



HAL
open science

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

Yvain Quéau, Jean Mélou, Fabien Castan, Jean-Denis Durou, Daniel Cremers

► **To cite this version:**

Yvain Quéau, Jean Mélou, Fabien Castan, Jean-Denis Durou, Daniel Cremers. A Variational Approach to Shape-from-shading Under Natural Illumination. 11th International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR 2017), Oct 2017, Venice, Italy. pp.342-357, 10.1007/978-3-319-78199-0_23 . hal-02118556

HAL Id: hal-02118556

<https://normandie-univ.hal.science/hal-02118556v1>

Submitted on 3 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Yvain QUÉAU¹, Jean MÉLOU^{2,3},
Fabien CASTAN³, Daniel CREMERS¹, and Jean-Denis DUROU²

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 an augmented Lagrangian approach for solving a generic PDE-based shape-from-shading model which handles directional or spherical harmonic lighting, orthographic or perspective projection, and greylevel or multi-channel images. 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 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 shape-from-shading [1]. Variational methods [2] ensure smoothness through regularization. Handling priors is easy, but tuning the regularization may be tedious. Alternatively, methods based on the exact resolution of a nonlinear PDE [3], which implicitly enforce differentiability (almost everywhere), 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 worthwhile for SFS under natural illumination.**

Contributions – This work proposes a generic numerical framework for SFS under natural illumination, which can be employed to achieve either pure shape-from-shading or shading-aware depth refinement (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. A variational approach for solving the arising PDE is proposed in Section 4, which includes optional regularization terms for incorporating a shape prior or enforcing smoothness. Numerical solving is carried out using an ADMM algorithm. 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.

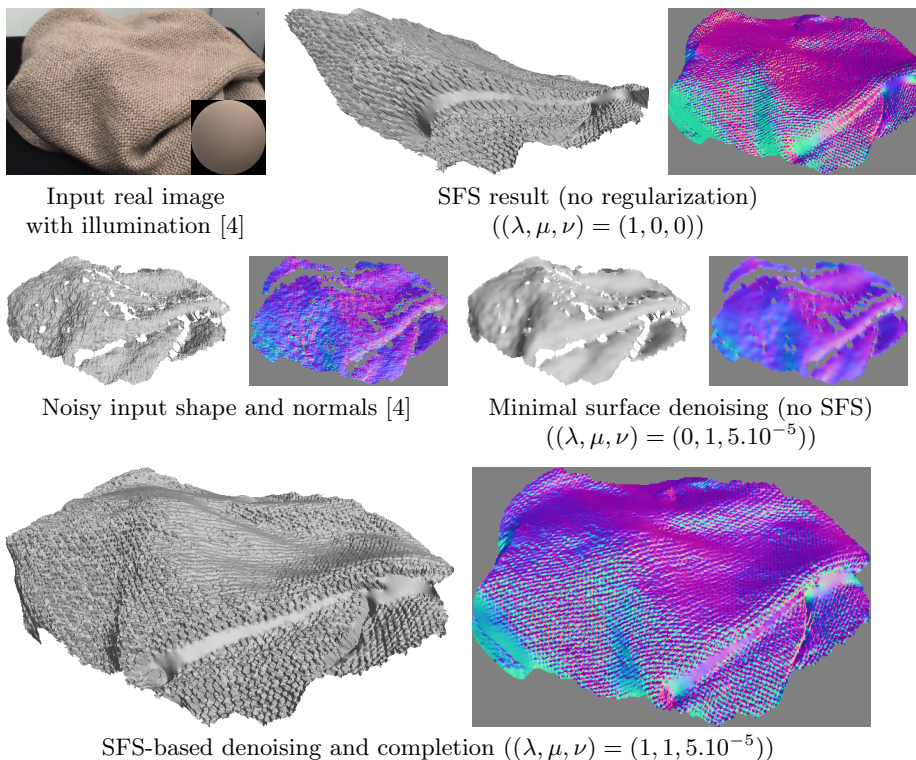


Fig. 1: We propose the generic variational framework (16) for shape-from-shading (SFS) under natural illumination (top row). It is able to estimate a smooth surface (out of infinitely many), which almost exactly solves the generic SFS model (2). To disambiguate SFS and improve robustness, prior surface knowledge (middle row, left) and minimal surface regularization (middle row, right) 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).

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 coinciding with the optical axis, such that z is oriented towards the scene. We denote $I : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}^C$, $(x, y) \mapsto I(x, y) = [I^1(x, y), \dots, I^C(x, y)]^\top$ 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 completely characterized by the albedo ρ . We further consider a second-order spherical harmonic model for the lighting vector \mathbf{l} . To deal with the spectral dependencies of reflectance and lighting, we assume a general model where both ρ and \mathbf{l} are channel-dependent. The albedo is thus a function $\rho : \Omega \rightarrow \mathbb{R}^C$, $(x, y) \mapsto \rho(x, y) = [\rho^1(x, y), \dots, \rho^C(x, y)]^\top$, 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 $\mathbf{n} : \Omega \rightarrow \mathbb{S}^2 \subset \mathbb{R}^3$, $(x, y) \mapsto \mathbf{n}(x, y) = [n_1(x, y), n_2(x, y), n_3(x, y)]^\top$ be the field of unit-length outward normals to the surface. The image formation model is then written as the following extension of a well-known model [5]:

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

However, let us remark that the writing (1), where both the reflectance and the lighting are channel-dependent, is abusive. Since the camera response function is also channel-dependent, this model is indeed justified only for white surfaces ($\rho^c = \rho$, $\forall c \in \{1, \dots, C\}$) under colored lighting, or for colored surfaces under white lighting ($\mathbf{l}^c = \mathbf{l}$, $\forall c \in \{1, \dots, C\}$). See, for instance, [6] for some discussion. In the following, we still consider the general model (1), with a view in designing a generic SFS solver handling both situations. However, in the experiments we will only consider white surfaces.

In SFS, both the reflectance values $\{\rho^c\}_{c \in \{1, \dots, C\}}$ and the lighting vectors $\{\mathbf{l}^c\}_{c \in \{1, \dots, C\}}$ are assumed to be known. The goal is to recover the object shape, represented in (1) by the normal field \mathbf{n} . 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, for instance, surface smoothness. When $C > 1$ and the lighting vectors are non-coplanar, ambiguities theoretically disappear [7]. However, under natural illumination 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 harmonic lighting. First-order harmonics have been considered in [8,9], but they only capture up to 90% of natural illumination, while this rate is over 99% using second-order harmonics [10]. The latter have been used in [11], 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 multi-scale or regularization mechanisms, and not only for disambiguation or for handling noise. For instance, SIRFS “fails badly” [11] without a multi-scale strategy, and the method of [9] becomes unstable without depth regularization (see Figure 2). Although regularization mechanisms somewhat circumvent such numerical instabilities 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 worthwhile. In the next section, we thus rewrite (1) as a nonlinear PDE.

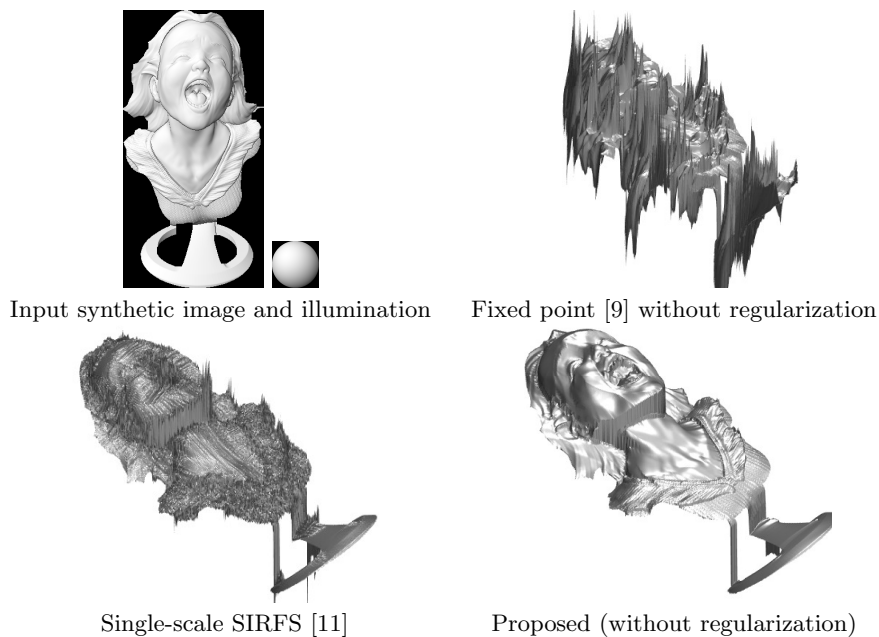


Fig. 2: Greylevel shape-from-shading using first-order spherical harmonics. Linearization strategies such as the fixed point one used in [9] fail if regularization is not employed. Similar issues arise in SIRFS [11] when the multi-scale approach is not used. Our SFS method can use regularization for disambiguation and for 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 our experiments, the albedo is assumed uniformly white and colored lighting is estimated from a gross surface approximation). These assumptions are usual in the SFS literature. They could be relaxed by simultaneously estimating shape, illumination and reflectance [11], 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 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 projections, the image formation model (1) can be rewritten as the following nonlinear PDE in z :*

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

with $z : \Omega \rightarrow \mathbb{R}$ a map characterizing the shape, $\nabla z : \Omega \rightarrow \mathbb{R}^2$ its gradient, and where $\mathbf{a}_{(\nabla z)}^c : \Omega \rightarrow \mathbb{R}^2$ and $b_{(\nabla z)}^c : \Omega \rightarrow \mathbb{R}$ are a vector field and a scalar field, respectively, which depend in a nonlinear way on ∇z .

Proof. The 3D-shape can be represented as a patch over the image domain, which associates each pixel $(x, y) \in \Omega$ to its conjugate 3D-point $\mathbf{x}(x, y)$ on the surface:

$$\mathbf{x} : \Omega \rightarrow \mathbb{R}^3$$

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

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, which relates the normal field to the depth map:

$$\mathbf{n} : \Omega \rightarrow \mathbb{S}^2 \subset \mathbb{R}^3$$

$$(x, y) \mapsto \frac{1}{d_{(\nabla z)}(x, y)} \begin{bmatrix} f \nabla z(x, y) \\ -1 - [\tilde{x}, \tilde{y}]^\top \cdot \nabla z(x, y) \end{bmatrix}, \quad (4)$$

where

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

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

$$d_{(\nabla z)} : \Omega \rightarrow \mathbb{R} \\ (x, y) \mapsto \sqrt{f^2 \|\nabla z(x, y)\|^2 + \left(1 + [\tilde{x}, \tilde{y}]^\top \cdot \nabla z(x, y)\right)^2}. \quad (6)$$

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:

$$\mathbf{a}_{(\nabla z)}^c : \Omega \rightarrow \mathbb{R}^2 \\ (x, y) \mapsto \frac{\rho^c(x, y)}{d_{(\nabla z)}(x, y)} \begin{bmatrix} f l_1^c - \tilde{x} l_3^c \\ f l_2^c - \tilde{y} l_3^c \end{bmatrix}, \quad (7)$$

$$b_{(\nabla z)}^c : \Omega \rightarrow \mathbb{R} \\ (x, y) \mapsto \rho^c \cdot \begin{bmatrix} \frac{-1}{d_{(\nabla z)}(x, y)} \\ 1 \\ \frac{f^2 z_x(x, y) z_y(x, y)}{(d_{(\nabla z)}(x, y))^2} \\ \frac{f z_x(x, y) (-1 - (\tilde{x}, \tilde{y}) \cdot \nabla z(x, y))}{(d_{(\nabla z)}(x, y))^2} \\ \frac{f z_y(x, y) (-1 - (\tilde{x}, \tilde{y}) \cdot \nabla z(x, y))}{(d_{(\nabla z)}(x, y))^2} \\ \frac{f^2 (z_x(x, y)^2 - z_y(x, y)^2)}{(d_{(\nabla z)}(x, y))^2} \\ \frac{3(-1 - (\tilde{x}, \tilde{y}) \cdot \nabla z(x, y))^2}{(d_{(\nabla z)}(x, y))^2} - 1 \end{bmatrix}. \quad (8)$$

□

When $C = 1$, the camera is orthographic and the lighting is directional and frontal (*i.e.*, $l_3 < 0$ is the only non-zero lighting component), then (2) becomes the *eikonal equation* $\frac{\rho |l_3|}{\sqrt{1 + \|\nabla z\|^2}} = I$. Efficient numerical methods for solving this nonlinear PDE have been suggested, using for instance semi-Lagrangian schemes [12]. Such techniques can also handle perspective camera projection and/or nearby point light source illumination [13]. 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 harmonic lighting, and greylevel or multi-channel images, has not been tackled so far. A variational solution to this generic SFS problem, which is inspired by the classical method of Horn and Brooks [2], is presented in the next section.

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:

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

In recent works on shading-based refinement [9], 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 frozen. Even in less trivial cases, Figure 2 shows that this strategy is unstable, which explains why regularization is employed in [9]. We will also resort to regularization later on, 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). Let us first sketch the proposed solver in the regularization-free case, and discuss its connection with Horn and Brooks' variational approach to SFS.

4.1 Horn and Brooks' Method Revisited

In [2], Horn and Brooks introduce a variational approach for solving the eikonal SFS model, which is a special case of (2). They promote a two-stages shape recovery method, which first estimates the gradient and then integrates it into a depth map. That is to say, the energy (9) is first minimized in terms of the gradient $\theta := \nabla z$, and then θ is integrated into a depth map z .

Since local gradient estimation is ambiguous, they put forward the introduction of the so-called integrability constraint for the first stage. Indeed, θ is the gradient of a function z : it must be a conservative field. This implies that it should be irrotational (zero-curl condition). Introducing the divergence operator $\nabla \cdot$, the latter condition reads

$$\underbrace{\left\| \nabla \cdot \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \theta \right\|_{\ell^2(\Omega)}}_{:=\mathcal{I}(\theta)} = 0. \quad (10)$$

In practice, they convert the hard-constraint (10) into a regularization term, introducing two hyper-parameters $(\lambda, \mu) > (0, 0)$ to balance the adequation to the images and the integrability of the estimated field:

$$\hat{\theta} = \underset{\theta: \Omega \rightarrow \mathbb{R}^2}{\operatorname{argmin}} \lambda \mathcal{E}(\theta; I) + \mu \mathcal{I}(\theta). \quad (11)$$

After solving (11), $\widehat{\theta}$ is integrated into a depth map z , by solving $\nabla z = \widehat{\theta}$. However, since integrability is not strictly enforced but only used as regularization, there is no guarantee for $\widehat{\theta}$ to be integrable. Therefore, Horn and Brooks recast the integration task as another variational problem (see [14] for an overview of this problem):

$$\min_{z: \Omega \rightarrow \mathbb{R}} \left\| \nabla z - \widehat{\theta} \right\|_{\ell^2(\Omega)}^2. \quad (12)$$

This two-stages approach consisting in solving (11), and then (12), is however prone to bias propagation: any error during gradient estimation may have dramatic consequences for the integration stage. We argue that such a sequential approach is not necessary. Indeed, θ is conservative *by construction*, and hence integrability should not even have to be invoked. We put forward an integrated approach, which infers shape clues from the image using local gradient estimation as in Horn and Brooks’ method, but which explicitly constrains the gradient to be conservative. That is to say, we simultaneously estimate the depth map and its gradient, by turning the minimization of (9) into a constrained variational problem:

$$\begin{aligned} \min_{\substack{\theta: \Omega \rightarrow \mathbb{R}^2 \\ z: \Omega \rightarrow \mathbb{R}}} \mathcal{E}(\theta; I) \\ \text{s.t. } \nabla z = \theta \end{aligned} \quad (13)$$

The variational problem (13) can be solved using an augmented Lagrangian approach. In comparison with Horn and Brooks’ method, this avoids tuning the hyper-parameter μ in (11), as well as bias propagation due to the two-stages approach. Besides, this approach is easily extended in order to handle regularization terms, if needed. In the next paragraph, we consider two types of regularization: one which represents prior knowledge of the surface, and another one which ensures its smoothness.

4.2 Regularized Variational Model

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

$$\mathcal{P}(z; z^0) = \left\| z - z^0 \right\|_{\ell^2(\Omega^0)}^2, \quad (14)$$

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 also 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 [15] 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:

$$\mathcal{S}(\nabla z) = \|d_{(\nabla z)}\|_{\ell^1(\Omega)}. \quad (15)$$

Putting altogether the pieces (9), (14) and (15), and using the same change of variable $\theta := \nabla z$ as in (13), we obtain the following constrained variational approach to shape-from-shading:

$$\begin{aligned} \min_{\substack{\theta: \Omega \rightarrow \mathbb{R}^2 \\ z: \Omega \rightarrow \mathbb{R}}} & \lambda \mathcal{E}(\theta; I) + \mu \mathcal{P}(z; z^0) + \nu \mathcal{S}(\theta) \\ \text{s.t.} & \nabla z = \theta \end{aligned} \quad (16)$$

which is a regularized version of the pure shape-from-shading model (13) 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 (16) yields a pure SFS model if $\mu = \nu = 0$, a depth denoising model similar to that in [15] if $\lambda = 0$ and $\Omega^0 = \Omega$, and a shading-aware joint depth refinement and completion if $\lambda > 0$, $\mu > 0$ and $\Omega^0 \subsetneq \Omega$.

4.3 Numerical Solution

The change of variable $\theta := \nabla z$ in (16) has a major advantage when it comes to numerical solving: it separates the difficulty induced by the nonlinearity (shape-from-shading model and minimal surface prior) from that induced by the global nature of the problem (dependency upon the depth gradient).

Optimization can then be carried out by alternating nonlinear, yet local gradient estimation and global, yet linear depth estimation. To this end, we make use of the ADMM procedure, a standard approach to constrained optimization which dates back to the 70s [16]. We refer the reader to [17] for a recent overview of this method.

The augmented Lagrangian functional associated to (16) is defined as follows:

$$\mathcal{L}_\beta(\theta, z, \Psi) = \lambda \mathcal{E}(\theta; I) + \mu \mathcal{P}(z; z^0) + \nu \mathcal{S}(\theta) + \langle \Psi, \nabla z - \theta \rangle + \frac{\beta}{2} \|\nabla z - \theta\|_2^2, \quad (17)$$

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

ADMM iterations are then written:

$$\theta^{(k+1)} = \underset{\theta}{\operatorname{argmin}} \mathcal{L}_{\beta^{(k)}}(\theta, z^{(k)}, \Psi^{(k)}), \quad (18)$$

$$z^{(k+1)} = \underset{z}{\operatorname{argmin}} \mathcal{L}_{\beta^{(k)}}(\theta^{(k+1)}, z, \Psi^{(k)}), \quad (19)$$

$$\Psi^{(k+1)} = \Psi^{(k)} + \beta^{(k)} \left(\nabla z^{(k+1)} - \theta^{(k+1)} \right). \quad (20)$$

where $\beta^{(k)}$ can be determined automatically [17].

Problem (18) is a pixelwise non-trivial optimization problem which is solved using a Newton method with an L-BFGS stepsize. As for (19), it is discretized by first-order, forward finite differences with Neumann boundary condition. This yields a linear least-squares problem whose normal equations provide a 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 (16) falls below 10^{-3} .

This ADMM algorithm can be interpreted as follows. During the θ -update (18), 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 [9], 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 -update (19), these surface normals are integrated into a new depth map, which should stay close to 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, taking inspiration from the recent studies on non-convex ADMM [18,19]. 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 [11], 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, as advised in [1].

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

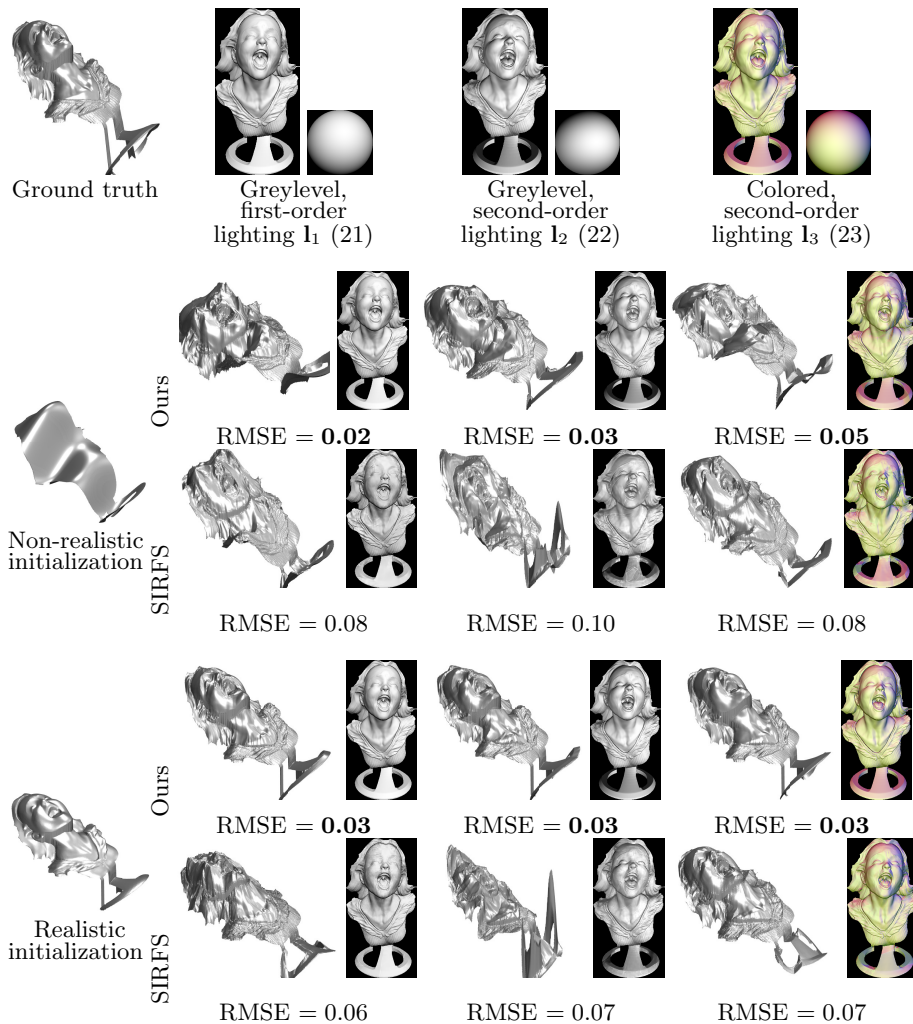


Fig. 3: Evaluation of our SFS approach against the multi-scale one from SIRFS [11], in three different lighting situations and using two different initial 3D-shapes (the first one is Matlab’s “peaks” function, while the second one is a smoothed version of the ground truth). For each experiment, we show 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.

second-order lighting vectors. Eventually, we consider a colored, second-order lighting vector. These lighting vectors are defined, respectively, by:

$$\mathbf{l}_1 = [0.1, -0.25, -0.7, 0.2, 0, 0, 0, 0, 0]^\top, \quad (21)$$

$$\mathbf{l}_2 = [0.2, 0.3, -0.7, 0.5, -0.2, -0.2, 0.3, 0.3, 0.2]^\top, \quad (22)$$

$$\mathbf{l}_3 = \begin{bmatrix} -0.2 & -0.2 & -1 & 0.4 & 0.1 & -0.1 & -0.1 & -0.1 & 0.05 \\ 0 & 0.2 & -1 & 0.3 & 0 & 0.2 & 0.1 & 0 & 0.1 \\ 0.2 & -0.2 & -1 & 0.2 & -0.1 & 0 & 0 & 0.1 & 0 \end{bmatrix}^\top. \quad (23)$$

To illustrate the underlying ambiguities, we consider two different initial 3D-shapes: 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 as soon as a good initialization is available.

In all the experiments, the images are better explained using our framework, which shows that the proposed numerical strategy 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.

In Figure 4, we illustrate the influence of the hyper-parameters μ and ν which control, respectively, the shape prior and the smoothness term. 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 might drift from the initial 3D-shape and interpret image noise as unwanted geometric artifacts. If $\mu \rightarrow +\infty$, the initial estimate (which exhibits reasonable low-frequency components, but no geometric detail) is not modified. If $\nu \rightarrow +\infty$, then only the minimal surface term matters, hence the result is over-smoothed. In this experiment, we also evaluate the accuracy of the 3D-reconstructions 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 since minimizing the latter comes down to estimating geometric details explaining the image noise.

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: in this case a low value of μ should be used. On the other hand, natural illumination is generally colored, so the three image channels provide redundant information: regularization is less important and a low value of the smoothness parameter ν can be used. We found that $(\lambda, \mu, \nu) = (1, 1, 5 \cdot 10^{-5})$ provides qualitatively nice results in all our real-world experiments.

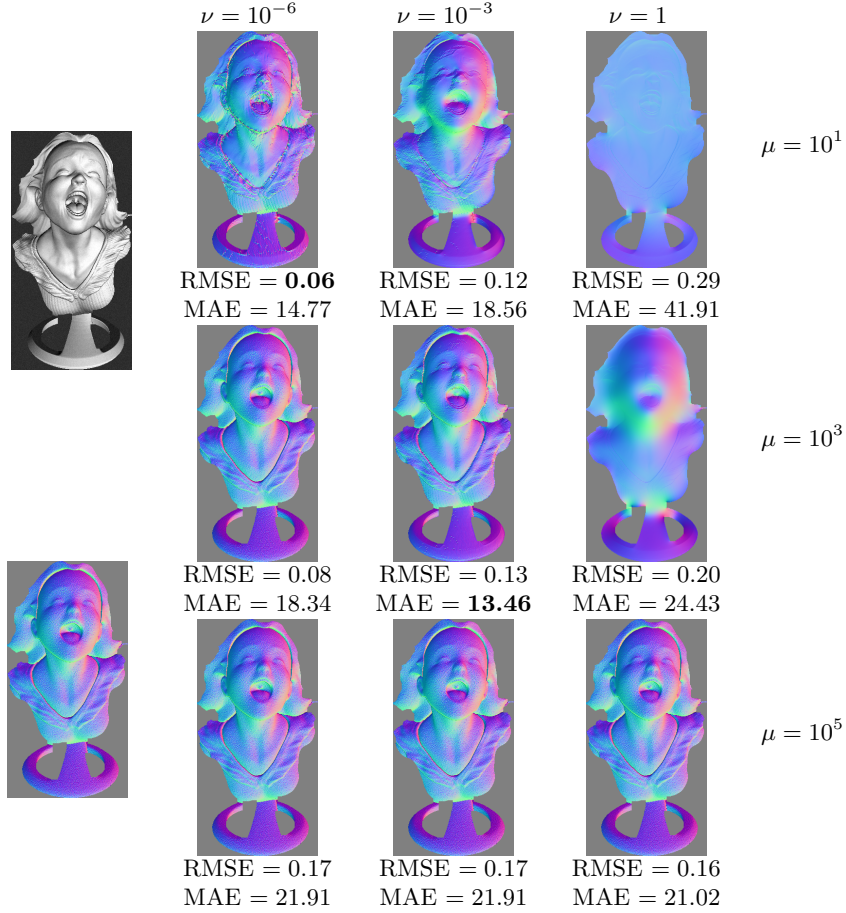


Fig. 4: Left: input noisy image ($\sigma_I = 2\%$ of the maximum greylevel) and noisy prior shape ($\sigma_z = 0.2\%$ of the maximum depth), represented by a normal map to emphasize the details. Right: estimated shape with $\lambda = 1$ and various values of μ and ν . The RMSE between the image and the reprojection is minimal when μ and ν are 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 both the inherent ambiguities of SFS, and the need for depth regularization.

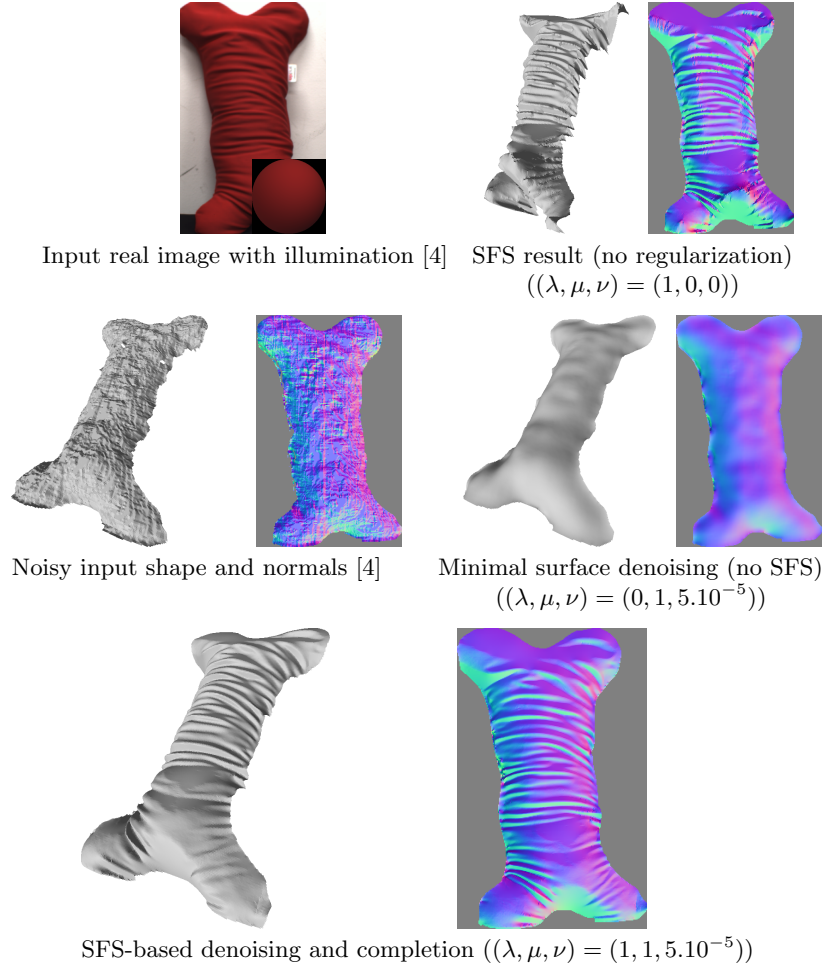


Fig. 5: Results on three computer vision problems: SFS, “blind” (not shading-based) depth refinement, and shading-based depth refinement. The shape estimated by SFS is distorted (due to the ambiguities of SFS), and artifacts occur (due to noise), but it contains the fine-scale details. Although the depth map provided by the RGB-D sensor is denoised without considering shading, thin structures are missed. With the proposed method, noise is removed and fine details are revealed.

In order to illustrate the practical disambiguation of SFS using a shape prior, we next consider as initialization $z^{(0)}$ and prior z^0 the depth provided by the RGB-D sensor. It is both 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 [15].

Eventually, Figure 6 demonstrates an application to stereovision, using a real-world dataset from [20]. This time, the initial depth map is obtained by a multi-view stereo (MVS) algorithm [21]. 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 MVS. 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 concavities are not well recovered, and the discontinuities are partly smoothed. Still, nice details are recovered, and the overall surface seems reasonable.

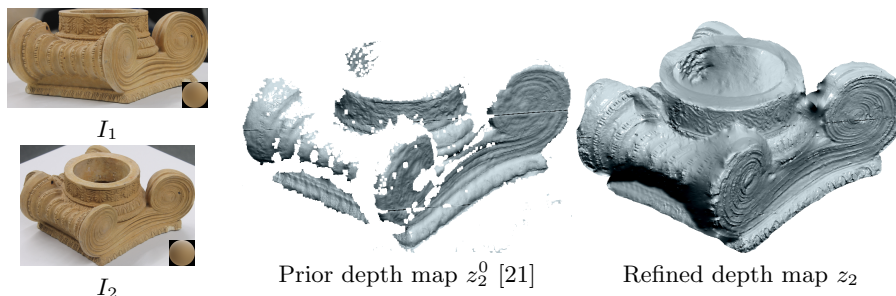


Fig. 6: Left: two (out of 30) images I_1 and I_2 of the “Figure” object [20]. Middle: depth map z_2^0 obtained by the CPMVS method [21] (before meshing). Right: refined and completed depth map z_2 .

6 Conclusion and Perspectives

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 models 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 disambiguation of SFS as well as practical 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.

References

1. Durou, J.D., Falcone, M., Sagona, M.: Numerical methods for shape-from-shading: A new survey with benchmarks. *Comput. Vis. Image Underst.* **109** (2008) 22–43
2. Horn, B.K.P., Brooks, M.J.: The variational approach to shape from shading. *Computer Vision, Graphics, and Image Processing* **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: *Proc. ICCV*. (2013)
5. Basri, R., Jacobs, D.P.: Lambertian reflectances and linear subspaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **25** (2003) 218–233
6. Quéau, Y., Mecca, R., Durou, J.D.: Unbiased Photometric Stereo for Colored Surfaces: A Variational Approach. In: *Proc. CVPR*. (2016)
7. Johnson, M.K., Adelson, E.H.: Shape estimation in natural illumination. In: *Proc. CVPR*. (2011)
8. Huang, R., Smith, W.A.P.: Shape-from-shading under complex natural illumination. In: *Proc. ICIP*. (2011)
9. Or-El, R., Rosman, G., Wetzler, A., Kimmel, R., Bruckstein, A.: RGBD-Fusion: Real-Time High Precision Depth Recovery. In: *Proc. CVPR*. (2015)
10. Frolova, D., Simakov, D., Basri, R.: Accuracy of spherical harmonic approximations for images of Lambertian objects under far and near lighting. In: *Proc. ECCV*. (2004)
11. Barron, J.T., Malik, J.: Shape, illumination, and reflectance from shading. *IEEE Trans. Pattern Anal. Mach. Intell.* **37** (2015) 1670–1687
12. Cristiani, E., Falcone, M.: Fast semi-Lagrangian schemes for the eikonal equation and applications. *SIAM J. Numer. Anal.* **45** (2007) 1979–2011
13. Breuß, M., Cristiani, E., Durou, J.D., Falcone, M., Vogel, O.: Perspective shape from shading: Ambiguity analysis and numerical approximations. *SIAM J. Imaging Sci.* **5** (2012) 311–342
14. Quéau, Y., Durou, J.D., Aujol, J.F.: Normal Integration: A Survey. *J. Math. Imaging Vision* (2017)
15. Graber, G., Balzer, J., Soatto, S., Pock, T.: Efficient minimal-surface regularization of perspective depth maps in variational stereo. In: *Proc. CVPR*. (2015)
16. Glowinski, R., Marroco, A.: Sur l’approximation, par éléments finis d’ordre un, et la résolution, par pénalisation-dualité d’une classe de problèmes de Dirichlet non linéaires. *Revue française d’automatique, informatique, recherche opérationnelle. Analyse numérique* **9** (1975) 41–76
17. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Found. Trends in Mach. Learn.* **3** (2011) 1–122
18. Li, G., Pong, T.K.: Global convergence of splitting methods for nonconvex composite optimization. *SIAM J. Optimiz.* **25** (2015) 2434–2460
19. Hong, M., Luo, Z.Q., Razaviyayn, M.: Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. *SIAM J. Optimiz.* **26** (2016) 337–364
20. 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
21. Jancosek, M., Pajdla, T.: Multi-view reconstruction preserving weakly-supported surfaces. In: *Proc. CVPR*. (2011)