Abstract

Photometric stereo (PS) techniques nowadays remain constrained to an ideal laboratory setup where modeling and calibration of lighting is amenable. This work aims to eliminate such restrictions. To this end, we introduce an efficient principled variational approach to uncalibrated PS under general illumination, which is approximated through a second-order spherical harmonic expansion. The joint recovery of shape, reflectance and illumination is formulated as a variational problem where 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. We provide a tailored numerical scheme to solve the resulting non-convex problem efficiently and robustly. On a variety of evaluations, our method consistently reduces the mean angular error 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 photometric stereo applications to laboratory setups where careful calibration of lighting must be carried out prior to any 3D-reconstruction.

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. 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.

After reviewing related works in Section 2, we discuss in Section 3 the spherical harmonics-based image formation model considered in this work. It can be inverted using the variational approach to the joint recovery of depth, albedo and lighting which is discussed 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.
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 photographs, 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 photographs 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 [41], contours [10], texture [46], template [6], structured light [16], motion [32], focus [33], silhouettes [19], 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. [23], 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 [47], 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 [28, 30]. 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 [18], 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 [45] for a recent survey. A natural way to enforce integrability consists in following a differential approach to photometric stereo [11, 29] 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 [39]. In addition to the theoretical interest of enforcing integrability in order to limit ambiguities, differential approaches to photometric stereo have the advantage of bypassing the problem of integrating the estimated normal field, which is by itself a non-trivial problem [38]. 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 [45] 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 [40]. 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 [42] (see also [21] 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 [25, 34, 48], although it is still an active research area [26]. 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 [20]. Another one is to consider a non-directional illumination model to represent natural illumination, as for instance in [24]. 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 [21].

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 [43] or relying on the prior knowledge of a rough geometry [4, 37, 44]. Uncalibrated photometric stereo under natural illumination has been revisited recently in [31], 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 [8], 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 \( \{I^c_i\}_{i=1}^M \), each \( I^c : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}^C \) representing a multi-channel image (i.e., \( C \geq 1 \)) over a masked pixel domain \( \Omega \). Assuming that the object being pictured is Lambertian, the surface’s reflectance is represented by the albedo \( \rho \), and the general image formation model is as follows, for all \( i \in \{1, ..., M\} \), \( c \in \{1, ..., C\} \), and \( p \in \Omega \):

\[
I^c_i(p) = \int_{\mathbb{S}^2} \rho_c(p) \ell^c_c(\omega) \max\{\omega \cdot n(p), 0\} \, d\omega. \tag{1}
\]

Here \( \mathbb{S}^2 \) is the unit sphere in \( \mathbb{R}^3 \), \( \ell^c_c : \mathbb{S}^2 \rightarrow \mathbb{R}_+ \) represents the channel-wise intensity of the incident light, and \( \rho_c(p) \in \mathbb{R}_+ \) and \( n(p) \in \mathbb{S}^2 \) are the channel-wise albedos and the unit-length surface normals, respectively, at the surface point conjugate to pixel \( p \in \Omega \). 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 \( \omega \). In the setup of uncalibrated PS, the quantities \( \{\ell^c_c\}, \{\rho_c\} \), in addition to \( n \), are unknown.

Equivalent directional lighting \([22]\) approximates (1) via

\[
I^c_i(p) = \rho_c(p) \tilde{\ell}^c_i(p) \cdot n(p),
\]

\[
\tilde{\ell}^c_i(p) := \int_{\{\omega \in \mathbb{S}^2 : n(p) \geq 0\}} \ell^c(\omega) \omega \, d\omega.
\tag{2}
\]

where \( \tilde{\ell}^c_i(p) \) represents the mean lighting over the visible hemisphere at \( p \). The field \( \tilde{\ell}^c_i \) 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 \([18]\). State-of-the-art patch-wise methods \([31]\) 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 \([7, 8]\). By defining the half-cosine kernel \( k \) as

\[
k(\omega, n) := \max\{\omega \cdot n, 0\},
\tag{3}
\]

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

\[
I^c_i(p) = \rho_c(p) \int_{\mathbb{S}^2} k(\omega, n(p)) \ell^c_c(\omega) \, d\omega. \tag{4}
\]

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

\[
\int_{\mathbb{S}^2} k(\omega, n(p)) \ell^c_c(\omega) \, d\omega = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} (k_n \ell^{c,n}_n) h_{n,m}(n(p)). \tag{5}
\]

Here the spherical harmonics \( \{h_{n,m}\} \) form an orthonormal basis on \( \mathbb{S}^2 \), and \( \{k_n\} \) and \( \{\ell^{c,n}_n\} \) are the expansion coefficients of \( k \) and \( \ell^c_i \) with respect to \( \{h_{n,m}\} \). Since most energy in the expansion (5) concentrates on low-order terms \([7]\), we obtain the second-order SHA by truncating the series up to the first nine terms (i.e., \( 0 \leq n \leq 2 \)):

\[
\int_{\mathbb{S}^2} k(\omega, n(p)) \ell^c_c(\omega) \, d\omega \approx \sum_{n=0}^{2} \sum_{m=-n}^{n} (k_n \ell^{c,n}_n) h_{n,m}(n(p)). \tag{6}
\]

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 \([7]\) 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 \([7]\) into (4), we finalize our image formation model as:

\[
I^c_i(p) \approx \rho_c(p) \ell^c_i \cdot h[n](p), \tag{7}
\]

\[
h[n] := [1, n_1, n_2, n_3, n_1n_2, n_1n_3, n_2n_3, n_1^2 - n_2^2, 3n_3^2 - 1]^T. \tag{8}
\]

Here \( h[n] : \Omega \rightarrow \mathbb{R}^9 \) represents the second-order harmonic images, and \( \ell^c_i \in \mathbb{R}^9 \) represents the harmonic lighting vector whose entries have absorbed \( \{k_n \ell^{c,n}_n\} \) 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 \( \{\ell^c_i\} \), which yields a less ill-posed inverse problem \([8]\). 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 \([36]\) 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 \(\Omega\). Under perspective projection, the surface geometry is modeled as a map \(x : p = (u, v) \in \Omega \mapsto x(u, v) \in \mathbb{R}^3\) given by

\[
x(u, v) = z(u, v)K^{-1}[u, v, 1]^T,
\]

with \(z : \Omega \mapsto \mathbb{R}_+\) the depth map and

\[
K := \begin{bmatrix} f_u & 0 & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}
\]

the calibrated camera’s intrinsic matrix. In the following we denote for convenience \((\tilde{u}, \tilde{v}) := (u - u_0, v - v_0)\).

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

\[
n[z](u, v) = \frac{\tilde{n}[z](u, v)}{|\tilde{n}[z](u, v)|},
\]

\[
\tilde{n}[z](u, v) := \begin{bmatrix} f_u \partial_u z(u, v) \\ f_v \partial_v z(u, v) \\ -z(u, v) - \tilde{u} \partial_u z(u, v) - \tilde{v} \partial_v z(u, v) \end{bmatrix}.
\]

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

\[
|s|_\gamma := \begin{cases} |s|^2/(2\gamma) & \text{if } |s| \leq \gamma, \\ |s| - \gamma/2 & \text{if } |s| > \gamma, \end{cases}
\]

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).

Note that the dependence of \(\tilde{n}[z]\) 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:

\[
\min_{\{\rho_c\}, \{l_i^c\}, z} \sum_{i=1}^M \sum_{c=1}^C \int_{\Omega} \phi_\lambda \left( \rho_c(u, v) l_i^c \cdot h[n[z]](u, v) \right) du dv + \mu \sum_{c=1}^C \int_{\Omega} |\nabla \rho_c(u, v)|_{\gamma}^\gamma du dv.
\]

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

\[
\phi_\lambda(s) = \lambda^2 \log(1 + s^2/\lambda^2),
\]

The first eight iterations we freeze the second-order spherical harmonics coefficients \((l_i^c)_5 = (l_i^c)_6 = \ldots = (l_i^c)_9 = 0\) i.e., we reconstruct using only first-order spherical harmonic approximation as a warm start.

Based on the observation that most real-world scenes are 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 \(n[z]\) and, hence, zero entries in the initial harmonic images \(h[n[z]]\). This would cause non-meaningful updates on albedos \(\{\rho_c\}\) and lighting vectors \((l_i^c)\); cf. Figure 3 for an illustration.

To solve this issue, we specialize the depth initialization which undergoes two phases:
1. Following [35], 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 [38].

Phase 1 is pursued via the following constrained convex optimization:

\[
\min_{z_o} \int_{\Omega} \sqrt{1 + |\nabla z_o|^2} \, du \, dv \\
\text{s.t.} \int_{\Omega} z_o \, du \, dv = V.
\]  

(16)

Intuitively, the model seeks for \( z_o \) which has minimal surface area subject to a constant volume \( V \). A global minimizer of this model can be efficiently computed by simple projected gradient iterations:

\[
z_o^{(k+1/2)} = z_o^{(k)} - \tau \nabla^T \frac{1}{\sqrt{1 + |\nabla z_o^{(k)}|^2}} \nabla z_o^{(k)}, \quad (17)
\]

\[
z_o^{(k+1)} = z_o^{(k+1/2)} + \left( \frac{V - \int_{\Omega} z_o^{(k+1/2)} \, du \, dv}{\int_{\Omega} z_o^{(k+1/2)} \, du \, dv} \right) \cdot 1_{\Omega},
\]

(18)

where \( 1_{\Omega}(u,v) \equiv 1 \) 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 \( \hat{z} \) is represented by

\[
\hat{z}(u,v) = [u, v, z_o(u,v)]^T,
\]

and the corresponding surface normal \( \hat{n} \) to the surface at \( \hat{z} \) conjugate to pixel \( \hat{p} = (u,v) \) is given by

\[
\hat{n}(u,v) = \frac{1}{\sqrt{\nabla z_o(u,v)^2 + 1}} \left[ \nabla z_o(u,v), -1 \right]^T.
\]  

(20)

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

\[
\hat{n}(u,v) \propto \left[ f_u \theta_u \hat{z}_p(u,v) \nu_x \hat{\theta}_u \hat{z}_p(u,v) \right], \quad (21)
\]

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

\[
\nabla \hat{z}_p(u,v) = \frac{-1}{f_u \theta_u \hat{z}_p(u,v) + \nu_x \theta(u,v) + \hat{n}_u(u,v)} \left[ \frac{f_x \hat{n}_u(u,v)}{f_u} \right],
\]

(22)

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

1. \((z_o \to \hat{n})\): Compute \( \hat{n} \) by (20).
2. \((\hat{n} \to \nabla \hat{z}_p)\): Compute \( \nabla \hat{z}_p \) by (22).
3. \((\nabla \hat{z}_p \to z_p)\): Perform integration [38] to obtain \( \hat{z}_p \).

Return \( z_p = \exp \hat{z}_p \) as the initialized (perspective) depth.

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.

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 \( h[n|z] \) and the Cauchy’s M-estimator \( \phi_\lambda \), 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 \( \theta_j = |\hat{n}_j(z)| \). This enables us to rewrite (11) as \( n_j(z) = \hat{n}_j(z)/\theta_j \). Then we formulate the following constrained optimization problem:

\[
\min_{\theta \in \{\theta_j \}_{j=1}^N, \{\gamma\}_{j=1}^N} \sum_{i=1}^M \sum_{c=1}^C \sum_{j=1}^N \phi_\lambda \left( r_{i,c,j}(\theta_j, \rho_c, \lambda, z) \right) + \mu \sum_{c=1}^C \sum_{j=1}^N |(\nabla \rho_c)_j| \gamma, \quad (23)
\]

s.t. \( \theta_j = |\hat{n}_j(z)|, \forall j \in \{1, \ldots, N\} \),

where \( r_{i,c,j} \) is the residual function defined by:

\[
r_{i,c,j}(\theta_j, \rho_c, \lambda, z) = \rho_c \lambda \hat{n}_j(z)/\theta_j - l_c(z). \quad (24)
\]

Upon initialization, the proposed LBCD proceeds as follows. At iteration \( k \), we lag \( \theta \) one iteration behind, i.e.,

\[
\theta_j^{(k+1)} := |\hat{n}_j(z^{(k)})|, \quad \forall j \in \{1, \ldots, N\}, \quad (25)
\]

and then sequentially update each of the three blocks (namely \( \rho_c \), \( l_c \) 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
  \[
  r_i^{(k+1/3)} := r_i^{(k+1/3)}(\theta_j^{(k+1)}, \rho_c^{(k)}, \lambda c_j^{(k)}, z^{(k)}),
  \]
  and then set up the (lagged) weight factors for both the Cauchy loss and the Huber loss as
  \[
  w_i^{(k+1/3)} := \phi'(r_i^{(k+1/3)})/r_i^{(k+1/3)},
  \quad \text{and}
  \quad q_i^{(k+1)} := 1/ \max \{ \gamma, |(\nabla p_c^{(k)})_j| \}.
  \]
  The albedos \( \{ \rho_c \} \) are updated as the solution to the following linear weighted least-squares problem:
  \[
  \{ \rho_c^{(k+1)} \} := \arg \min_{\{ \rho_c \}} \mu \sum_{i,c,j} w_i^{(k+1/3)} \frac{1}{2} (\nabla p_c)_j^2
  + \sum_{i,c,j} w_i^{(k+1/3)} \frac{1}{2} r_i^{(k+1/3)}(\theta_j^{(k+1)}, \rho_c^{(k)}, \lambda c_j^{(k)}, z^{(k)})^2,
  \]
  which is carried out by conjugate gradient (CG).

- **(Update \( \{ \lambda c_j \} \))**: The lighting subproblem is similar to the one for albedos, except for absence of the Huber TV term. Upon evaluation of the residual \( r_i^{(k+2/3)} \) and the weight factor \( w_i^{(k+2/3)} \), we update \( \{ \lambda c_j \} \) by solving the following linear weighted least-squares problem via CG:
  \[
  \{ \lambda c_j^{(k+1)} \} := \arg \min_{\lambda c_j} \sum_{i,c,j} w_i^{(k+2/3)} \frac{1}{2} (\nabla p_i)_j^2.
  \]

- **(Update \( z \))**: The depth subproblem requires additional efforts. With \( r_i^{(k+1)} \) and \( w_i^{(k+1)} \) evaluated after the \( \{ \lambda c_j \} \)-update, we are faced with the following weighted least squares problem:
  \[
  \min_z \frac{1}{2} \sum_{i,c,j} w_i^{(k+1)} \frac{1}{2} r_i^{(k+1)}(\theta_j^{(k+1)}, \rho_c^{(k)}, \lambda c_j^{(k+1)}, z)^2,
  \]
  where the dependence of \( r_i^{(k+1)} \) on \( z \) is still nonlinear. Therefore, we further linearize \( r_i^{(k+1)} \) with respect to \( z \) and arrive at the following update:
  \[
  z^{(k+1)} := \arg \min_z \frac{1}{2} \sum_{i,c,j} w_i^{(k+1)} \frac{1}{2} r_i^{(k+1)}(\theta_j^{(k+1)}, \rho_c^{(k)}, \lambda c_j^{(k+1)}, z)^2 + J_z(z^{(k)})(z - z^{(k)})^2,
  \]
  where \( J_z(z^{(k)}) \) is the Jacobian of the map \( z \mapsto r_i^{(k+1)}(\theta_j^{(k+1)}, \rho_c^{(k)}, \lambda c_j^{(k+1)}, z) \) 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 monotonically decrease of the energy.

### 6. Experimental Validation

This section is concerned with the validation 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 \( \mu \), 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” [27], “Armadillo” [27] and “Thai Statue” [27]) and nine different albedo maps and each of those 36 combinations is rendered as described in (1) using \( M = 25 \) different environment maps, cf. Figure 4. The resulting 25 RGB images per dataset are used as input, along with the intrinsic camera parameters and a binary mask \( \Omega \). A quantitative evaluation on the triplet \( (V, \mu, M) \) 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 \( (V, \mu, M) \) 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 \( \kappa \) is dataset-dependent. For synthetic datasets we always selected this optimal value, yet for real-world data no such evaluation is possible and \( \kappa \) 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 \( \kappa \) can result in appealing geometries.

Next, we evaluate the impact of \( \mu \), cf. Figure 5 (right). As can be seen, the depth estimate seems to deteriorate for too small and too large values of \( \mu \), 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.

---

2Environment maps are downloaded from [http://www.hdrlabs.com/sibl/archive.html](http://www.hdrlabs.com/sibl/archive.html)
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 [31], 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 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.

<table>
<thead>
<tr>
<th>Approach</th>
<th>[15]</th>
<th>[37]</th>
<th>[31]</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>27.16</td>
<td>21.14</td>
<td>34.06</td>
<td>9.17</td>
</tr>
<tr>
<td>Mean</td>
<td>34.15</td>
<td>21.18</td>
<td>35.53</td>
<td>10.72</td>
</tr>
</tbody>
</table>

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 [17]. 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 $\Omega$. 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 [37] 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

---

3Code associated with [15] and [37] can be found online, and the results obtained by [31] were provided by the authors.
the objects’ rough shape (Ovenmitt, Shirt, Tabletcase, Vase) a meaningful geometry is recovered. The approach of [31] 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.
References


