A Variational Approach to Shape-from-shading Under Natural Illumination

Yvain Quéau[1], Jean Mélou[2,3],
Fabien Castan[3], Daniel Cremers[1], and Jean-Denis Durou[2]

[1] Department of Informatics, Technical University Munich, Germany
[2] IRIT, UMR CNRS 5505, Université de Toulouse, France
[3] Mikros Image, Levallois-Perret, France

Abstract. A numerical solution to shape-from-shading under natural illumination is presented. It builds upon a generic PDE-based formulation which handles directional or spherical harmonics lighting, orthographic or perspective projection, and greylevel or multi-channel images. An augmented Lagrangian solver is proposed, which separates the global, nonlinear optimization problem into a sequence of simpler (either local or linear) ones. Real-world applications to shading-aware depth map denoising, refinement and completion are presented.

1 Introduction

Standard 3D-reconstruction pipelines are based on sparse 3D-reconstruction by structure-from-motion (SfM), densified by multi-view stereo (MVS). Both these techniques require unambiguous correspondences based on local color variations. Assumptions behind this requirement are that the surface of interest is Lambertian and well textured. This has proved to be suitable for sparse reconstruction, but problematic for dense reconstruction: dense matching is impossible in textureless areas. In contrast, shape-from-shading (SfS) techniques explicitly model the reflectance of the object surface. The brightness variations observed in a single image provide dense geometric clues, even in textureless areas. SfS may thus eventually push back the limits of MVS.

However, most shape-from-shading methods require a highly controlled illumination and thus they may fail when deployed outside the lab. **Numerical methods for SFS under natural illumination are still lacking**. Besides, SFS remains a classic ill-posed problem with well-known ambiguities such as the concave/convex ambiguity. Solving such ambiguities for real-world applications requires **handling priors on the surface**. There exist two main numerical strategies for solving and disambiguating shape-from-shading [1]. Variational methods [2] ensure smoothness through regularization. Handle priors is easy, but tuning the regularization may be tedious. Alternatively, methods based on the exact resolution of a nonlinear PDE [3] implicitly enforce differentiability (almost everywhere). These methods do not require any tuning, but they lack robustness and they require a boundary condition. To combine the advantages of each approach, **a variational solution based on PDEs would be worthwile for SFS under natural illumination**.

Contributions – This work proposes a generic numerical framework for SFS under natural illumination (see Figure 1). After reviewing existing solutions in Section 2, we introduce in Section 3 a new PDE-based model for SFS, which handles various illumination and camera models. Whatever these models, the same ADMM-based solution, described in Section 4, can be employed. It reformulates SFS as a sequence of easier subproblems: local estimation of the surface normals (possibly, with a smoothness prior), and then integration of surface normals into a depth map (possibly, with a shape prior). Experiments on synthetic datasets are presented in Section 5, as well as real-world applications to depth refinement and completion for RGB-D cameras or stereovision systems. Our achievements are eventually summarized in Section 6.

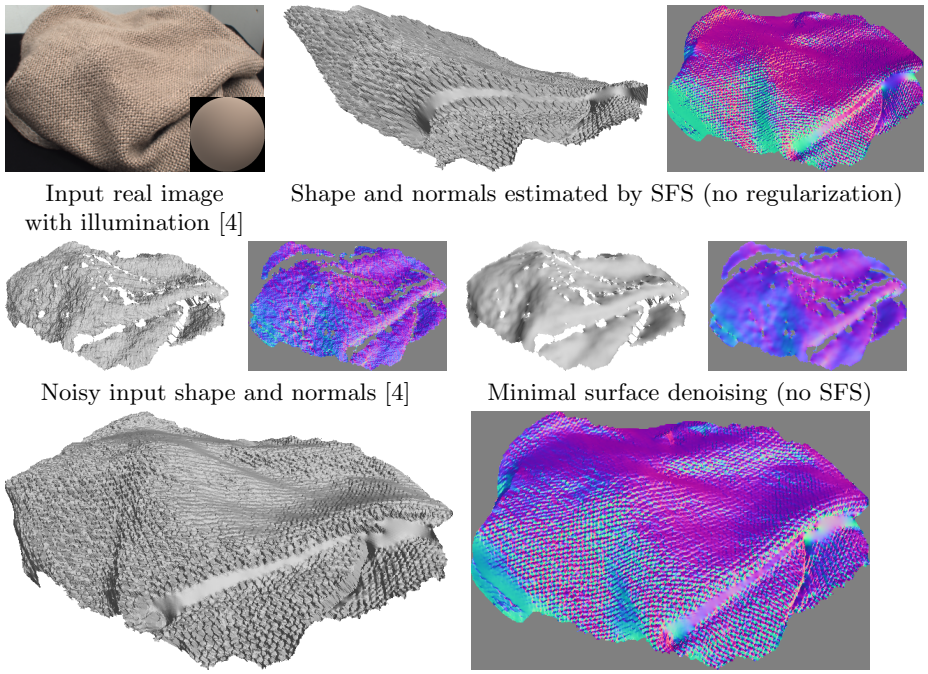


Fig. 1: We propose the generic variational framework (12) for shape-from-shading (SFS) under natural illumination (top row: $(\lambda, \mu, \nu) = (1, 0, 0)$). It is able to estimate a smooth surface (out of infinitely many), which almost exactly solves the generic SFS model (2). To disambiguate SFS and improve robustness, prior surface knowledge (middle row, left: $(\lambda, \mu, \nu) = (0, 1, 0)$) and minimal surface regularization (middle row, right: $(\lambda, \mu, \nu) = (0, 1, 5.10^{-5})$) can be further included in the variational framework. These building blocks can be put together for shading-aware joint depth denoising, refinement and completion (bottom row: $(\lambda, \mu, \nu) = (1, 1, 5.10^{-5})$).

2 Image Formation Model and Related Works

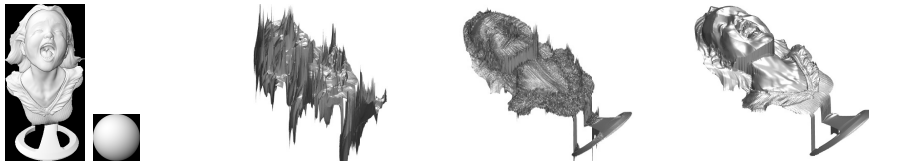
In the following, a 3D-frame $(Oxyz)$ is attached to the camera, O being the optical center and the axis Oz being parallel to the optical axis, such that z is oriented towards the scene. We denote $I : \Omega \subset \mathbb{R}^2 \to \mathbb{R}^C$, $(x, y) \mapsto I(x, y) = \left[I^1(x, y), \dots, I^C(x, y)\right]^\top$ a greylevel ($C = 1$) or multi-channel ($C > 1$) image of a surface, where Ω represents a "mask" of the object being pictured. We assume that the surface is Lambertian, so its reflectance is characterized by the albedo ρ. We further consider a second-order spherical harmonics model for the lighting vector $\mathbf{l}$. To deal with the spectral dependencies of reflectance and lighting, we assume both ρ and $\mathbf{l}$ are channel-dependent. The albedo is thus a function $\rho : \Omega \to \mathbb{R}^C$, $(x, y) \mapsto \rho(x, y) = \left[\rho^1(x, y), \dots, \rho^C(x, y)\right]^\top$, and the lighting in each channel $c \in \{1, \dots, C\}$ is represented as a vector $\mathbf{l}^c = [l_1^c, l_2^c, l_3^c, l_4^c, l_5^c, l_6^c, l_7^c, l_8^c, l_9^c]^\top \in \mathbb{R}^9$. Eventually, let $\mathbf{n} : \Omega \to \mathbb{S}^2 \subset \mathbb{R}^3$, $(x, y) \mapsto \mathbf{n}(x, y) = [n_1(x, y), n_2(x, y), n_3(x, y)]^\top$ be the field of unit-length outward normals to the surface. The image formation model is then written [5]:

$$I^c(x, y) = \rho^c(x, y)\, \mathbf{l}^c \cdot \begin{bmatrix} \mathbf{n}(x, y) \\ 1 \\ n_1(x, y) n_2(x, y) \\ n_1(x, y) n_3(x, y) \\ n_2(x, y) n_3(x, y) \\ n_1(x, y)^2 - n_2(x, y)^2 \\ 3 n_3(x, y)^2 - 1 \end{bmatrix}, \ (x, y) \in \Omega, \ c \in \{1, \dots, C\}. \quad (1)$$

The goal of SFS is to recover the object shape, given its image, its albedo and the lighting. Each unit-length normal vector $\mathbf{n}(x, y)$ has two degrees of freedom, thus each Equation (1), $(x, y) \in \Omega$, $c \in \{1, \dots, C\}$, is a nonlinear equation with two unknowns. If $C = 1$, it is impossible to solve such an equation locally: all these equations must be solved together, by coupling the surface normals in order to ensure surface smoothness. When $C > 1$ and the lighting vectors are non-coplanar, ambiguities theoretically disappear [6]. However, under natural lighting these vectors are close to being collinear, and thus locally solving (1) is numerically challenging (bad conditioning). Again, a global solution should be preferred but this time, for robustness reasons.

There is a large amount of literature on numerical SFS, in the specific case where $C = 1$ and lighting is directional ($l_4^c = \dots = l_9^c = 0$), see for instance [1]. However, few SFS methods deal with more general spherical harmonics lighting. First-order harmonics have been considered in [7,8], but they only capture up to 90% of natural lighting, while this rate is over 99% using second-order harmonics [9]. The latter have been used in [10], where the challenging problem of shape, illumination and reflectance from shading (SIRFS) is tackled (this method is also applicable to SFS if albedo and lighting are fixed). However, all these works heavily rely on regularization mechanisms, and not only for disambiguation. For instance, SIRFS "fails badly" [10] without a multi-scale strategy,

and the method of [8] becomes unstable without depth regularization (see Figure 2). Although regularization mechanisms somewhat circumvent these issues in practice, an ideal numerical solver would rely on regularization only for disambiguation and for handling noise, not for enforcing numerical stability. In order to design such a solver, a variational approach based on PDEs may be worthwile. Let us thus first rewrite (1) as a nonlinear PDE.



Input synthetic image and illumination | Fixed point [8] without regularization | Single-scale SIRFS [10] | Proposed (without regularization)

Fig. 2: Greylevel shape-from-shading using first-order spherical harmonics. Linearization strategies such as the fixed point one used in [8] fail if regularization is not employed. Similar issues arise in SIRFS [10] when the multi-scale approach is not used. Our SFS method can use regularization for disambiguation and improving robustness, but it remains stable even without. In these three experiments, the same initial shape was used (the "Realistic initialization" of Figure 3).

3 A Generic PDE-based Model for Shape-from-shading

We assume hereafter that lighting and albedo are known (in practice, it is enough to assume that the albedo is uniform, and to estimate lighting from a gross surface approximation). These assumptions are usual in the SFS literature. They could be relaxed by simultaneously estimating shape, illumination and reflectance [10], but we leave this as future work and focus only on shape estimation. This is the most challenging part anyways, since (1) is linear in the lighting and the albedo, but it is generally nonlinear in the normal.

In order to comply with the discussion above, Equation (1) should be solved *globally* over the entire domain Ω. To this end, we do not estimate the normals but rather the underlying depth map, through a PDE-based approach [3]. This has the advantage of implicitly enforcing smoothness (almost everywhere) without requiring any regularization term (regularization will be introduced in Section 4, but only for the sake of disambiguation and robustness against noise). We show in this section the following result:

Proposition 1. *Under both orthographic and perspective projection, the image formation model* (1) *can be rewritten as the following nonlinear PDE in* z*:*

$$\mathbf{a}^c_{(\nabla z)} \cdot \nabla z + b^c_{(\nabla z)} = I^c \quad \text{over } \Omega,\ c \in \{1, \dots, C\} \tag{2}$$

with $z : \Omega \to \mathbb{R}$ *a map characterizing the shape,* $\nabla z : \Omega \to \mathbb{R}^2$ *its gradient, and where* $\mathbf{a}^c_{(\nabla z)} : \Omega \to \mathbb{R}^2$ *and* $b^c_{(\nabla z)} : \Omega \to \mathbb{R}$ *depend in a nonlinear way on* ∇z*.*

Proof. The 3D-shape can be represented as a patch over the image domain:

$$\begin{aligned}\mathbf{x}: \quad \Omega \quad &\to \mathbb{R}^3 \\ (x,y) &\mapsto \begin{cases} [x, y, \tilde{z}(x,y)]^\top & \text{under orthographic projection,} \\ \tilde{z}(x,y)\left[\frac{x-x_0}{\tilde{f}}, \frac{y-y_0}{\tilde{f}}, 1\right]^\top & \text{under perspective projection,} \end{cases}\end{aligned} \tag{3}$$

with $\tilde{z}$ the *depth map*, $\tilde{f} > 0$ the *focal length*, and $(x_0, y_0) \in \Omega$ the coordinates of the *principal point* in the image plane.

Using this parameterization, the normal to the surface in a surface point $\mathbf{x}(x,y)$ is the unit-length, outgoing vector proportional to the cross product $\mathbf{x}_x(x,y) \times \mathbf{x}_y(x,y)$, where $\mathbf{x}_x$ (resp. $\mathbf{x}_y$) is the partial derivative of $\mathbf{x}$ along the x (resp. y)-direction. After a bit of algebra, the following formula is obtained:

$$\begin{aligned}\mathbf{n}: \quad \Omega \quad &\to \mathbb{S}^2 \subset \mathbb{R}^3 \\ (x,y) &\mapsto \frac{1}{d_{(\nabla z)}(x,y)} \begin{bmatrix} f\,\nabla z(x,y) \\ -1 - [\tilde{x}, \tilde{y}]^\top \cdot \nabla z(x,y) \end{bmatrix},\end{aligned} \tag{4}$$

where

$$(z, f, \tilde{x}, \tilde{y}) = \begin{cases} (\tilde{z}, 1, 0, 0) & \text{under orthographic projection,} \\ (\log \tilde{z}, \tilde{f}, x - x_0, y - y_0) & \text{under perspective projection,} \end{cases} \tag{5}$$

and where the map $d_{(\nabla z)}$ ensures the unit-length constraint:

$$\begin{aligned} d_{(\nabla z)}: \quad \Omega \quad &\to \mathbb{R} \\ (x,y) &\mapsto \sqrt{f^2 \|\nabla z(x,y)\|^2 + \left(1 + (\tilde{x}, \tilde{y}) \cdot \nabla z(x,y)\right)^2}. \end{aligned} \tag{6}$$

Note that $\|d_{(\nabla z)}\|_{\ell^1(\Omega)}$ is the total area of the surface, which will be used in Section 4 for designing a regularization term.

By plugging (4) into (1), we obtain the nonlinear PDE (2), if we denote:

$$\begin{aligned} \mathbf{a}^c_{(\nabla z)}: \quad \Omega \quad &\to \mathbb{R}^2 \\ (x,y) &\mapsto \frac{\rho^c(x,y)}{d_{(\nabla z)}(x,y)} \begin{bmatrix} f\, l^c_1 - \tilde{x}\, l^c_3 \\ f\, l^c_2 - \tilde{y}\, l^c_3 \end{bmatrix}, \end{aligned} \tag{7}$$

$$\begin{aligned} b^c_{(\nabla z)}: \quad \Omega \quad &\to \mathbb{R} \\ (x,y) &\mapsto \rho^c \begin{bmatrix} l^c_3 \\ l^c_4 \\ l^c_5 \\ l^c_6 \\ l^c_7 \\ l^c_8 \\ l^c_9 \end{bmatrix} \cdot \begin{bmatrix} \frac{-1}{d_{(\nabla z)}(x,y)} \\ 1 \\ \frac{f^2 z_x(x,y) z_y(x,y)}{\left(d_{(\nabla z)}(x,y)\right)^2} \\ \frac{f z_x(x,y)(-1-(\tilde{x},\tilde{y})\cdot\nabla z(x,y))}{\left(d_{(\nabla z)}(x,y)\right)^2} \\ \frac{f z_y(x,y)(-1-(\tilde{x},\tilde{y})\cdot\nabla z(x,y))}{\left(d_{(\nabla z)}(x,y)\right)^2} \\ \frac{f^2\left(z_x(x,y)^2 - z_y(x,y)^2\right)}{\left(d_{(\nabla z)}(x,y)\right)^2} \\ \frac{3(-1-(\tilde{x},\tilde{y})\cdot\nabla z(x,y))^2}{\left(d_{(\nabla z)}(x,y)\right)^2} - 1 \end{bmatrix}. \end{aligned} \tag{8}$$

□

When $C = 1$, camera is orthographic and lighting is directional and frontal (*i.e.*, l_3 is the only non-zero lighting component), then (2) becomes the *eikonal equation* $\frac{\rho^c |l_3|}{\sqrt{1+\|\nabla z\|^2}} = I^c$. Efficient numerical methods for solving this nonlinear PDE have been suggested, using for instance semi-Lagrangian schemes [11]. Such techniques can also handle perspective camera projection and/or nearby point light source illumination [12]. Still, existing PDE-based methods require a boundary condition, or at least a state constraint, which is rarely available in practice. In addition, the more general PDE-based model (2), which handles both orthographic or perspective camera, directional or second-order spherical harmonics lighting, and greylevel or color images, has not been tackled so far. A variational solution to this generic SFS problem is now presented.

4 Variational Formulation and Optimization

The C PDEs in (2) are in general incompatible due to noise. Thus, an approximate solution must be sought. If we assume that the image formation model (1) is satisfied up to an additive, zero-mean and homoskedastic, Gaussian noise, then the maximum likelihood solution is attained by estimating the depth map z which minimizes the following least-squares cost function:

$$\mathcal{E}(\nabla z; I) = \sum_{c=1}^{C} \left\| \mathbf{a}^c_{(\nabla z)} \cdot \nabla z + b^c_{(\nabla z)} - I^c \right\|^2_{\ell^2(\Omega)} . \tag{9}$$

In recent works on shading-based refinement [8], it is suggested to minimize a cost function similar to (9) iteratively, by freezing the nonlinear fields $\mathbf{a}^c$ and b^c at each iteration. This strategy must be avoided. For instance, it cannot handle the simplest case of orthographic projection and directional, frontal lighting: this yields $\mathbf{a}^c \equiv \mathbf{0}$ according to (7), and thus (9) does not even depend on the unknown depth z if b^c is freezed. Even in less trivial cases, Figure 2 shows that this strategy is unstable, which explains why regularization is employed in [8]. We also resort to regularization, but only for the sake of disambiguating SFS and handling noise: our proposal yields a stable solution even in the absence of regularization (see Figure 2). In this work, we consider two types of regularization: one which represents prior knowledge of the surface, and one which ensures its smoothness.

4.1 Regularized Variational Model

In some applications such as RGB-D sensing, or MVS, a depth map z^0 is available. This depth map is usually noisy and incomplete, but it may represent a useful "guide" for SFS. We may thus consider the following prior term:

$$\mathcal{P}(z; z^0) = \left\| z - z^0 \right\|^2_{\ell^2(\Omega^0)} , \tag{10}$$

where $\Omega^0 \subseteq \Omega \subset \mathbb{R}^2$ is the image region for which prior information is available.

In order not to interpret noise in the image as geometric artifacts, one may want to improve robustness by explicitly including a smoothness term. However, standard total variation regularization, which is often considered in image processing, may tend to favor piecewise fronto-parallel surfaces and thus induce staircasing. We rather penalize the total area of the surface, which has recently been shown in [13] to be better suited for depth map regularization. To this end, let us remark that in differential geometry terms, the map $d_{(\nabla z)}$ defined in (6) is the square root of the determinant of the first fundamental form of function z (metric tensor). Its integral over Ω is exactly the area of the surface, and thus the following smoothness term may be considered:

$$\mathcal{S}(\nabla z) = \left\| d_{(\nabla z)} \right\|_{\ell^1(\Omega)}. \tag{11}$$

Putting altogether the pieces (9), (10) and (11), we obtain the following variational problem:

$$\min_{z\colon \Omega \to \mathbb{R}} \lambda\, \mathcal{E}(\nabla z; I) + \mu\, \mathcal{P}(z; z^0) + \nu\, \mathcal{S}(\nabla z), \tag{12}$$

where $(\lambda, \mu, \nu) \geq (0, 0, 0)$ are user-defined parameters controlling the respective influence of each term.

Let us remark that our variational model (12) yields a pure SFS model if $\mu = \nu = 0$, a depth denoising model similar to that in [13] if $\lambda = 0$ and $\Omega^0 = \Omega$, and a shading-aware joint depth refinement and completion if $\lambda > 0$, $\mu > 0$ and $\Omega^0 \subsetneq \Omega$.

4.2 Numerical Solution

The variational problem (12) is difficult to solve, because the shading term $\mathcal{E}(\nabla z; I)$ and the regularization term $\mathcal{S}(\nabla z)$ are not only nonlinear, but they also depend on the depth gradient. We propose to separate the difficulty induced by the nonlinearity from that induced by the dependency on the gradient. To this end, we introduce an auxiliary variable $\theta : \Omega \to \mathbb{R}^2$, and rewrite (12) as a constrained optimization problem:

$$\begin{aligned} &\min_{\substack{z\colon \Omega \to \mathbb{R} \\ \theta\colon \Omega \to \mathbb{R}^2}} \lambda\, \mathcal{E}(\theta; I) + \mu\, \mathcal{P}(z; z^0) + \nu\, \mathcal{S}(\theta) \\ &\text{s.t. } \nabla z = \theta. \end{aligned} \tag{13}$$

We solve (13) using an ADMM algorithm. The augmented Lagrangian functional associated to (13) is defined as

$$\mathcal{L}_\beta(z, \theta, \Psi) = \lambda\, \mathcal{E}(\theta; I) + \mu\, \mathcal{P}(z; z^0) + \nu\, \mathcal{S}(\theta) + \langle \Psi, \nabla z - \theta \rangle + \frac{\beta}{2} \left\| \nabla z - \theta \right\|_2^2, \tag{14}$$

with $\Psi : \Omega \to \mathbb{R}^2$ the field of Lagrange multipliers, $\langle \cdot\, \rangle$ the scalar product induced by $\| \cdot \|_2$ over Ω, and $\beta > 0$.

ADMM iterations [14] are then written:

$$\theta^{(k+1)} = \underset{\theta}{\text{argmin}} \ \mathcal{L}_{\beta^{(k)}}(z^{(k)}, \theta, \Psi^{(k)}), \tag{15}$$

$$z^{(k+1)} = \underset{z}{\text{argmin}} \ \mathcal{L}_{\beta^{(k)}}(z, \theta^{(k+1)}, \Psi^{(k)}), \tag{16}$$

$$\Psi^{(k+1)} = \Psi^{(k)} + \beta^{(k)} \left(\nabla z^{(k+1)} - \theta^{(k+1)} \right). \tag{17}$$

where $\beta^{(k)}$ can be determined automatically [14].

Problem (15) is a pixelwise nonlinear least-squares problem which is solved using a Newton method with a L-BFGS stepsize. As for (16), it is discretized by first-order, forward finite differences. This yields a linear least-squares problem whose normal equations provide symmetric, positive definite (semi-definite if $\mu = 0$) linear system. It is sparse, but too large to be solved directly: conjugate gradient iterations should be preferred. In our experiments, the algorithm stops when the relative variation of the energy in (12) falls below 10^{-3}.

This ADMM algorithm can be interpreted as follows. During the θ-update (15), local estimation of the gradient (*i.e.*, of the surface normals) is carried out based on SFS, while ensuring that the gradient map is smooth and close to the gradient of the current depth map. Unlike in the fixed point approach [8], local surface orientation is inferred from the whole model (2), and not only from its linear part. In practice, we observed that this yields a much more stable algorithm (see Figure 2). In the z step (16), these surface normals are integrated into a new depth map, which should stay close to both the previous one and the prior.

Given the non-convexity of the shading term $\mathcal{E}$ and of the smoothness term $\mathcal{S}$, convergence of the ADMM algorithm is not guaranteed. However, in practice we did not observe any particular convergence-related issue, so we conjecture that a convergence proof could eventually be provided. However, we leave this as future work and focus in this proof of concept work on sketching the approach and providing preliminary empirical results. The next section shows quantitatively the effectiveness of the proposed ADMM algorithm for solving SFS under natural illumination, and introduces qualitative results on real-world datasets.

5 Experiments

5.1 Quantitative Evaluation of the Proposed SFS Framework

We first validate in Figure 3 the ability of the proposed variational framework to solve SFS under natural illumination *i.e.*, to solve (1). Our approach is compared against SIRFS [10], which is the only method for SFS under natural illumination whose code is freely available. For fair comparison, albedo and lighting estimations are disabled in SIRFS, and its multi-scale strategy is used, in order to avoid the artifacts shown in Figure 2.

Since we only want to compare here the ability of both methods to explain a shaded image, our regularization terms are disabled ($\mu = \nu = 0$), as well as those from SIRFS. To quantify this ability, we measure the RMSE between the input images and the reprojected ones.

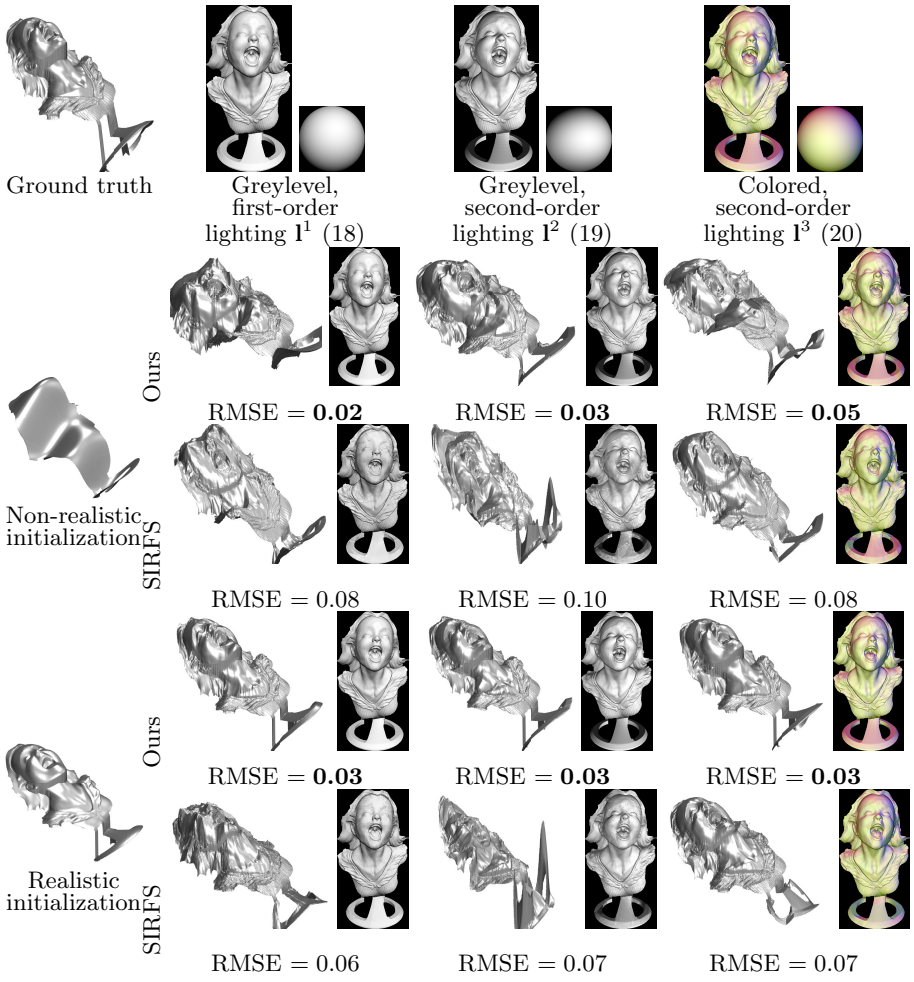


Fig. 3: Evaluation of our SFS approach against the multi-scale one from SIRFS [10], in three different lighting situations and using two different initial estimates (the first one is Matlab's "peaks" function, the second one is a smoothed version of the ground truth). For each experiment, we show the estimated depth map and the reprojected image, and provide the root mean square error (RMSE) between the input synthetic image and the reprojection (the input images are scaled between 0 and 1). Our variational framework solves SFS under natural illumination more accurately than state-of-the-art.

To create these datasets, we use the public domain "Joyful Yell" 3D-shape, considering orthographic projection for fair comparison (SIRFS cannot handle perspective projection). Noise-free images are simulated under three lighting scenarios. We first consider greylevel images, with a single-order and then a second-order lighting vector. Eventually, we consider a colored, second-order lighting vector. These lighting vectors are defined, respectively, by:

$$\mathbf{l}^1 = [0.1, -0.25, -0.7, 0.2, 0, 0, 0, 0, 0]^\top, \tag{18}$$

$$\mathbf{l}^2 = [0.2, 0.3, -0.7, 0.5, -0.2, -0.2, 0.3, 0.3, 0.2]^\top, \tag{19}$$

$$\mathbf{l}^3 = \begin{bmatrix} -0.2 & -0.2 & -1 & 0.4 & 0.1 & -0.1 & -0.1 & -0.1 & 0.05 \\ 0 & 0.2 & -1 & 0.3 & 0 & 0.2 & 0.1 & 0 & 0.1 \\ 0.2 & -0.2 & -1 & 0.2 & -0.1 & 0 & 0 & 0.1 & 0 \end{bmatrix}^\top . \tag{20}$$

To illustrate the underlying ambiguities, we consider two different initial estimates: one very different from the ground truth (Matlab's "peaks" function), and one close to it (obtained by applying a Gaussian filter to the ground truth). Interestingly, although $\mu = 0$ for the tests in Figure 3, our method does not drift too much from the latter: the shape is qualitatively satisfactory in all the experiments when a good initialization is available.

In all the experiments, the images are better explained using our framework, which shows that the proposed numerical framework solves the challenging, highly nonlinear SFS model (1) in a more accurate manner than state-of-the-art. Besides, the runtimes of both methods are comparable: a few minutes in all cases (on a standard laptop using Matlab codes), for images having around 150.000 pixels inside Ω. Unsurprisingly, initialization matters a lot, because of the inherent ambiguities of SFS.

Disambiguation can be carried out through regularization. This is illustrated in Figure 4, where we consider the same dataset as in the second experiment of Figure 3, but with additive, zero-mean, homoskedastic Gaussian noise on the image and on the depth forming the shape prior (we use the "Realistic initialization" as prior). If $\lambda = 1$ and $(\mu, \nu) = (0, 0)$, then pure SFS is carried out: high-frequency details are perfectly recovered, but the surface may drift from the initial estimate and interpret image noise as unwanted geometric artifacts. If $\mu \to +\infty$, the initial estimate (which exhibits reasonable low-frequency components, but no geometric detail) is not modified. If $\nu \to +\infty$, then only the minimal surface term matters, hence the result is over-smoothed. In this experiment, we also evaluate the accuracy of the 3D-reconstruction through the mean angular error (MAE) on the normals: it is minimal when the parameters are tuned appropriately, not when the image error (RMSE) is minimal.

The appropriate tuning of μ and ν depends on how trustworthy the image and the shape prior are. Typically, in RGB-D sensing, the depth may be noisier than in this synthetic experiment so one may want to use a lower value of μ. On the other hand, natural lighting is generally colored, so the three image channels provide redundant information: regularization is less important and the smoothness parameter ν can be reduced. We found that $(\lambda, \mu, \nu) = (1, 1, 5.10^{-5})$ prodives qualitatively nice results in all our real-world experiments.

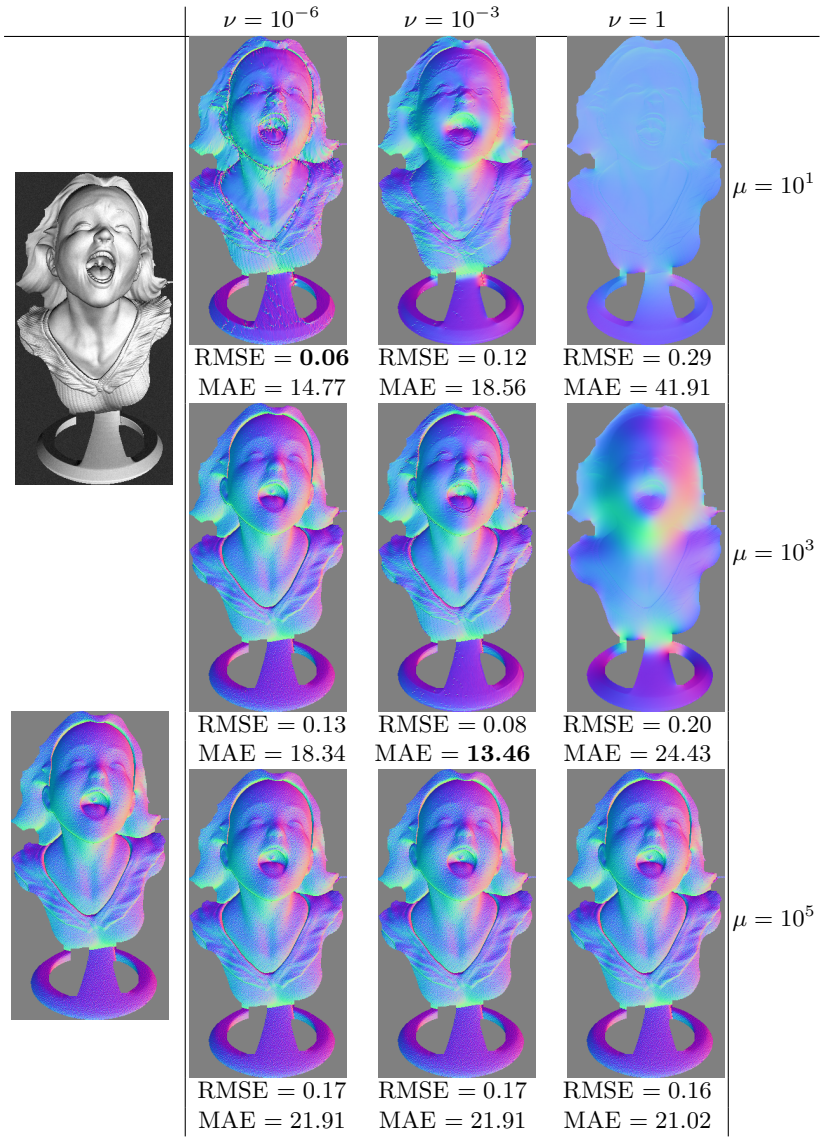


Fig. 4: Left: input noisy image ($\sigma_I = 0.02$) and noisy prior shape ($\sigma_z = 0.2\%\|z\|_\infty$), represented by a normal map to emphasize details. Right: estimated shape with $\lambda = 1$ and various values of μ and ν. The RMSE between the image and the reprojection is minimal when regularization is minimal, but the mean angular error (MAE, in degrees) between the estimated shape and the ground truth one is not.

5.2 Qualitative Evaluation on Real-world Datasets

The importance of initialization is further confirmed in the top rows of Figures 1 and 5. In these experiments, our SFS method ($\mu = \nu = 0$) is evaluated, under perspective projection, on real-world datasets obtained using an RGB-D sensor [4], considering a fronto-parallel surface as initialization. Although fine details are revealed, the results present an obvious low-frequency bias and artifacts due to the image noise occur. This illustrates the inherent ambiguities of SFS, and the need for regularization.

In order to illustrate the practical disambiguation of SFS using regularization, we next consider as initialization and prior z^0 the depth provided by the RGB-D sensor. It is either noisy and incomplete, but with our framework it can be denoised, refined and completed in a shading-aware manner, by tuning the parameters μ (prior) and ν (smoothness). Second and third rows of Figures 1 and 5 illustrate the interest of SFS for depth refinement, in comparison with "blind" methods based solely on depth regularization [13].

Eventually, Figure 6 demonstrates an application to stereovision, using a real-world dataset from [15]. This time, the initial depth map is obtained by a multi-view stereo algorithm [16]. We estimated lighting from this initial depth map, assuming uniform albedo. Then, we let our algorithm recover the thin geometric structures, which are missed by multi-view stereo. The initial depth map contains a lot of missing data and discontinuities, which is challenging for our algorithm: ambiguities arise inside the large holes, and our model favors smooth surfaces. Indeed, the concavity is not very nicely recovered, and the discontinuities are partly smoothed. Still, nice details are recovered, and the overall surface is reasonable.

6 Conclusion and Perspective

We have introduced a generic variational framework for SFS under natural illumination, which can be applied in a broad range of scenarios. It relies on a tailored PDE-based SFS formulation which handles a variety of scenarios for the camera and the lighting. To solve the resulting system of PDEs, we introduce an ADMM algorithm which separates the difficulty due to nonlinearity from that due to the dependency upon the gradient. Shape prior and nonlinear smoothing terms are easily included in this variational framework, allowing practical disambiguation of SFS as well as natural applications to depth map refinement and completion for RGB-D sensors or stereovision systems.

As future work, we plan to investigate the convergence of the proposed ADMM algorithm for our non-convex problem, and to include reflectance and lighting estimation. With these extensions, we have good hope that the proposed variational framework may be useful in other computer vision applications, such as shading-aware dense multi-view stereo.

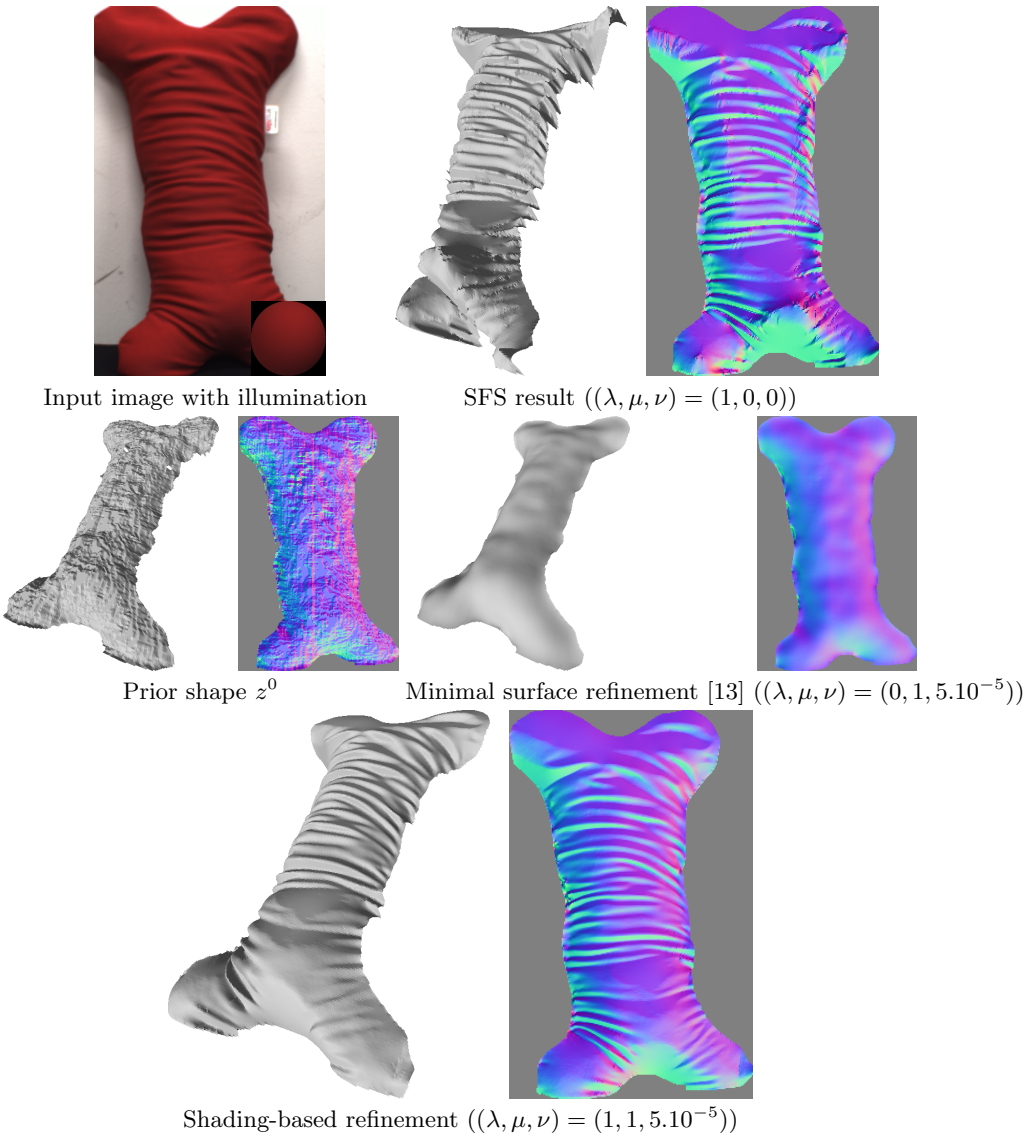


Fig. 5: Result obtained using our variational framework for three computer vision problems: pure SFS, "blind" (not shading-based) depth refinement, and shading-based depth refinement. The shape estimated by SFS is distorted (due to the ambiguities of SFS), and artifacts occur (due to noise and the absence of regularization), but it contains the fine-scale details. The depth map provided by the RGB-D sensor is nicely denoised without considering shading, but thin structures are missed. With the proposed method, noise is removed and fine details are revealed.

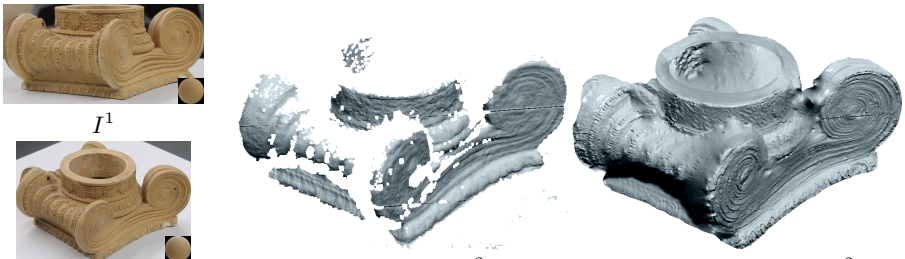


Fig. 6: Left: two (out of $N = 30$) images I^1 and I^2 of the "Figure" object [15]. Middle: depth map $z^{0,2}$ (same viewpoint as image I^2) obtained by the CMPMVS method [16] (before meshing). Right: refined and completed depth map z^2.

References

1. Durou, J.D., Falcone, M., Sagona, M.: Numerical methods for shape-from-shading: A new survey with benchmarks. CVIU **109** (2008) 22–43
2. Horn, B.K.P., Brooks, M.J.: The variational approach to shape from shading. CVGIP **33** (1986) 174–208
3. Lions, P.L., Rouy, E., Tourin, A.: Shape-from-shading, viscosity solutions and edges. Numerische Mathematik **64** (1993) 323–353
4. Han, Y., Lee, J.Y., Kweon, I.S.: High quality shape from a single RGB-D image under uncalibrated natural illumination. In: ICCV. (2013)
5. Basri, R., Jacobs, D.P.: Lambertian reflectances and linear subspaces. PAMI **25** (2003) 218–233
6. Johnson, M.K., Adelson, E.H.: Shape estimation in natural illumination. In: CVPR. (2011)
7. Huang, R., Smith, W.A.P.: Shape-from-shading under complex natural illumination. In: ICIP. (2011)
8. Or-El, R., Rosman, G., Wetzler, A., Kimmel, R., Bruckstein, A.: RGBD-Fusion: Real-Time High Precision Depth Recovery. In: CVPR. (2015)
9. Frolova, D., Simakov, D., Basri, R.: Accuracy of spherical harmonic approximations for images of Lambertian objects under far and near lighting. In: ECCV. (2004)
10. Barron, J.T., Malik, J.: Shape, illumination, and reflectance from shading. PAMI **37** (2015) 1670–1687
11. Cristiani, E., Falcone, M.: Fast semi-Lagrangian schemes for the eikonal equation and applications. SIAM JNA **45** (2007) 1979–2011
12. Breuß, M., Cristiani, E., Durou, J.D., Falcone, M., Vogel, O.: Perspective shape from shading: Ambiguity analysis and numerical approximations. SIAM JIS **5** (2012) 311–342
13. Graber, G., Balzer, J., Soatto, S., Pock, T.: Efficient minimal-surface regularization of perspective depth maps in variational stereo. In: CVPR. (2015)
14. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Founds. Trends in Mach. Learn. **3** (2011) 1–122
15. Zollhöfer, M., Dai, A., Innman, M., Wu, C., Stamminger, M., Theobalt, C., Nießner, M.: Shading-based refinement on volumetric signed distance functions. ACM Trans. Graph. **34** (2015) 96:1–96:14
16. Jancosek, M., Pajdla, T.: Multi-view reconstruction preserving weakly-supported surfaces. In: CVPR. (2011)