v)\equiv 1$ and $\tau = 0.8/\|\nabla\|_{\text{spec}}$ with $\|\cdot\|_{\text{spec}}$ the spectral norm. The volume constant V is a hyperparameter which is empirically chosen, see Section 6 for discussion.

Next, we convert the orthographic depth z_o to a perspective depth z_p. Note that z_o complies with the orthographic projection, under which a 3D-point $\widehat{\mathbf{x}}$ is represented by

$$\widehat{\mathbf{x}}(u,v) = [u, v, z_o(u,v)]^\top, \tag{19}$$

and the corresponding surface normal $\widehat{\mathbf{n}}$ to the surface at $\widehat{\mathbf{x}}$ conjugate to pixel $\widehat{\mathbf{p}} = (u,v)$ is given by

$$\widehat{\mathbf{n}}(u,v) = \frac{1}{\sqrt{|\nabla z_o(u,v)|^2+1}}[\nabla z_o(u,v), -1]^\top. \tag{20}$$

Since $\widehat{\mathbf{n}}$ is invariant to the projection model, Eq. (11) also implies that

$$\widehat{\mathbf{n}}(u,v) \propto \begin{bmatrix} f_u\partial_u\widehat{z}_p(u,v) \\ f_v\partial_v\widehat{z}_p(u,v) \\ -1 - \widetilde{u}\partial_u\widehat{z}_p(u,v) - \widetilde{v}\partial_v\widehat{z}_p(u,v)\end{bmatrix}, \tag{21}$$

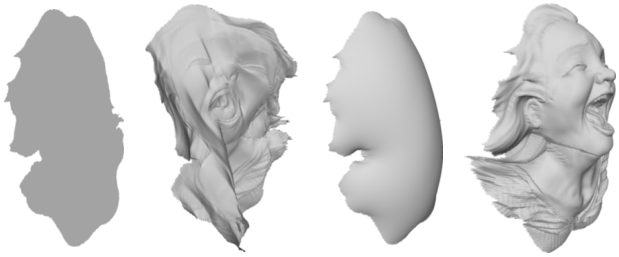
where $\widehat{z}_p(u,v) = \log z_p(u,v)$ stands for the log-perspective depth. This further implies the formula for $\nabla\widehat{z}_p$:

$$\nabla\widehat{z}_p(u,v) = \frac{-1}{\frac{\widetilde{u}\widehat{\mathbf{n}}_1(u,v)}{f_u} + \frac{\widetilde{v}\widehat{\mathbf{n}}_2(u,v)}{f_v} + \widehat{\mathbf{n}}_3(u,v)}\begin{bmatrix}\frac{1}{f_u}\widehat{\mathbf{n}}_1(u,v) \\ \frac{1}{f_v}\widehat{\mathbf{n}}_2(u,v)\end{bmatrix}, \tag{22}$$

which can be integrated to obtain $\widehat{z}_p$ (and hence z_p). The overall pipeline in Phase 2 is summarized as follows:

1. ($z_o \to \widehat{\mathbf{n}}$): Compute $\widehat{\mathbf{n}}$ by (20).
2. ($\widehat{\mathbf{n}} \to \nabla\widehat{z}_p$): Compute $\nabla\widehat{z}_p$ by (22).
3. ($\nabla\widehat{z}_p \to z_p$): Perform integration [41] to obtain $\widehat{z}_p$. Return $z_p = \exp\widehat{z}_p$ as the initialized (perspective) depth.

As discussed in [18] the perspective surface area depends linearly on the depth z. This complicates direct perspective ballooning, since the depth is driven towards zero and hence yields numerical instability. For this reason, we opted for the two-step approach which bypasses the issue.



Trivial $z_p \equiv 1$ and its result z Our z_p and its result z

Figure 3. Impact of depth initialization: a trivial constant initialization on the left vs. our initialization on the right and its corresponding resulting geometry estimates. Further results from varying initializations can be found in the supplementary material.

5.2. Lagged Block Coordinate Descent

Even with a reasonable initialization, the numerical resolution of Problem (23) remains challenging. Due to the appearances of the spherical harmonic approximation $\mathbf{h}[\mathbf{n}[z]]$ and the Cauchy's M-estimator ϕ_λ, the objective in (23) is highly nonlinear and nonconvex. To tackle these challenges, here we present a lagged block coordinate descent (LBCD) method which performs efficiently in practice.

To derive LBCD, we introduce an auxiliary variable $\theta \in \mathbb{R}^N$ such that $\theta_j = |\widetilde{\mathbf{n}}_j[z]|$. This enables us to rewrite (11) as $\mathbf{n}_j[z] = \widetilde{\mathbf{n}}_j[z]/\theta_j$. Then we formulate the following constrained optimization problem:

$$\begin{aligned} \min_{\theta,\{\rho_c\},\{\mathbf{l}_c^i\},z} &\sum_{i=1}^{M}\sum_{c=1}^{C}\sum_{j=1}^{N}\phi_\lambda\left(r_{i,c,j}(\theta_j,\rho_{c,j},\mathbf{l}_c^i,z)\right) \\ &+\mu\sum_{c=1}^{C}\sum_{j=1}^{N}|(\nabla\rho_c)_j|_\gamma, \\ \text{s.t. }\ &\theta_j = |\widetilde{\mathbf{n}}_j[z]|,\ \forall j\in\{1,\dots,N\}, \end{aligned} \tag{23}$$

where $r_{i,c,j}$ is the residual function defined by:

$$r_{i,c,j}(\theta_j,\rho_{c,j},\mathbf{l}_c^i,z) = \rho_{c,j}\,\mathbf{l}_c^i\cdot\mathbf{h}_j[\widetilde{\mathbf{n}}_j[z]/\theta_j] - I_{c,j}^i. \tag{24}$$

Upon initialization, the proposed LBCD proceeds as follows. At iteration k, we lag θ one iteration behind, i.e.,

$$\theta_j^{(k+1)} := |\widetilde{\mathbf{n}}_j[z^{(k)}]|, \quad \forall j\in\{1,...,N\}, \tag{25}$$

and then sequentially update each of the three blocks (namely $\{\rho_c\}$, $\{\mathbf{l}_c^i\}$ and z). In each resulting subproblem,

we solve (lagged) weighted least squares problems as an approximation of the Cauchy loss and/or the Huber loss. This is detailed in the following:

- (Update $\{\rho_c\}$): We evaluate the residual

$$r_{i,c,j}^{(k+1/3)} := r_{i,c,j}(\theta_j^{(k+1)}, \rho_{c,j}^{(k)}, \mathbf{l}_c^{i,(k)}, z^{(k)}), \tag{26}$$

and then set up the (lagged) weight factors for both the Cauchy loss and the Huber loss as

$$w_{i,c,j}^{(k+1/3)} := \phi_\lambda'(r_{i,c,j}^{(k+1/3)})/r_{i,c,j}^{(k+1/3)}, \tag{27}$$

$$q_{c,j}^{(k+1/3)} := 1/\max\{\gamma, |(\nabla\rho_c^{(k)})_j|\}. \tag{28}$$

The albedos $\{\rho_c\}$ are updated as the solution to the following linear weighted least-squares problem:

$$\begin{aligned}\{\rho_c^{(k+1)}\} := \arg\min_{\{\rho_c\}} \mu \sum_{c,j} q_{c,j}^{(k+1/3)} |(\nabla\rho_c)_j|^2 \\ + \sum_{i,c,j} w_{i,c,j}^{(k+1/3)} |r_{i,c,j}(\theta_j^{(k+1)}, \rho_j^c, \mathbf{l}_c^{i,(k)}, z^{(k)})|^2,\end{aligned} \tag{29}$$

which is carried out by conjugate gradient (CG).

- (Update $\{\mathbf{l}_c^i\}$): The lighting subproblem is similar to the one for albedos, except for absence of the Huber TV term. Upon evaluation of the residual $r_{i,c,j}^{(k+2/3)}$ and the weight factor $w_{i,c,j}^{(k+2/3)}$, we update $\{\mathbf{l}_c^i\}$ by solving the following linear weighted least-squares problem via CG:

$$\begin{aligned}\{\mathbf{l}_c^{i,(k+1)}\} = \arg\min_{\mathbf{l}_c^i} \sum_{i,c,j} w_{i,c,j}^{(k+2/3)} \cdot \\ |r_{i,c,j}(\theta_j^{(k+1)}, \rho_{c,j}^{(k+1)}, \mathbf{l}_c^i, z^{(k)})|^2.\end{aligned} \tag{30}$$

- (Update z): The depth subproblem requires additional efforts. With $r_{i,c,j}^{(k+1)}$ and $w_{i,c,j}^{(k+1)}$ evaluated after the $\{\mathbf{l}_c^i\}$-update, we are faced with the following weighted least squares problem:

$$\min_z \sum_{i,c,j} w_{i,c,j}^{(k+1)} |r_{i,c,j}(\theta_j^{(k+1)}, \rho_{c,j}^{(k+1)}, \mathbf{l}_c^{i,(k+1)}, z)|^2, \tag{31}$$

where the dependence of $r_{i,c,j}$ on z is still nonlinear. Therefore, we further linearize $r_{i,c,j}$ with respect to z and arrive at the following update:

$$\begin{aligned}z^{(k+1)} = \arg\min_z \sum_{i,c,j} w_{i,c,j}^{(k+1)} \cdot \\ |r_{i,c,j}^{(k+1)} + J_r(z^{(k)})(z - z^{(k)})|^2,\end{aligned} \tag{32}$$

where $J_r(z^{(k)})$ is the Jacobian of the map $z \mapsto r_{i,c,j}(\theta_j^{(k+1)}, \rho_{c,j}^{(k+1)}, \mathbf{l}_c^{i,(k+1)}, z)$ at $z = z^{(k)}$. The resulting linearized least-squares problem is again solved by CG. In our experiments, we additionally incorporate backtracking line search in the z-update to ensure a monotonic decrease of the energy.

6. Experimental Validation

This section is concerned with the evaluation of the proposed nonconvex variational approach to uncalibrated photometric stereo under general lighting.

6.1. Synthetic Experiments

To validate the impact of the initial volume V in (16), the tunable hyper-parameter μ, and the number of input images M in (13), we consider 36 challenging synthetic datasets. We use four different depth maps ("Joyful Yell" [1], "Lucy" [30], "Armadillo" [30] and "Thai Statue" [30]) and nine different albedo maps and each of those 36 combinations is rendered as described in (1) using $M = 25$ different environment maps[2], cf. Figure 4. The resulting 25 RGB images per dataset are used as input, along with the intrinsic camera parameters and a binary mask Ω. A quantitative evaluation on the triplet (V, μ, M) is carried out on four randomly chosen datasets (Armadillo & White albedo, Joyful Yell & Ebsd albedo, Lucy & Hippie albedo, and Thai Statue & Voronoi albedo), comparing the impact of each value of (V, μ, M) on the resulting mean angular error (MAE) between ground truth and estimated normals.

First, we validate the choice of the input volume V using the initially fixed values of $\mu = 2 \cdot 10^{-6}$ and $M = 25$. As the volume depends on the size of the mask, we consider a linear parametrization $V(\kappa) = \kappa|\Omega| = \kappa N$ and evaluate a range of ratios $\kappa \in [1, 10^3]$. Figure 5 (left) indicates that the optimal value of κ is dataset-dependent. For synthetic datasets we always selected this optimal value, yet for real-world data no such evaluation is possible and κ must be tuned manually. Since the ballooning-based depth initialization can be carried out in real-time (implementation is parallelized in CUDA), the user has an immediate feedback on the initial depth and thus a plausible initial shape is easily drawn. Humans excel at estimating size and shape of objects [5] and real-world experiments will show that a manual choice of κ can result in appealing geometries.

Next, we evaluate the impact of μ, cf. Figure 5 (right). As can be seen, the depth estimate seems to deteriorate for too small and too large values of μ, whereas $\mu \in [10^{-6}, 10^{-5}]$ seems to provide good depth estimates across all albedo maps. Therefore we fix $\mu = 2 \cdot 10^{-6}$ for all our upcoming experimental evaluation.

Unsurprisingly, the MAE is inversely proportional to the number M of input images, but runtime increases (linearly) with M, cf. Figure 6. We found that $M \in [15, 25]$ represents a good trade-off between runtime and accuracy, and fix $M = 20$ for all our further experiments. Our Matlab implementation needs about 1–2 minutes on a computer with an Intel $i7$ processor.

[2]Environment maps are downloaded from http://www.hdrlabs.com/sibl/archive.html

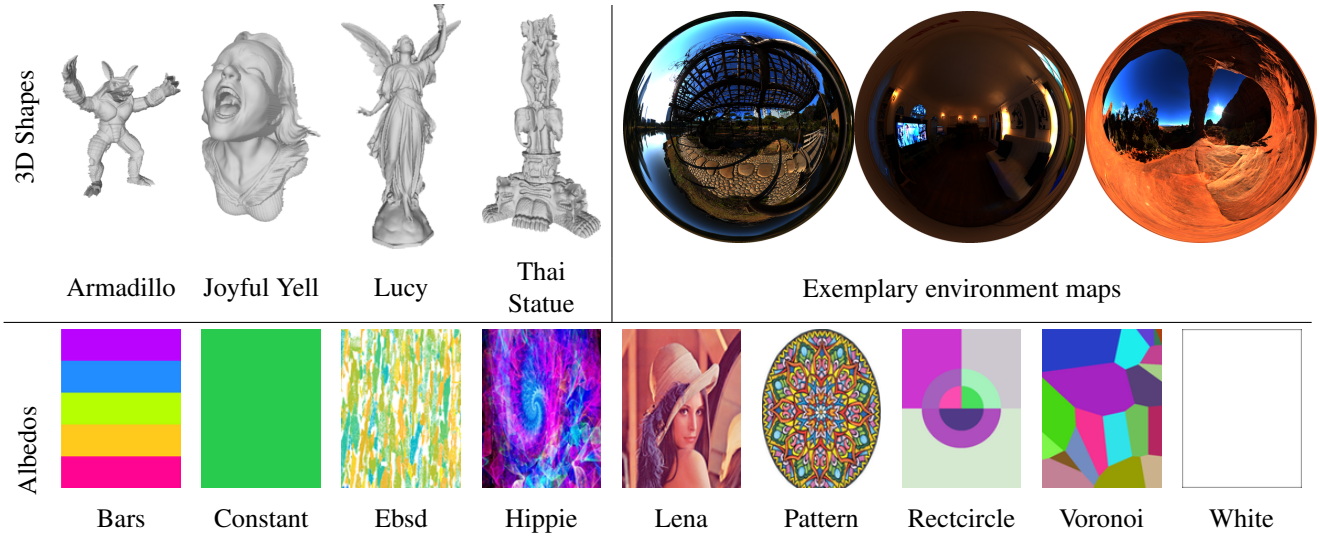


Figure 4. The four 3D-shapes and nine albedo maps we used to create 36 (3D-shape, albedo) datasets. For each dataset, $M = 25$ images were rendered using different environment maps such as those shown on the top right.

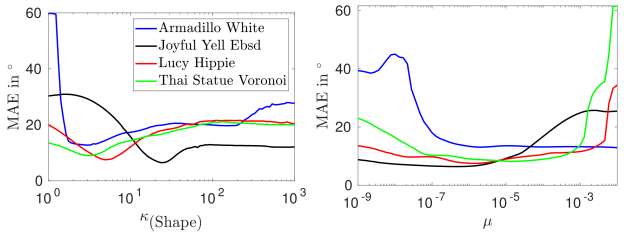


Figure 5. Impact of the initial volume $V_{\text{(Shape)}}$ as well as μ on the accuracy of the estimated depth. Based on these experiments we choose $\kappa_{\text{(Armadillo)}} = 2.84$, $\kappa_{\text{(Joyful Yell)}} = 24.77$, $\kappa_{\text{(Lucy)}} = 4.98$, $\kappa_{\text{(Thai Statue)}} = 3.05$ and $\mu = 2 \cdot 10^{-6}$ for all experiments, where $V_{\text{(Shape)}} = \kappa_{\text{(Shape)}} N_{\text{(Shape)}}$.

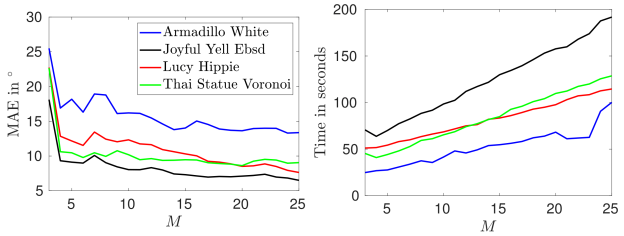


Figure 6. Impact of the number of images M on the mean angular error (MAE) and the runtime. Based on these insights we choose $M = 20$ for our experiments.

Having fixed the choice of (V, μ, M), we can now evaluate our approach against other state-of-the-art methods. We compare our results against those obtained by an uncalibrated photometric stereo approach assuming directional lighting [15], and another one assuming general (first-order spherical harmonics) illumination yet relying on an input shape prior (e.g., from an RGB-D sensor) [40]. As this limiting assumption on the access to a sensor-based depth prior is not always given and to make comparison fair, we input as depth prior to this method the ballooning initialization described in Section 5.1. Furthermore, we compare against another uncalibrated photometric stereo work under natural illumination [34][3], which resorts to the equivalent directional lighting instead of spherical harmonics, cf. Section 3. Table 1 shows the median and mean MAEs over all 36 datasets (a more detailed table can be found in the supplementary material). On these datasets, it can be seen that our method quantitatively outperforms the current state-of-the-art by a factor of 2–3. This gain is also evaluated qualitatively in Figure 7, which shows a selection of two results.

Approach	[15]	[40]	[34]	Ours
Median	27.16	21.14	34.06	**9.17**
Mean	34.15	21.18	35.53	**10.72**

Table 1. Median and mean of the mean angular errors (MAE) over all 36 datasets. The proposed approach overcomes the state-of-the-art by a factor of 2–3.

6.2. Real-World Experiments

For real-world data we use the publicly available dataset of [19]. It offers eight challenging real-world datasets of objects with complex geometry and albedo captured under daylight and a freely moving LED, along with intrinsics matrix K and masks Ω. Results are presented in Figure 8. Despite relying on a directional lighting model, the approach of [15] produces reasonable results on some datasets (Face1, Ovenmitt or Shirt), but it fails on others. As [40] assumes a reliable prior on depth in order to perform a photometric refinement, this approach is biased towards its initialization and thus, only when the depth prior is very close to

[3] Code associated with [15] and [40] can be found online, and the results obtained by [34] were provided by the authors.

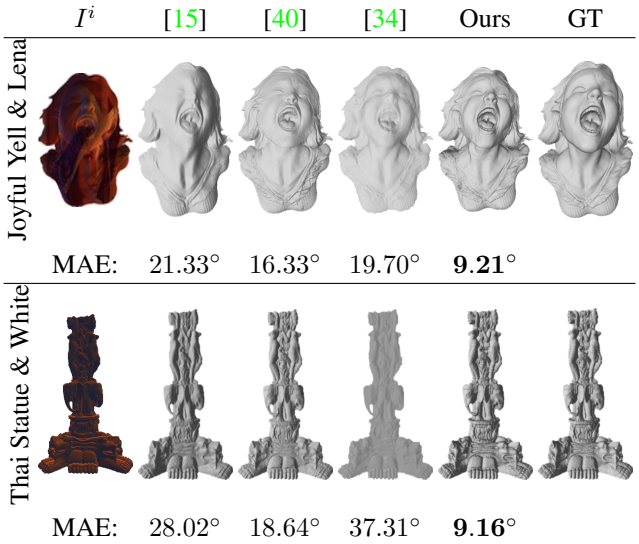


Figure 7. Results of state-of-the-art approaches and our approach on two out of the 36 synthetic datasets. Numbers show the mean angular error (MAE) in degrees.

the objects' rough shape (Ovenmitt, Shirt, Tabletcase, Vase) a meaningful geometry is recovered. The approach of [34] estimates a possibly non-integrable normal field only, and it can be seen that after integration the depth map might not be satisfactory. As our approach optimizes over depth directly, such issues are not apparent and we are able to recover fine-scale geometric details throughout all tests.

7. Conclusion

We proposed a variational approach to uncalibrated photometric stereo (PS) under general lighting. Assuming a perspective camera setup, our method jointly estimates shape, reflectance and lighting in a robust manner. The possible non-integrability of normals is bypassed by the direct estimation of the underlying depth map, and robustness is ensured by resorting to Cauchy's M-estimator and Huber-TV albedo regularization. Although the problem is nonconvex and thus numerically challenging and initialization-dependent, we tackled it efficiently through a tailored lagged block coordinate descent algorithm and ballooning-based depth initialization. Over a series of evaluations on synthetic and real data, we demonstrated that our method outperforms existing methods in terms of MAE by a factor of 2–3 and provides highly detailed reconstructions even in challenging real-world settings.

In future research, a more automated balloon-like depth initialization is desirable. Exploring the theoretical foundations (uniqueness of a solution) of differential perspective uncalibrated PS under spherical harmonic lighting and analyzing the convergence properties of the proposed numerical scheme constitute two other promising perspectives.

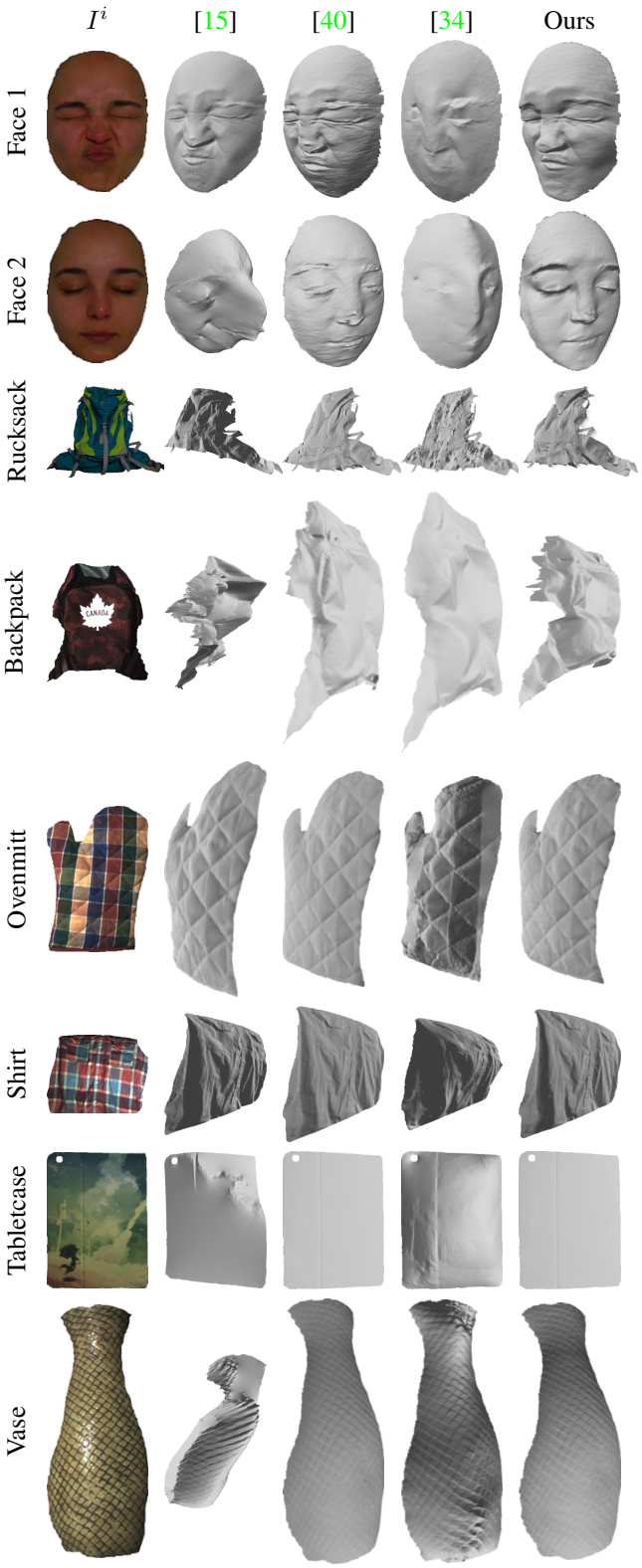


Figure 8. Results of state-of-the-art approaches and our approach on challenging real-world datasets. While the competing approaches fail on some datasets, our approach consistently yields satisfactory results.

References

[1] The Joyful Yell. 2015. https://www.thingiverse.com/thing:897412. 6

[2] A. Abrams, C. Hawley, and R. Pless. Heliometric Stereo: Shape from Sun Position. In *European Conference on Computer Vision (ECCV)*, volume 7573 of *Lecture Notes in Computer Science*, pages 357–370, 2012. 2

[3] J. Ackermann, F. Langguth, S. Fuhrmann, and M. Goesele. Photometric stereo for outdoor webcams. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 262–269, 2012. 2

[4] J. Ackermann, M. Ritz, A. Stork, and M. Goesele. Removing the example from example-based photometric stereo. In *Trends and Topics in Computer Vision (ECCV Workshops)*, volume 6554 of *Lecture Notes in Computer Science*, pages 197–210. 2012. 2

[5] J. Baldwin, A. Burleigh, R. Pepperell, and N. Ruta. The perceived size and shape of objects in peripheral vision. *i-Perception*, 7(4):2041669516661900, 2016. 6

[6] A. Bartoli, Y. Gérard, F. Chadebecq, T. Collins, and D. Pizarro. Shape-from-template. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(10):2099–2118, 2015. 2

[7] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision*, 72(3):239–257, 2007. 2, 3

[8] R. Basri and D. W. Jacobs. Lambertian reflectances and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):218–233, 2003. 3

[9] P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille. The bas-relief ambiguity. *International Journal of Computer Vision*, 35(1):33–44, 1999. 2, 3

[10] M. Brady and A. Yuille. An extremum principle for shape from contour. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(3):288–301, 1984. 2

[11] M. Chandraker, J. Bai, and R. Ramamoorthi. On Differential Photometric Reconstruction for Unknown, Isotropic BRDFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12):2941–2955, Dec 2013. 2

[12] T. Collins and A. Bartoli. 3D Reconstruction in Laparoscopy with Close-Range Photometric Stereo. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 7511 of *Lecture Notes in Computer Science*, pages 634–642. 2012. 2

[13] J.-D. Durou, M. Falcone, and M. Sagona. Numerical Methods for Shape-from-shading: A New Survey with Benchmarks. *Computer Vision and Image Understanding*, 109(1):22–43, 2008. 2

[14] A. R. Farooq, M. L. Smith, L. N. Smith, and S. Midha. Dynamic photometric stereo for on line quality control of ceramic tiles. *Computers in industry*, 56(8-9):918–934, 2005. 2

[15] P. Favaro and T. Papadhimitri. A closed-form solution to uncalibrated photometric stereo via diffuse maxima. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 821–828, 2012. 7, 8, 11, 14, 15

[16] J. Geng. Structured-light 3D surface imaging: a tutorial. *Advances in Optics and Photonics*, 3(2):128–160, 2011. 2

[17] P. F. Gotardo, T. Simon, Y. Sheikh, and I. Matthews. Photogeometric scene flow for high-detail dynamic 3d reconstruction. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 846–854, 2015. 2

[18] G. Graber, J. Balzer, S. Soatto, and T. Pock. Efficient minimal-surface regularization of perspective depth maps in variational stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 511–520, 2015. 5

[19] B. Haefner, S. Peng, A. Verma, Y. Quéau, and D. Cremers. Photometric depth super-resolution. *Arxiv preprint 1809.10097*, 2018. 7

[20] H. Hayakawa. Photometric stereo under a light source with arbitrary motion. *Journal of the Optical Society of America A*, 11(11):3079–3089, 1994. 2, 3

[21] C. Hernández. *Stereo and Silhouette Fusion for 3D Object Modeling from Uncalibrated Images Under Circular Motion*. Thèse de doctorat, École Nationale Supérieure des Télécommunications, 2004. 2

[22] Y. Hold-Geoffroy, P. F. Gotardo, and J.-F. Lalonde. Deep photometric stereo on a sunny day. *Arxiv preprint 1803.10850*, 2018. 2

[23] Y. Hold-Geoffroy, J. Zhang, P. F. Gotardo, and J.-F. Lalonde. What is a good day for outdoor photometric stereo? In *International Conference on Computational Photography*, 2015. 2

[24] Y. Hold-Geoffroy, J. Zhang, P. F. Gotardo, and J.-F. Lalonde. x-hour outdoor photometric stereo. In *International Conference on 3D Vision*, 2015. 3

[25] B. K. Horn. *Shape From Shading: A Method for Obtaining the Shape of a Smooth Opaque Object From One View*. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1970. 2

[26] J. Jung, J.-Y. Lee, and I. S. Kweon. One-Day Outdoor Photometric Stereo Using Skylight Estimation. *International Journal of Computer Vision*, 2019. (to appear). 2

[27] I. Kemelmacher-Shlizerman and R. Basri. 3d face reconstruction from a single image using a single reference face shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(2):394–405, 2010. 2

[28] R. Kozera. On Shape Recovery from Two Shading Patterns. *International Journal of Pattern Recognition and Artificial Intelligence*, 6(4):673–698, 1993. 2

[29] R. Kozera and A. Prokopenya. Second-order algebraic surfaces and two image photometric stereo. In *International Conference on Computer Vision and Graphics*, pages 234–247, 2018. 2

[30] M. Levoy, J. Gerth, B. Curless, and K. Pull. The stanford 3d scanning repository. 2005. http://www-graphics.stanford.edu/data/3dscanrep. 6

[31] F. Logothetis, R. Mecca, and R. Cipolla. Semi-calibrated near field photometric stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 941–950, 2017. 2

[32] R. Mecca, A. Tankus, A. Wetzler, and A. M. Bruckstein. A direct differential approach to photometric stereo with perspective viewing. *SIAM Journal on Imaging Sciences*, 7(2):579–612, 2014. 2

[33] R. Mecca, A. Wetzler, A. M. Bruckstein, and R. Kimmel. Near field photometric stereo with point light sources. *SIAM Journal on Imaging Sciences*, 7(4):2732–2770, 2014. 2

[34] Z. Mo, B. Shi, F. Lu, S.-K. Yeung, and Y. Matsushita. Uncalibrated photometric stereo under natural illumination. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2936–2945, 2018. 2, 3, 7, 8, 11, 14, 15, 17

[35] T. Moons, L. Van Gool, and M. Vergauwen. 3D Reconstruction from Multiple Images. *Foundations and Trends in Computer Graphics and Vision*, 4(4):287–404, 2008. 2

[36] S. N. Y. Nakagawa and S. Nayar. Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):824–831, 1994. 2

[37] R. Onn and A. Bruckstein. Integrability disambiguates surface recovery in two-image photometric stereo. *International Journal of Computer Vision*, 5(1):105–113, 1990. 2

[38] M. R. Oswald, E. Toeppe, and D. Cremers. Fast and Globally Optimal Single View Reconstruction of Curved Objects. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 534–541, 2012. 4

[39] T. Papadhimitri and P. Favaro. A new perspective on uncalibrated photometric stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1474–1481, 2013. 3

[40] S. Peng, B. Haefner, Y. Quéau, and D. Cremers. Depth super-resolution meets uncalibrated photometric stereo. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, pages 2961–2968, 2017. 2, 7, 8, 11, 14, 15

[41] Y. Quéau, J.-D. Durou, and J.-F. Aujol. Normal Integration: A Survey. *Journal of Mathematical Imaging and Vision*, 60(4):576–593, 2018. 2, 4, 5

[42] Y. Quéau, T. Wu, F. Lauze, J.-D. Durou, and D. Cremers. A Non-Convex Variational Approach to Photometric Stereo under Inaccurate Lighting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 350–359, 2017. 2, 4

[43] Y. Sato and K. Ikeuchi. Reflectance analysis under solar illumination. In *Workshop on Physics-Based Modeling in Computer Vision (ICCV Workshops)*, pages 180–187, 1995. 2

[44] S. A. Shafer and T. Kanade. Using shadows in finding surface orientations. *Computer Vision, Graphics, and Image Processing*, 22(1):145–176, 1983. 2

[45] F. Shen, K. Sunkavalli, N. Bonneel, S. Rusinkiewicz, H. Pfister, and X. Tong. Time-lapse photometric stereo and applications. *Computer Graphics Forum*, 33(7):359–367, 2014. 2

[46] L. Shen and P. Tan. Photometric stereo and weather estimation using internet images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1850–1857, 2009. 2

[47] B. Shi, K. Inose, Y. Matsushita, P. Tan, S.-K. Yeung, and K. Ikeuchi. Photometric stereo using internet images. In *International Conference on 3D Vision*, volume 1, pages 361–368, 2014. 2

[48] B. Shi, Z. Wu, Z. Mo, D. Duan, S.-K. Yeung, and P. Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2):271–284, 2019. 2

[49] A. P. Witkin. Recovering surface shape and orientation from texture. *Artificial Intelligence*, 17(1):17–45, 1981. 2

[50] R. J. Woodham. Reflectance Map Techniques for Analyzing Surface Defects in Metal Castings. Technical Report MIT AITR-457, 1978. 2

[51] J. Yang, N. Ohnishi, and N. Sugie. Two image photometric stereo method. In *Intelligent Robots and Computer Vision XI: Biological, Neural Net, and 3D Methods*, volume 1826 of *Proceedings of the International Society for Optical Engineering*, pages 452–463, 1992. 2

A. Further Details on Synthetic Experiments

To provide further insights on the synthetic experiments (in Section 6.1), we visualize the environment lightings ℓ^i, $i = 1 \dots 25$, used to render each image. Figure 9 shows all 25 environment maps[4]. The impact of each incident lighting ℓ^i, $i = 1 \dots 25$, is illustrated in Figure 10 showing the Joyful Yell with a White ($\rho \equiv 1$) albedo. Thus, color changes in the images are caused by lighting only, as depicted in model (1) and (7) in the main paper.

Table 2 shows the mean angular error (MAE) of each dataset on the state-of-the-art approaches [15, 34, 40] and our proposed methodology. It can be seen that our approach consistently overcomes [15, 34, 40] by a factor of 2–3. Only the Pattern albedo seems to bias the resulting depth negatively, yet even in this case our approach estimates the geometry more faithfully than the current state-of-the-art.

Two more qualitative results on synthetic data are shown in Figure 11. While [15] gives more meaningful results on Armadillo with Constant albedo, depth deteriorates strongly on Lucy with Hippie albedo. Methods of [34, 40] both result in rather flattened shapes (cf. Lucy). Most accurate results are achieved using the proposed method where fine geometric details, as well as non flattened depth estimates are shown.

Additional to the depth results, Figure 12 shows estimated lightings and albedos along with the ground truths. Although lighting estimates show less shadowed areas and seem brighter compared to ground truths, this does not seem to affect reflectance estimations much. The estimated albedos are satisfactory, although some shading information is slightly visible.

The initialization is indeed crucial for the whole algorithm. Here, we show two different non-trivial initializations for our algorithm in Table 2: 1) Hemisphere, we first compute the circumscribed sphere for the 3D points of ground truth. The projection of each point onto this sphere is considered as initialization; 2) Initialization by [34], we simply refine the result from [34] by our algorithm. In Figure 13, we show visualized results. In certain special cases, the initialization from [34] is slightly better. However, our minimal surface strategy is stable for all cases, and our algorithm improves the results from [34]) in most cases.

B. Further Details on Real-World Results

Supplementary to the real-world experiments (in Section 6.2), Figures 14 and 15 show alternative viewpoints of the real-world results. The estimated albedos, which are mapped onto the surfaces, appear satisfactory. Correspondingly, we also show the estimated albedos and lightings. In view of the multiplicative ambiguity between lightings and albedos, all visualized albedos are normalized to have maximum value 1.

[4] All environment maps were downloaded from http://www.hdrlabs.com/sibl/archive.html

Figure 9. All environment maps ℓ^i (360° view) used throughout the synthetic evaluation.

Figure 10. Illustration of the input data. The Joyful Yell dataset with White albedo to show the impact of the different environment maps used throughout the synthetic experimental validation.

Dataset		[15]	[40]	[34]	Our approach with different initializations		
Shape	Albedo				Hemisphere	Using [34]	Minimal surface (Sec. 5.1)
Armadillo	Bars	26.22	27.84	36.91	79.54	20.08	**16.78**
	Constant	25.84	26.64	36.87	83.01	18.81	**13.97**
	Ebsd	25.34	26.88	27.80	82.53	15.99	**14.26**
	Hippie	28.21	27.30	25.82	79.12	**12.56**	14.52
	Lena	27.07	27.33	28.36	84.24	17.79	**14.78**
	Pattern	45.87	26.82	24.01	82.59	19.39	**19.06**
	Rectcircle	26.97	26.71	36.23	80.68	19.64	**14.06**
	Voronoi	25.62	26.91	50.70	79.65	55.29	**14.07**
	White	26.19	26.64	52.04	83.04	56.74	**14.13**
Joyful Yell	Bars	21.84	16.26	31.80	21.21	28.82	**8.69**
	Constant	23.95	14.93	33.47	16.85	29.31	**5.96**
	Ebsd	26.08	15.63	15.91	17.63	7.49	**7.28**
	Hippie	28.67	16.23	22.96	17.68	**7.47**	7.49
	Lena	21.33	16.33	19.70	20.11	13.16	**9.21**
	Pattern	26.07	18.76	26.67	18.76	21.03	**16.97**
	Rectcircle	35.27	15.19	52.41	16.27	61.77	**7.34**
	Voronoi	22.27	16.42	45.74	18.62	54.78	**6.57**
	White	27.12	14.32	33.06	17.70	28.99	**6.20**
Lucy	Bars	49.13	21.90	36.51	40.55	26.15	**8.16**
	Constant	54.98	19.89	36.57	41.00	25.74	**8.71**
	Ebsd	62.33	20.81	23.56	40.80	13.36	**9.61**
	Hippie	58.61	21.29	32.38	39.93	8.10	**7.87**
	Lena	64.01	22.24	30.93	40.16	19.14	**9.56**
	Pattern	48.83	22.25	32.68	40.11	20.56	**17.78**
	Rectcircle	24.68	20.99	43.13	41.17	10.01	**8.98**
	Voronoi	61.53	22.10	48.14	40.39	71.32	**7.59**
	White	64.43	19.33	44.76	41.54	72.45	**8.76**
Thai Statue	Bars	25.53	21.91	66.17	78.72	8.94	**8.55**
	Constant	27.20	18.91	38.47	81.14	24.26	**9.58**
	Ebsd	27.85	20.22	34.11	79.58	19.23	**9.47**
	Hippie	21.91	21.86	30.62	77.27	12.78	**8.83**
	Lena	33.53	19.66	34.00	79.43	19.55	**9.19**
	Pattern	26.77	22.06	28.81	83.92	16.69	**15.27**
	Rectcircle	29.36	19.92	43.86	81.88	79.88	**8.84**
	Voronoi	30.65	21.56	36.58	78.92	25.21	**8.69**
	White	28.02	18.64	37.31	81.54	24.94	**9.16**
Median		27.16	21.14	34.06	59.41	19.86	**9.17**
Mean		34.15	21.18	35.53	55.20	27.43	**10.72**

Table 2. Quantitative comparison between our method and other state-of-the-art methods on challenging synthetic datasets. The last three columns refer to the results with different initializations for our approach.

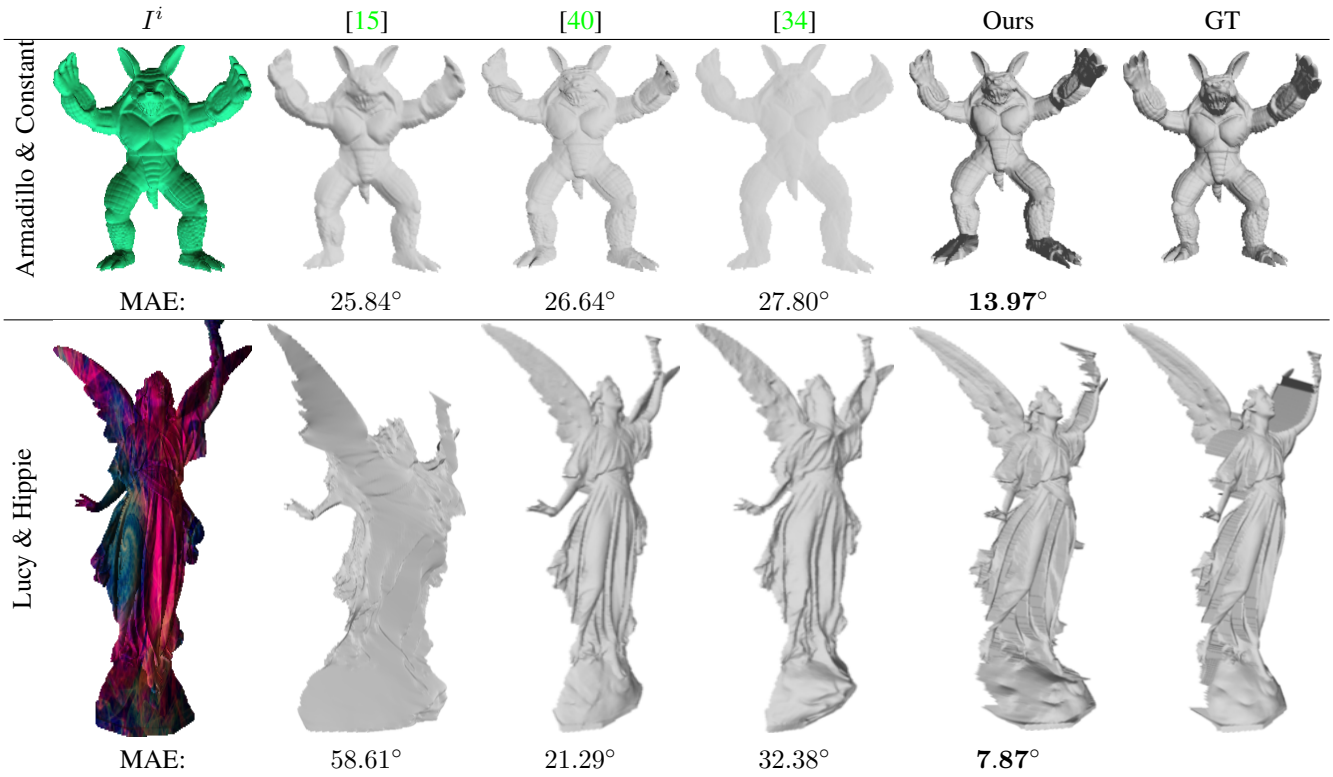


Figure 11. Results of state-of-the-art approaches and our approach on two out of the 36 synthetic datasets. Numbers show the mean angular error (MAE) in degrees.

Figure 12. Our estimated albedos and lighting next to the ground truth. Lighting estimates show less shadowed areas and seem brighter compared to ground truth, yet this does not seem to affect reflectance and geometry estimation much, cf. Figure 7 in main paper and Figure 11 in the supplementary material. The estimated albedos are satisfactory, although some shading information is slightly visible.

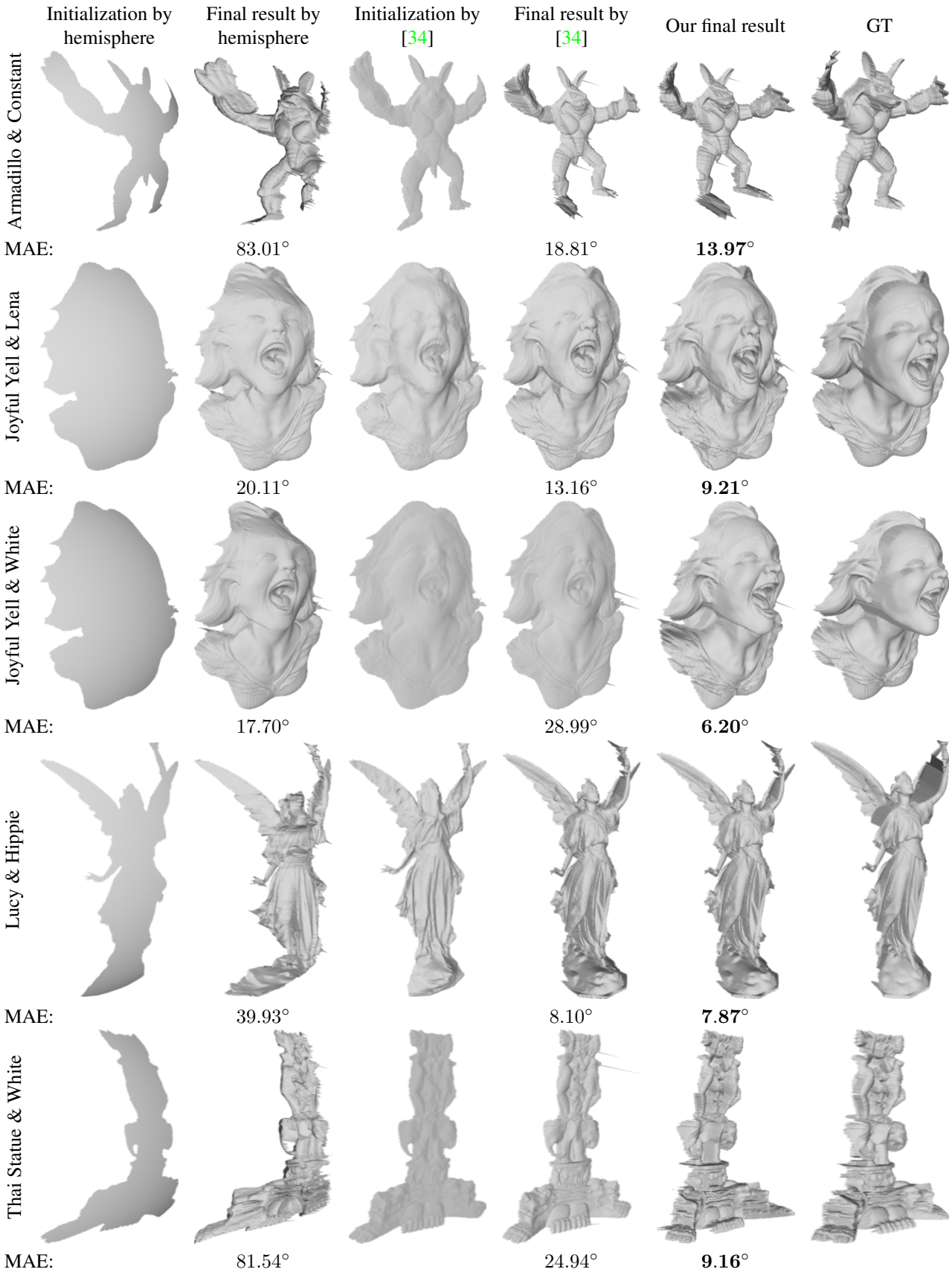


Figure 13. Our results compared those from two different initializations of our algorithm. Numbers show the mean angular error (MAE) in degrees. Though the initialization by [34] achieves comparable result to ground truth on "Lucy & Hippie" dataset, its performance is not stable across different datasets.

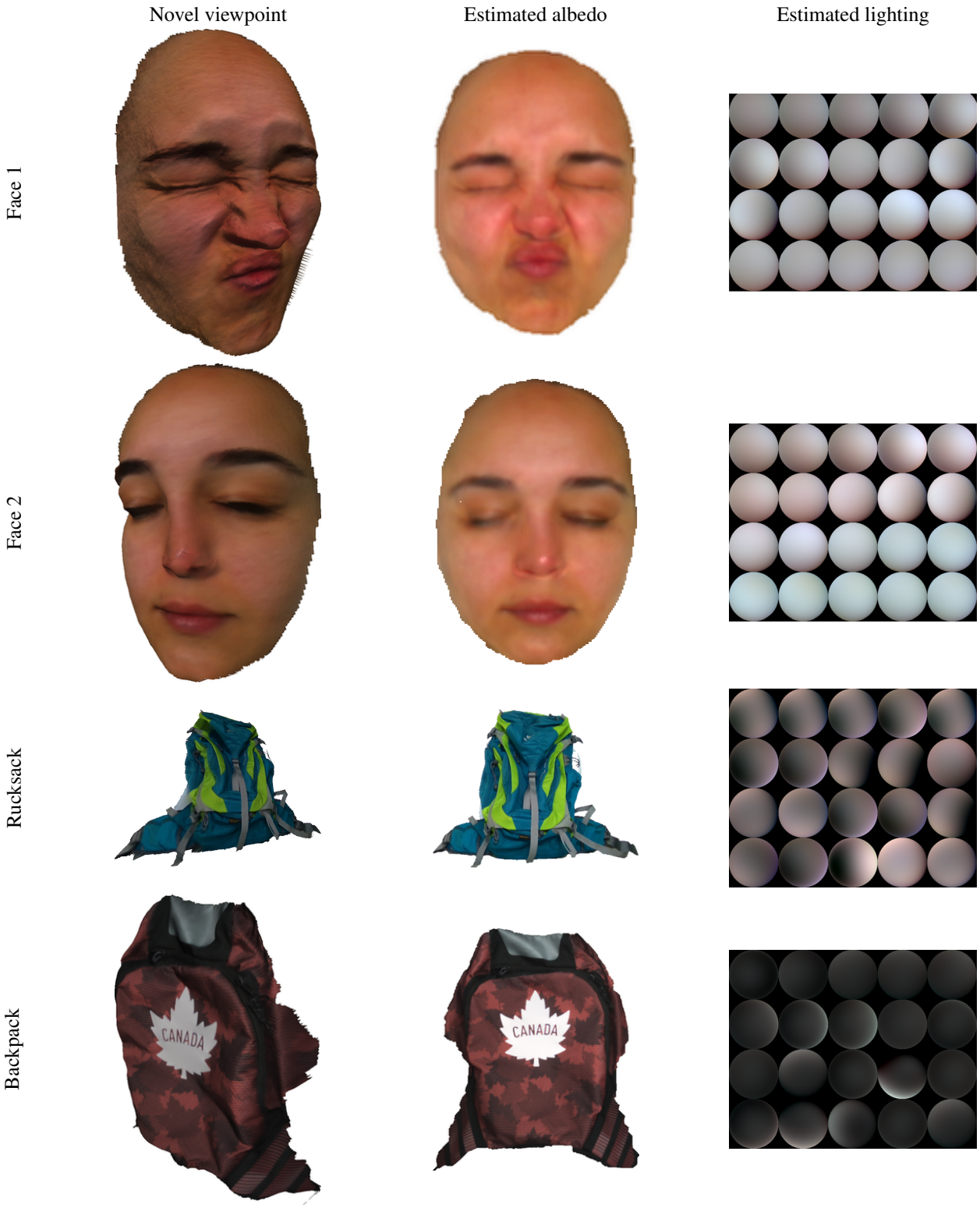


Figure 14. Real-world results: (left) estimated albedos mapped onto estimated surfaces rendered under a novel viewpoint, (middle) estimated albedos, (right) estimated lightings for all $M = 20$ input images.

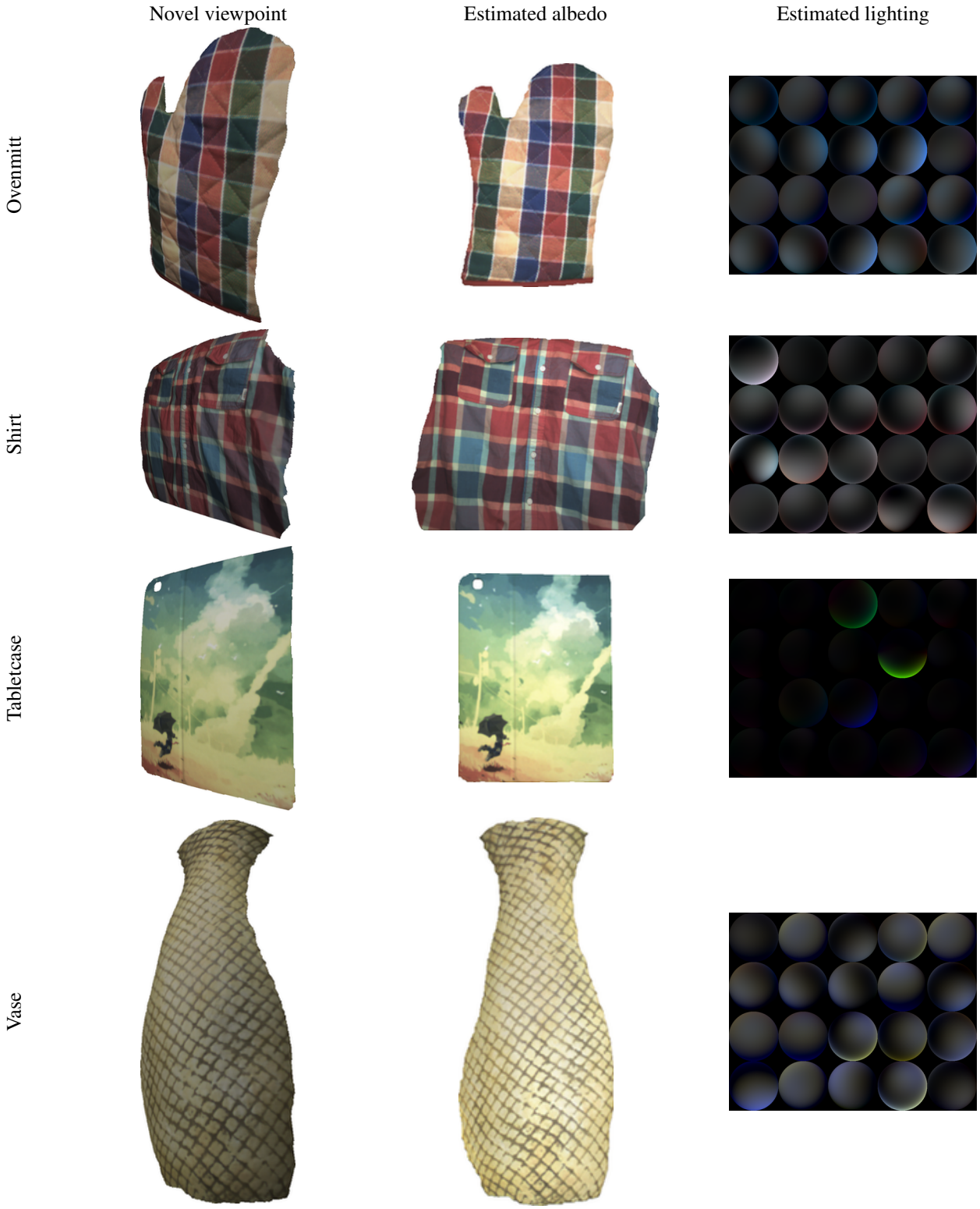


Figure 15. More real-world results: (left) estimated albedos mapped onto estimated surfaces rendered under a novel viewpoint, (middle) estimated albedos, (right) estimated lightings for all $M = 20$ input images.