

Photometric Depth Super-Resolution

Bjoern Haefner*, Songyou Peng*, Alok Verma*, Yvain Quéau, and Daniel Cremers

Abstract—This study explores the use of photometric techniques (shape-from-shading and uncalibrated photometric stereo) for upsampling the low-resolution depth map from an RGB-D sensor to the higher resolution of the companion RGB image. A single-shot variational approach is first put forward, which is effective as long as the target’s reflectance is piecewise-constant. It is then shown that this dependency upon a specific reflectance model can be relaxed by focusing on a specific class of objects (e.g., faces), and delegate reflectance estimation to a deep neural network. A multi-shot strategy based on randomly varying lighting conditions is eventually discussed. It requires no training or prior on the reflectance, yet this comes at the price of a dedicated acquisition setup. Both quantitative and qualitative evaluations illustrate the effectiveness of the proposed methods on synthetic and real-world scenarios.

Index Terms—RGB-D cameras, depth super-resolution, shape-from-shading, photometric stereo, variational methods, deep learning.



1 INTRODUCTION

RGB-D sensors have become very popular for 3D-reconstruction, in view of their low cost and ease of use. They deliver a colored point cloud in a single shot, but the resulting shape often misses thin geometric structures. This is due to noise, quantisation and, more importantly, the coarse resolution of the depth map. In comparison, the quality and resolution of the companion RGB image are substantially better. For instance, the Asus Xtion Pro Live device delivers 1280×1024 RGB images, but only up to 640×480 depth maps. The depth map thus needs to be up-sampled to the same resolution of the RGB image, and the latter could be analysed photometrically to reveal fine-scale details.

However, super-resolution of a solitary depth map without additional constraints is an ill-posed problem, and retrieving geometry from either a single color image (shape-from-shading) or from a sequence of color images acquired under unknown, varying lighting (uncalibrated photometric stereo) is another ill-posed problem. The present study explores the resolution of both these ill-posedness issues by jointly performing depth super-resolution and photometric 3D-reconstruction. We call this combined approach *photometric depth super-resolution*.

The choice of jointly solving both these classic inverse problems is motivated by the observation that ill-posedness in depth super-resolution and in photometric 3D-reconstruction have different peculiarities and origins. In depth super-resolution, constraints on high-frequency shape variations are missing (there exist infinitely many ways to interpolate between two measurements), while low-frequency (e.g., concave-convex or bas-relief) ambiguities

arise in photometric 3D-reconstruction. Therefore, the low-frequency geometric information necessary to disambiguate photometric 3D-reconstruction should be extracted from the low-resolution depth measurements and, symmetrically, the high-resolution photometric clues in the RGB data should provide the high-frequency information required to disambiguate depth super-resolution. One hand thus washes the other: ill-posedness in depth super-resolution is fought using photometric 3D-reconstruction, and vice-versa.

As we shall see in Section 2, the photometric depth super-resolution problem comes down to simultaneously inferring high-resolution depth and reflectance maps, given the low-resolution depth and the high-resolution RGB images. As depicted in Figure 1, this study explores three different strategies for such a task¹. The rest of this paper discusses them by increasing order of efficiency which, unfortunately, is inversely proportional to the amount of required resources. 1) If the available resources consist of a single RGB-D frame, then a variational approach to shape-from-shading can be followed. This approach, presented in Section 3, has no particular requirement in terms of acquisition setup or offline processing, yet it is effective only as long as the surface’s reflectance is piecewise-constant. 2) Section 4 then discusses a solution for eliminating this dependency upon a specific reflectance model. Pre-training a neural network for reflectance estimation allows to handle surfaces with more complex reflectance within the same variational framework. Yet, additional resources are required for offline training and the target has to resemble the objects used in the training phase (we thus focus in this section on human faces). 3) If multiple pairs of images can be acquired from the same viewing angle but under varying lighting, then one can resort to uncalibrated photometric stereo. This last strategy, discussed in Section 5, requires neither an assumption on the reflectance, nor offline training for a specific class of objects. However, it requires capturing more data online. Section 6 eventually recalls the main conclusions of this study and suggests future research directions.

* Equal contribution

- B. Haefner, A. Verma, and D. Cremers are with the Department of Computer Science, Technical University of Munich, 80333, Germany.
E-mail: {bjoern.haefner,alok.verma,cremers}@tum.de
- S. Peng is with Advanced Digital Sciences Center, University of Illinois at Urbana-Champaign, Singapore, 138602.
E-mail: songyou.peng@adsc-create.edu.sg
- Y. Quéau is with the GREYC laboratory, UMR CNRS 6072, Caen, France.
E-mail: yvain.queau@ensicaen.fr

Manuscript received 24 Sept. 2018; revised 30 May 2019.

1. Codes and data can be found in <https://vision.in.tum.de/data/datasets/photometricdepthsr>.


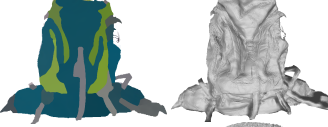


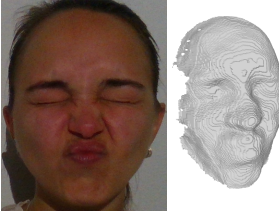




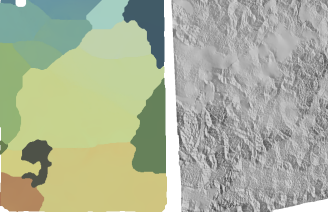
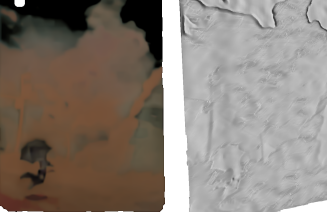

	Approach Required data Albedo	SfS (Section 3) 1 RGB-D frame Piecewise-constant	SfS + reflectance learning (Section 4) 1 RGB-D frame + training dataset Learned (e.g., faces)	UPS (Section 5) $n \geq 4$ RGB-D frames Arbitrary
Rucksack				
Face 1				
Tablecase				
	I z^0	ρ z	ρ z	ρ z

Fig. 1: Photometric depth super-resolution of a low-resolution depth map z^0 to the higher resolution of the companion image **I** (first column, Rucksack and Face 1 datasets were acquired using an Intel Realsense D415, and Tablecase using an Asus Xtion Pro Live). Second column: shape-from-shading (SfS) recovers high-resolution albedo (ρ) and depth (z) from a single RGB-D frame, assuming piecewise-constant albedo. If this assumption is not satisfied (e.g., Face 1 and Tablecase), shape estimation deteriorates. Third column: this can be circumvented by learning reflectance, an approach which is efficient as long as the target resembles the training data (here, training was carried out on human faces). Fourth column: uncalibrated photometric stereo (UPS) requires no training and handles arbitrary albedo, but it requires $n \geq 4$ input frames acquired under varying illumination. See Section 6 in the supplementary material for additional comparisons.

2 PROBLEM STATEMENT

A generic RGB-D sensor is considered, which consists of a depth sensor and an RGB camera with parallel optical axes and optical centers lying on a plane orthogonal to these axes (see Figure 2). The images of the surface on the focal planes of the depth and the color cameras are denoted respectively by $\Omega_{LR} \subset \mathbb{R}^2$ and $\Omega_{HR} \subset \mathbb{R}^2$. In a single shot, the RGB-D sensor provides two 2D-representations of the surface:

- A geometric one, taking the form of a mapping $z^0 : \Omega_{LR} \rightarrow \mathbb{R}$ between pixels in Ω_{LR} and the depth of their conjugate 3D-points on the surface;
- A photometric one, taking the form of a mapping $\mathbf{I} : \Omega_{HR} \rightarrow \mathbb{R}^3$ between pixels in Ω_{HR} and the radiance (relatively to the red, green and blue channels of the color camera) of their conjugate 3D-point.

In real-world scenarios, the sets Ω_{LR} and Ω_{HR} are discrete, and the cardinality $|\Omega_{LR}|$ of Ω_{LR} is lower than that $|\Omega_{HR}|$ of Ω_{HR} . To obtain the richest surface representation, one should thus project the depth measurements z^0 from Ω_{LR} to Ω_{HR} , i.e. estimate a new, high-resolution depth map $z : \Omega_{HR} \rightarrow \mathbb{R}$. To this end, we next introduce constraints arising from depth super-resolution and from photometric 3D-reconstruction.

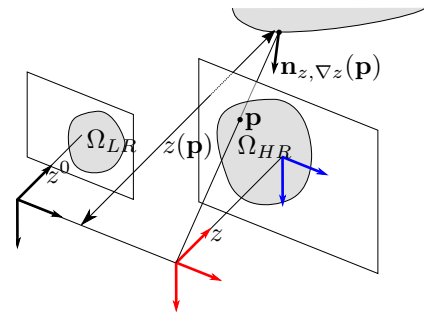


Fig. 2: Geometric setup. Depth measurements z^0 are available over a low-resolution set Ω_{LR} , and color measurements **I** over a high-resolution set Ω_{HR} . Photometric depth super-resolution consists in estimating a high-resolution depth map z out of these geometric and photometric measurements, which are connected through the surface normals $\mathbf{n}_{z, \nabla z}$, see Equations (1) to (3).

2.1 Geometric and Photometric Constraints

Given the assumptions above on the alignment of the sensors, and neglecting occlusions, the low-resolution depth map z^0 can be considered as a downsampled version of the sought high-resolution one z , after warping and averaging:

$$z^0 = Kz + \eta_z, \quad (1)$$

with η_z the realisation of a stochastic process representing measurement errors and quantisation, and K a non-invertible injective linear operator combining warping, blurring and downsampling [1], which can be calibrated beforehand [2]. Solving (1) in terms of the high-resolution depth map z constitutes the *depth super-resolution* problem, which requires additional assumptions on the smoothness of the observed surface. In this work, the latter is assumed regular, i.e. the normal to the surface exists in every visible point. Denoting by $f > 0$ the focal length of the color camera, and by $\mathbf{p} : \Omega_{HR} \rightarrow \mathbb{R}^2$ the field of pixel coordinates with respect to its principal point (blue reference coordinates system in Figure 2), the surface normal is defined as the following $\Omega_{HR} \rightarrow \mathbb{S}^2 \subset \mathbb{R}^3$ field of unit-length vectors (see e.g., [3]):

$$\mathbf{n}_{z,\nabla z} = \frac{1}{\sqrt{|f \nabla z|^2 + (-z - \mathbf{p}^\top \nabla z)^2}} \begin{bmatrix} f \nabla z \\ -z - \mathbf{p}^\top \nabla z \end{bmatrix}. \quad (2)$$

We further assume that the surface is Lambertian and lit by a collection of infinitely-distant point light sources. Lighting can then be represented in a compact manner using first-order spherical harmonics, see [4], [5] and Section 2.1 in the supplementary material. The irradiance in channel $\star \in \{R, G, B\}$ then writes

$$\mathbf{I} = \mathbf{l}^\top \underbrace{\begin{bmatrix} \mathbf{n}_{z,\nabla z} \\ 1 \end{bmatrix}}_{:= \mathbf{m}_{z,\nabla z}} \boldsymbol{\rho} + \boldsymbol{\eta}_{\mathbf{I}}, \quad (3)$$

with $\boldsymbol{\eta}_{\mathbf{I}} : \Omega_{HR} \rightarrow \mathbb{R}^3$ the realisation of a stochastic process standing for noise, quantisation and outliers, $\mathbf{l} \in \mathbb{R}^4$ the “light vector”, $\boldsymbol{\rho} : \Omega_{HR} \rightarrow \mathbb{R}^3$ the albedo (Lambertian reflectance) map and $\mathbf{m}_{z,\nabla z} : \Omega_{HR} \rightarrow \mathbb{R}^4$ a normal-dependent vector field. Solving (3) in terms of the high-resolution depth map z constitutes the *photometric 3D-reconstruction* problem, where reflectance $\boldsymbol{\rho}$ and lighting \mathbf{l} represent hidden variables to estimate.

Photometric depth super-resolution aims at inferring z out of z^0 and \mathbf{I} , while ensuring consistency with the super-resolution constraint in (1) and with the photometric one in (3). Before elaborating on three strategies for solving this problem, let us first review related works.

2.2 Related Works

Single depth image super-resolution requires solving Equation (1) in terms of the high-resolution depth map z . Since K is not invertible, this is an ill-posed problem: there exist infinitely many choices for interpolating between observations, cf. Section 2.2 in the supplementary material. Disambiguation can be carried out by adding observations obtained from different viewing angles [6], [7], [8]. In the more challenging case of a single viewing angle, a smoothness prior on the high-resolution depth map can be added and a variational approach can be followed [1]. One may also resort to machine learning techniques relying on a dictionary of low- and high-resolution depth or edge patches [9], [10]. Such a dictionary can even be constructed from a single depth image by looking for self-similarities [11], [12]. Nevertheless, learning-based depth super-resolution methods remain prone to over-fitting [13], which can be avoided by combining the respective merits of machine learning and variational approaches [14], [15].

Shape-from-shading [16], [17], [18], [19] is another classic inverse problem which aims at inferring shape from a single image of a scene, by inverting an image formation model such as (3). Common numerical strategies for this task include variational [20], [21] and PDE methods [22], [23], [24], [25]. However, even when reflectance and lighting are known, shape-from-shading is still ill-posed due to the underlying concave / convex ambiguity, cf. Section 2.2 in the supplementary material. Obviously, even more ambiguities arise under more realistic lighting and reflectance assumptions: any image can be explained by a flat shape illuminated uniformly but painted in a complex manner, by a white and frontally-lit surface with a complex geometry, or by a white planar surface illuminated in a complex manner [26]. Shape-from-shading under uniform reflectance but natural lighting has been studied [27], [28], [29], [30], but the case with unknown reflectance requires the introduction of additional priors [31]. This can be avoided by actively controlling the lighting, a variant of shape-from-shading known as photometric stereo which allows to estimate both shape and reflectance [32]. The problem with uncalibrated lighting is however ill-posed: it can be solved only up to a linear ambiguity [33] which, assuming integrability of the normals, reduces to a generalised bas-relief (GBR) one under directional lighting [34], and to a Lorentz one under natural lighting [35]. Resolution of such ambiguities by resorting to additional priors [36], [37], [38], extensions to non-Lambertian reflectance [39] and natural illumination [40] remain active research topics for which public benchmarks exist [41]. Recent developments in this field include PDE-based variational methods [42] and machine learning solutions [43], [44].

Shape-from-shading has recently gained new life with the emergence of RGB-D sensors. Indeed, the rough depth map can be used as prior to “guide” shape-from-shading and thus circumvent its ambiguities. This has been achieved in both the multi-view [45], [46], [47] and the single-shot [48], [49], [50], [51], [52], [53] cases. Still, the resolutions of the input image and depth map are assumed equal, and the same holds for approaches resorting to photometric stereo instead of shape-from-shading [54], [55], [56], [57]. In fact, depth super-resolution and photometric 3D-reconstruction have been widely studied, but rarely together. Several methods were proposed to coalign the depth edges in the super-resolved depth map with edges in the high-resolution color image [2], [58], [59], [60], [61], [62], but such approaches only consider sparse color features and may thus miss thin geometric structures. Some authors super-resolve the photometric stereo results [63], and others generate high-resolution images using photometric stereo [64], but none employ low-resolution depth clues except those of [65], who combine calibrated photometric stereo with structured light sensing. However, this involves a non-standard setup and careful lighting calibration, and reflectance is assumed to be uniform. Such issues are circumvented in the building blocks [66] and [67] of this study, which deal with photometric depth super-resolution based on, respectively, shape-from-shading and photometric stereo. Let us present the former approach, which is a single-shot solution to photometric depth super-resolution based on a variational approach to shape-from-shading.

3 SINGLE-SHOT DEPTH SUPER-RESOLUTION USING SHAPE-FROM-SHADING

In this section, the input data consists of a single RGB-D frame, i.e. a high-resolution image \mathbf{I} and a low-resolution depth map z^0 . To obtain a high-resolution depth map z consistent with both the geometric constraint (1) and the photometric one (3), we consider a variational approach which comes down to solving the optimization problem (10). Following [68], such a variational formulation can be derived from a Bayesian rationale.

3.1 Bayesian-to-Variational Rationale

Besides the high-resolution depth map z , neither the reflectance ρ nor the lighting vector \mathbf{l} is known. We treat the joint recovery of these three quantities as a maximum a posteriori (MAP) estimation problem. To this end we aim at maximising the posterior distribution of \mathbf{I} and z^0 which, according to Bayes rule, writes

$$\mathcal{P}(z, \rho, \mathbf{l} | z^0, \mathbf{I}) = \frac{\mathcal{P}(z^0, \mathbf{I} | z, \rho, \mathbf{l}) \mathcal{P}(z, \rho, \mathbf{l})}{\mathcal{P}(z^0, \mathbf{I})}. \quad (4)$$

In (4), the denominator is the evidence, which is a constant with respect to the variables z, ρ and \mathbf{l} and can thus be neglected during optimisation. The numerator is the product of the likelihood $\mathcal{P}(z^0, \mathbf{I} | z, \rho, \mathbf{l})$ and the prior distribution $\mathcal{P}(z, \rho, \mathbf{l})$, which both need to be further discussed.

The measurements of depth and image observations being done using separate sensors, z^0 and \mathbf{I} are statistically independent and thus the likelihood factors out as $\mathcal{P}(z^0, \mathbf{I} | z, \rho, \mathbf{l}) = \mathcal{P}(z^0 | z, \rho, \mathbf{l}) \mathcal{P}(\mathbf{I} | z, \rho, \mathbf{l})$. Furthermore, we assume that the process of how the depth map z^0 is acquired is depending neither on lighting \mathbf{l} nor on reflectance ρ . Given this, the marginal likelihood for the depth map z^0 can be written as $\mathcal{P}(z^0 | z, \rho, \mathbf{l}) = \mathcal{P}(z^0 | z)$. Assuming that noise η_z in (1) is homoskedastic, zero-mean and Gaussian-distributed with variance σ_z^2 , we further have $\mathcal{P}(z^0 | z) \propto \exp\left\{-\frac{\|Kz - z^0\|_2^2}{2\sigma_z^2}\right\}$ (here $\|\cdot\|_2$ is the ℓ^2 -norm over Ω_{LR}). Concerning the marginal likelihood of \mathbf{I} , we assume the random variable $\eta_{\mathbf{I}}$ in (3) follows a homoskedastic Gaussian distribution with zero mean and covariance matrix $\text{diag}(\sigma_I^2, \sigma_I^2, \sigma_I^2) \in \mathbb{R}^{3 \times 3}$, thus $\mathcal{P}(\mathbf{I} | z, \rho, \mathbf{l}) \propto \exp\left\{-\frac{\|\mathbf{l}^\top \mathbf{m}_{z, \nabla z} \rho - \mathbf{I}\|_2^2}{2\sigma_I^2}\right\}$ (this time, $\|\cdot\|_2$ is the ℓ^2 -norm over Ω_{HR}). Therefore, the likelihood in (4) is given by

$$\mathcal{P}(z^0, \mathbf{I} | z, \rho, \mathbf{l}) \propto \exp\left\{-\frac{\|Kz - z^0\|_2^2}{2\sigma_z^2} - \frac{\|\mathbf{l}^\top \mathbf{m}_{z, \nabla z} \rho - \mathbf{I}\|_2^2}{2\sigma_I^2}\right\}. \quad (5)$$

The prior distribution $\mathcal{P}(z, \rho, \mathbf{l})$ in (4) can be derived in a similar manner. The Lambertian assumption implies independence of reflectance from geometry and lighting, and the distant-light assumption implies independence of geometry and lighting. Therefore, z, ρ and \mathbf{l} are statistically independent and the prior distribution factors out as

$$\mathcal{P}(z, \rho, \mathbf{l}) = \mathcal{P}(z) \mathcal{P}(\rho) \mathcal{P}(\mathbf{l}). \quad (6)$$

Regarding lighting, we do not want to favor any particular situation and thus we opt for an improper prior:

$$\mathcal{P}(\mathbf{l}) = \text{constant}. \quad (7)$$

The prior on z is slightly more evolved. As we want to prevent oversmoothing (Sobolev regularisation) and/or staircasing artefacts (total variation regularisation), we make use of a minimal surface prior [69]. To this end, a parametrisation $d\mathcal{A}_{z, \nabla z} : \Omega_{HR} \rightarrow \mathbb{R}$ mapping each pixel to the corresponding area of the surface element is required. This writes $d\mathcal{A}_{z, \nabla z} = \frac{z}{f^2} \sqrt{|f \nabla z|^2 + (-z - \mathbf{p}^\top \nabla z)^2}$, and the total surface area is then given by $\|d\mathcal{A}_{z, \nabla z}\|_1$ (here $\|\cdot\|_1$ is the ℓ^1 -norm over Ω_{HR}). Introducing a free parameter $\alpha > 0$ to control the surface smoothness, the minimal surface prior can then be stated as

$$\mathcal{P}(z) \propto \exp\left\{-\frac{\|d\mathcal{A}_{z, \nabla z}\|_1}{\alpha}\right\}. \quad (8)$$

Following the Retinex theory [70], reflectance ρ can be assumed piecewise-constant, resulting in a Potts prior

$$\mathcal{P}(\rho) \propto \exp\left\{-\frac{\|\nabla \rho\|_0}{\beta}\right\}, \quad (9)$$

with $\beta > 0$ controlling the degree of discontinuities in the reflectance ρ . Note that ρ is a vector field, thus for each pixel \mathbf{p} , $\nabla \rho(\mathbf{p}) = [\nabla \rho_R(\mathbf{p}), \nabla \rho_G(\mathbf{p}), \nabla \rho_B(\mathbf{p})]^\top \in \mathbb{R}^{3 \times 2}$, and we use the following definition of the ℓ^0 -“norm” over Ω_{HR} : $\|\nabla \rho\|_0 := \sum_{\mathbf{p} \in \Omega_{HR}} \begin{cases} 0 & \text{if } |\nabla \rho(\mathbf{p})|_F = 0, \\ 1 & \text{else} \end{cases}$, with $|\cdot|_F$ the Frobenius norm over $\mathbb{R}^{3 \times 2}$.

The MAP estimate for depth, reflectance and lighting is eventually attained by maximising the posterior distribution (4) or, equivalently, minimising its negative logarithm. Plugging Equations (5) to (9) into (4), and discarding all additive constants, this comes down to solving the following variational problem:

$$\min_{z, \rho, \mathbf{l}} \left\| \mathbf{l}^\top \mathbf{m}_{z, \nabla z} \rho - \mathbf{I} \right\|_2^2 + \mu \|Kz - z^0\|_2^2 + \nu \|d\mathcal{A}_{z, \nabla z}\|_1 + \lambda \|\nabla \rho\|_0, \quad (10)$$

where the trade-off parameters (μ, ν, λ) are given by

$$\mu = \frac{\sigma_I^2}{\sigma_z^2}, \quad \nu = \frac{\sigma_I^2}{\alpha}, \quad \lambda = \frac{\sigma_I^2}{\beta}. \quad (11)$$

3.2 Numerical Solving of (10)

The variational problem in (10) is not only nonconvex, but also inherits a nonlinear dependency upon the gradient of z , see (3) along with (2). Compared to other methods, which overcome this issue by either following a two-step approach via optimising over the normals and then fitting an integrable surface to it [48] (a strategy which may fail if the estimated normals are non-integrable), or by freezing the nonlinearity [51] (which may yield convergence issues, in view of the nonconvexity of the optimisation problem), we solve for the depth directly and without any approximation. To this end we follow [30] and turn the global-and-nonlinear problem (10) into a sequence of global-yet-linear and nonlinear-yet-local ones. This can be achieved by introducing an auxiliary vector field $\theta : \Omega_{HR} \rightarrow \mathbb{R}^3$ with $\theta := (z, \nabla z)$ and rewriting (10) as the following equivalent constrained optimisation problem:

$$\begin{aligned} \min_{z, \rho, \mathbf{l}, \theta} & \left\| \mathbf{l}^\top \mathbf{m}_\theta \rho - \mathbf{I} \right\|_2^2 + \mu \|Kz - z^0\|_2^2 + \nu \|d\mathcal{A}_\theta\|_1 + \lambda \|\nabla \rho\|_0 \\ \text{s.t. } & \theta = (z, \nabla z). \end{aligned} \quad (12)$$

To solve the nonconvex, non-smooth and constrained optimisation problem (12) we make use of a multi-block ADMM scheme [71], [72], [73]. This comes down to iterating a sequence consisting of minimisations of the augmented Lagrangian

$$\mathcal{L}(z, \rho, \mathbf{l}, \boldsymbol{\theta}, \mathbf{u}) = \left\| \mathbf{l}^\top \mathbf{m}_\theta \rho - \mathbf{I} \right\|_2^2 + \mu \|Kz - z^0\|_2^2 + \nu \|\text{d}\mathcal{A}\theta\|_1 + \lambda \|\nabla \rho\|_0 + (\boldsymbol{\theta} - (z, \nabla z))^\top \mathbf{u} + \frac{\kappa}{2} \|\boldsymbol{\theta} - (z, \nabla z)\|_2^2 \quad (13)$$

over the primal variables z , ρ , \mathbf{l} and $\boldsymbol{\theta}$, and one gradient ascent step over the dual variable $\mathbf{u} : \Omega_{HR} \rightarrow \mathbb{R}^3$ ($\kappa > 0$ can be viewed as a step size).

At iteration (k) , one sweep of this scheme writes as:

$$\rho^{(k+1)} = \underset{\rho}{\operatorname{argmin}} \left\| \mathbf{l}^{(k)\top} \mathbf{m}_{\theta^{(k)}} \rho - \mathbf{I} \right\|_2^2 + \lambda \|\nabla \rho\|_0, \quad (14)$$

$$\mathbf{l}^{(k+1)} = \underset{\mathbf{l}}{\operatorname{argmin}} \left\| \mathbf{l}^\top \mathbf{m}_{\theta^{(k)}} \rho^{(k+1)} - \mathbf{I} \right\|_2^2, \quad (15)$$

$$\boldsymbol{\theta}^{(k+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left\| \mathbf{l}^{(k+1)\top} \mathbf{m}_\theta \rho^{(k+1)} - \mathbf{I} \right\|_2^2 + \nu \|\text{d}\mathcal{A}\theta\|_1 + \frac{\kappa}{2} \left\| \boldsymbol{\theta} - (z, \nabla z)^{(k)} + \mathbf{u}^{(k)} \right\|_2^2, \quad (16)$$

$$z^{(k+1)} = \underset{z}{\operatorname{argmin}} \mu \|Kz - z^0\|_2^2 + \frac{\kappa}{2} \left\| \boldsymbol{\theta}^{(k+1)} - (z, \nabla z) + \mathbf{u}^{(k)} \right\|_2^2, \quad (17)$$

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \boldsymbol{\theta}^{(k+1)} - (z, \nabla z)^{(k+1)}. \quad (18)$$

The albedo subproblem (14) is solved using the primal-dual algorithm [74]. The lighting update step in (15) is done using the pseudo-inverse. The $\boldsymbol{\theta}$ -update (16) is a nonlinear optimisation subproblem, yet free of neighboring pixel dependency thanks to the proposed splitting. It can be solved independently in each pixel using the implementation [75] of the L-BFGS method [76]. Eventually, the conjugate gradient method is applied on the normal equations of (17), which is a sparse linear least squares problem.

Our initial values for $(k) = (0)$ are chosen to be $\rho^{(0)} = \mathbf{I}$, $\mathbf{l}^{(0)} = [0, 0, -1, 0]^\top$, $z^{(0)}$ an inpainted [77] and smoothed [78] version of z^0 followed by bicubic interpolation to upsample to the image domain Ω_{HR} , $\boldsymbol{\theta}^{(0)} = (z, \nabla z)^{(0)}$, $\mathbf{u}^{(0)} = 0$ and $\kappa = 10^{-4}$. Due to the problem being non-smooth and nonconvex, to date no convergence result has been established and we leave this as future work. Nevertheless, in our experiments we have never encountered any problem reaching convergence, which we consider as reached if the relative residual falls below some threshold:

$$r_{\text{rel}} := \frac{\|z^{(k+1)} - z^{(k)}\|_2}{\|z^{(0)}\|_2} < 10^{-5}, \quad (19)$$

and if the constraint $\boldsymbol{\theta} = (z, \nabla z)$ is numerically satisfied, i.e.

$$r_c := \left(\boldsymbol{\theta}^{(k+1)} - (z, \nabla z)^{(k+1)} \right)^\top \mathbf{u}^{(k+1)} + \frac{\kappa}{2} \left\| \boldsymbol{\theta}^{(k+1)} - (z, \nabla z)^{(k+1)} \right\|_2^2 < 5 \cdot 10^{-6}. \quad (20)$$

To ensure the latter, the step size κ is multiplied by a factor of 2 after each iteration.

The scheme is implemented in Matlab, except the albedo update (14) which is implemented in CUDA. Depending on the datasets, convergence is reached between 10s and 90s.

3.3 Experiments

Although the optimal value of each parameter can be deduced using (11), it can be difficult to estimate the noise statistics in practice, thus we consider (μ, ν, λ) as tunable hyperparameters. We first carried out a series of experiments on synthetic datasets, which showed that the set of parameters $(\mu, \nu, \lambda) = (0.1, 0.7, 1)$ seems appropriate, cf. Section 3.2 in the supplementary material. Using these values, we then conducted qualitative and quantitative comparison of our results against the state-of-the-art single-shot approaches [10], [51], [60], on synthetic datasets and publicly available real-world ones from [41], [46], [47]. The proposed method appeared to represent the best compromise between the recovery of high- and low-frequency geometric information. These experimental results can be found in Sections 3.3 to 3.6 in the supplementary material.

Next, we qualitatively evaluated our approach on data we captured ourselves with an Intel RealSense D415 (1280×720 RGB and 320×240 depth) and an Asus Xtion Pro Live camera (1280×1024 RGB and 320×240 depth). Data was captured indoor with an LED attached to the camera in order to reinforce shading in the RGB images. The objects of interest were manually segmented from background before processing. Figure 3 shows the resulting estimates of ρ and z (1D depth profiles highlighting the recovery of thin structures can be found in Section 3.6 in the supplementary material). In the simplest ‘‘Android’’ experiment, all shading information is explained with geometry since the Potts prior prevents shading information being propagated into reflectance. The ‘‘Basecap’’ experiment is slightly more challenging due to the presence of areas with very low intensity. However, in such cases minimal surface ensures robustness, while fine details such as the stitches on the peak or the rivet of the bottle opener can still be recovered. The geometry of the 3-dimensional ‘‘GUINNESS’’ stitching is also correctly explained in terms of geometric variations and not as albedo. Although under- and over-segmentation of reflectance can be observed in the ‘‘Minion’’ experiment (cf. the eyes, the ‘‘Gru’’ logo in the center of the dungaree, or the left foot), this does not seem to affect depth estimation too much.

Another interesting qualitative result is the ‘‘Rucksack’’ experiment in Figure 1, where the very thin wrinkles are appropriately interpreted in terms of slight geometric variations. However, our method fails whenever the reflectance of the pictured object does not fit the Potts prior, see for instance the ‘‘Face 1’’ and ‘‘Tabletcase’’ experiments in Figure 1. For such objects with smoothly varying reflectance the piecewise-constant albedo assumption induces bias which propagates to the estimated depth. Indeed, the prior forbids to explain thin brightness variations in terms of reflectance, and thus the depth is forced to account for them, which results in noisy high-resolution depth maps. These failure cases illustrate the difficulty of designing a Bayesian prior which would properly split geometry and albedo information. The rest of this manuscript discusses two different strategies to circumvent this issue: by replacing the albedo estimation brick of the proposed variational framework with a deep neural network, or by acquiring additional data. The former approach is described in the next section.

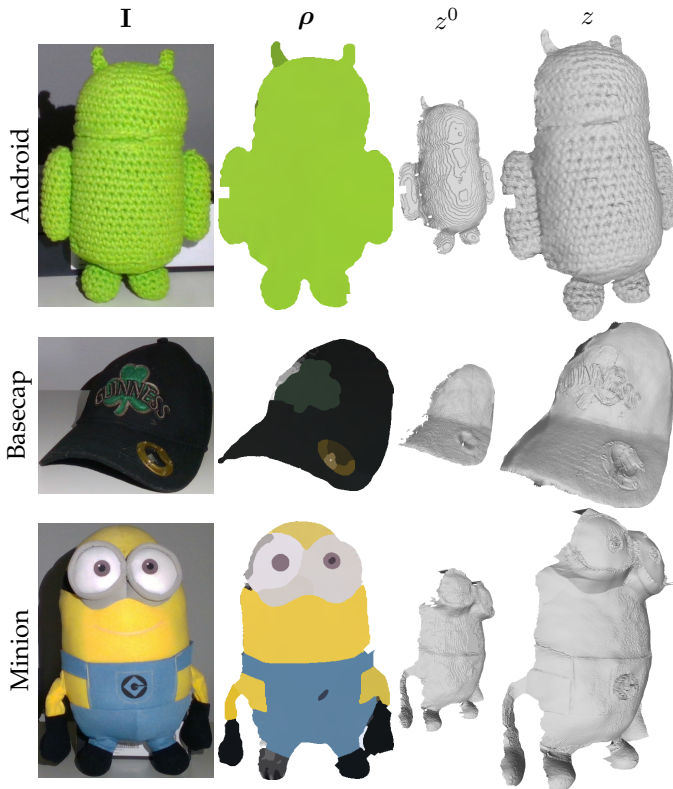


Fig. 3: Qualitative results obtained using the proposed single-shot approach on three real-world datasets captured with an Intel Realsense D415 camera. Even when intensity is very low (second row), or when under- or over-segmentation of reflectance happens (third row), the minimal surface prior prevents artefacts from arising while still allowing the recovery of thin geometric structures.

4 DEPTH SUPER-RESOLUTION USING SHAPE-FROM-SHADING AND REFLECTANCE LEARNING

The need for a strong prior on the target’s reflectance is a serious bottleneck in single-shot depth super-resolution using shape-from-shading. To circumvent this issue, we investigate in this section the combination of a deep learning strategy (to estimate reflectance) with a simplified version of the proposed variational framework (to carry out depth super-resolution, with pre-estimated reflectance).

4.1 Motivations and Construction of our Method

If we replace the assumption of a piecewise-constant albedo by the much stronger assumption of known albedo, the variational problem from the previous section comes down to jointly achieving depth super-resolution and low-order lighting estimation, and is thus substantially simplified. Yet, the task of designing a reflectance prior which is both realistic and numerically tractable is replaced with that of designing an efficient method for estimating a reflectance map out of a high-resolution RGB image. Luckily, this problem has long been investigated in the computer vision community: it is an intrinsic image decomposition problem. Some variational solutions exist [31], [79], yet they rely on explicit reflectance priors and thus suffer from the same

limitations as the previously proposed approach. One recent alternative is to rather resort to convolutional neural networks (CNNs), see for instance [80].

One important issue pertaining to CNN-based albedo estimation techniques is the lack of inter-class generalisation. Nevertheless, as long as the object to be analysed resembles those used during the training stage, the albedo estimates are satisfactory (see Section 2.3 in the supplementary material). Therefore, our proposal is to replace our man-made reflectance prior (piecewise-constantness) by a less explicit prior on the class of objects that the target belongs to. In this section, we focus on the class of human faces, as e.g., in [81], in view of both the richness of geometric details to recover and the complexity of the reflectance.

Let us emphasise that we resort to CNNs only for reflectance estimation and not for geometry refinement, although several deep learning strategies are able to provide shape clues [82], [83], [84], [85], [86]. Indeed, such methods have shown commendable results yet they are fraught with good-to-the-eye but possibly physically-incorrect geometry estimates, probably because during testing time they are unfettered by any concrete physics-based model and prior. Given that we do already have a physics-based depth refinement framework at hand, which furthermore makes use of the available low-resolution geometric clues from the depth sensor, we believe it is more sound to pick the best from both worlds - deep learning and variational methods. The solution we advocate thus contains two building blocks: a deep neural network prior-lessly learns the mapping from the input RGB image to reflectance for a particular class of objects (here, human faces), and then our variational framework based on shape-from-shading provides a physically-sound numerical framework for depth super-resolution.

4.2 Reflectance Learning

To train a CNN for the estimation of the human face’s reflectance, one needs at his disposal hundreds of facial images in vivid lighting and viewing conditions, along with the corresponding albedo maps (see Figure 4). This could be achieved using photometric stereo, yet the process would be very tedious. Training a neural network using synthetic images is a much simpler alternative: for instance, the approach from [87] resorts to the ShapeNet 3D-model library for estimating the albedo of inanimate objects. We follow a similar approach, but dedicated to human faces.

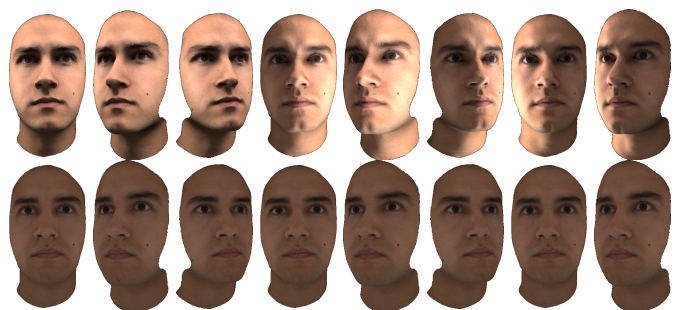


Fig. 4: Examples of human faces rendered under varying viewing and lighting conditions (top), along with the corresponding albedo maps (bottom).

We consider for this purpose the ICT-3DRFE database [88], [89], which comprises of 3D meshes of human faces, reflectance maps and normal maps. These databases were captured using a Light Stage, which provides fine-detailed shape and reflectance. Using a rendering software like Blender, one can then relight the faces and change viewing angles in order to obtain hundreds of shaded RGB images along with ground-truth albedo maps. Our training dataset consists of 21 faces, each enacting 15 different expressions. For each face and each expression, several images are acquired under varying lighting conditions induced by combining ten extended light sources. In practice, eight different lighting conditions are simulated by modulating the intensity of each light source, in accordance to the usual lighting in homes and offices e.g., light sources on the ceiling, walls, windows etc. Furthermore, rendering of the faces is done from three different viewing angles, i.e. center, slight left and slight right. Eventually, the images are generated using the Lambertian reflectance model. In total, after pruning the dataset and augmenting the faces for lighting, viewpoint and specularities, the training set comprises of 5175 images. Figure 4 shows some rendering examples, along with the corresponding ground-truth albedo maps.

A CNN is then trained to learn the mapping from the rendered face images to the corresponding ground-truth reflectance. Our network architecture is based on U-Net [90]. Generally, U-Net comprises of convolution and nonlinear layers which downsample the input to a 1D array and then upsample to the same input size using transpose convolution and nonlinear layers. Apart from these layers, an important architectural nuance of U-Net is the skip connections between downsampling and upsampling layers. This allows U-Net to produce sharp results, which is crucial for albedo estimation. Let us emphasise that the architecture of this network is remarkably simple, cf. Section 2.3 in the supplementary material. Once reflectance estimation is dropped out, the variational problem (10) for joint depth super-resolution and lighting estimation also becomes rather simple. Still, the appropriate combination of such simple frameworks does provide state-of-the-art results, as we shall see in the following.

4.3 Experiments

Since the numerical framework for estimating lighting and high-resolution depth is the same as the one discussed in Section 3, we use exactly the same parameters as in this section. Using these parameters, we carried out qualitative and quantitative comparison of our results against state-of-the-art methods which perform deep neural network-based depth super-resolution with the same kind of inputs as our method [62], and deep neural network-based shape-from-shading on low-resolution RGB data (without depth super-resolution) [86]. Our method appears to outperform the state-of-the-art both qualitatively and quantitatively on synthetic and publicly available real-world data from [41]. We also compared our reflectance learning-based approach with the previously discussed variational approach, and the learning-based method better refines the geometry of faces, which illustrates the benefit of dropping a handcrafted prior in favor of a more general learning framework (see Section 4.2 in the supplementary material).



Fig. 5: Results of the proposed variational approach to photometric depth super-resolution, using deep learning to estimate reflectance. Data was captured with an Intel Realsense D415 camera.

Next, we qualitatively evaluated our method on data we captured ourselves using an Intel RealSense D415 (1280 × 720 RGB and 640 × 360 depth). The results in Figure 5 illustrate the ability of the proposed approach to recover detail-preserving geometry with subtle wrinkles and teeth details, in contrast with pure deep learning methods which are less accurate (see Section 4.3 in the supplementary material). Eventually, comparing the result on the “Face 1” dataset (Figure 1) against the shape-from-shading result from Section 3 also confirms the interest of replacing a model-based prior by a learning framework. However, the “Rucksack” and “Tabletcase” experiments of Figure 1 also highlight the limitation of the proposed learning-based solution: whenever the object significantly departs from usual facial appearance, the reflectance fails and artifacts arise in the depth map. This can also be observed on objects from the DiLiGenT dataset [41] (see Section 4.4 in the supplementary material), although our approach still outperforms other learning-based ones. The only way to circumvent such an issue is to acquire more data in a photometric stereo manner, as discussed in the next section.

5 MULTI-SHOT DEPTH SUPER-RESOLUTION USING PHOTOMETRIC STEREO

Single-shot depth super-resolution requires some prior knowledge of the surface reflectance, either in terms of a piecewise-constant prior or of adequation to a learning database. The only way to get rid of such priors consists in acquiring multiple observations under varying lighting, i.e. performing uncalibrated photometric stereo.

Let us consider from now on a sequence of images $\{\mathbf{I}_i\}$, $i \in \{1, \dots, n\}$ and $n \geq 4$, captured under varying lighting conditions denoted by $\{\mathbf{l}_i\}$. The image formation model (3) is then turned into the following system of n equations:

$$\mathbf{I}_i = \mathbf{l}_i^\top \mathbf{m}_{z, \nabla z} \boldsymbol{\rho} + \boldsymbol{\eta}_i, \quad i \in \{1, \dots, n\}. \quad (21)$$

In (21), neither the depth z nor the reflectance map $\boldsymbol{\rho}$ depends on i . Hence, their estimation is much more constrained in comparison with shape-from-shading. Nevertheless, nescience of the lighting vectors $\{\mathbf{l}_i\}$ makes the joint estimation of shape, reflectance and lighting an ill-posed problem: as discussed in Section 2, the arising ambiguities cannot be resolved without the introduction of additional priors. As we shall see now, in the context of RGB-D sensing the need for such priors can be circumvented and a purely data-driven approach can be followed. In other words, the low-resolution depth information act as a natural disambiguation prior for uncalibrated photometric stereo and, equally, the tailored photometric based-prior implicitly ensures surface regularity for depth map super-resolution.

5.1 Maximum Likelihood-Based Solution

Let us recall that the single-shot approach discussed in Section 3 required priors on the regularity of both the depth and the reflectance maps. By considering *multiple* RGB-D frames $\{\mathbf{I}_i, z_i^0\}$, $i \in \{1, \dots, n\}$ of a static scene obtained under varying (though unknown) lighting, we hope to end up with a variational framework free of such man-made priors. To this end, we consider a maximum likelihood framework instead of a Bayesian one.

Considering again the independence of depth and image observations as well as the independence of shape from reflectance and lighting, the joint likelihood of the observations $\{\mathbf{I}_i, z_i^0\}$ can be factored out as follows:

$$\mathcal{P}(\{\mathbf{I}_i, z_i^0\} | z, \boldsymbol{\rho}, \{\mathbf{l}_i\}) = \mathcal{P}(\{\mathbf{I}_i\} | z, \boldsymbol{\rho}, \{\mathbf{l}_i\}) \mathcal{P}(\{z_i^0\} | z). \quad (22)$$

Under the assumption that the random variables $\boldsymbol{\eta}_i$ in (21) are homoskedastically distributed according to zero-mean Gaussian laws with the same covariance matrix $\text{diag}(\sigma_I^2, \sigma_I^2, \sigma_I^2)$, the marginal likelihood for $\{\mathbf{I}_i\}$ can be explicitly written as

$$\mathcal{P}(\{\mathbf{I}_i\} | z, \boldsymbol{\rho}, \{\mathbf{l}_i\}) \propto \exp \left\{ - \frac{\sum_i \|\mathbf{l}_i^\top \mathbf{m}_{z, \nabla z} \boldsymbol{\rho} - \mathbf{I}_i\|_2^2}{2\sigma_I^2} \right\}. \quad (23)$$

Assuming that the n low-resolution depth maps z_i^0 are consistent with the super-resolution model (1), and that the n corresponding random variables η_{z_i} follow a zero-mean Gaussian distribution with same variance σ_z^2 , the marginal likelihood for $\{z_i^0\}$ writes as

$$\mathcal{P}(\{z_i^0\} | z) \propto \exp \left\{ - \frac{\sum_i \|Kz - z_i^0\|_2^2}{2\sigma_z^2} \right\}. \quad (24)$$

Maximum likelihood estimation of depth, reflectance and lighting consists in maximising the joint likelihood (22) or, equivalently, minimising its negative logarithm. Neglecting all additive constants and plugging (23) and (24) into (22), this writes as the following variational problem:

$$\min_{z, \boldsymbol{\rho}, \{\mathbf{l}_i\}} \sum_i \|Kz - z_i^0\|_2^2 + \gamma \left\| \mathbf{l}_i^\top \mathbf{m}_{z, \nabla z} \boldsymbol{\rho} - \mathbf{I}_i \right\|_2^2, \quad (25)$$

with the trade-off parameter γ given by the ratio $\gamma = \frac{\sigma_z^2}{\sigma_I^2}$. Let us emphasise the simplicity of the photometric stereo-based variational model (25), in comparison with the one obtained using shape-from-shading, cf. (10). Although one may think that more data introduces more complexity to such problems, we can clearly see here that in fact Problem (25) is naturally easier by itself as it does not include non-smooth prior terms on the albedo and the depth, but only two data terms. As discussed next, this allows a much simpler numerical strategy to be followed.

5.2 Numerical Solving of (25)

Contrarily to the shape-from-shading problem (10), in (25) the nonlinearity arises only from the unit-length constraint on the normals. Therefore, we opt for a simpler numerical solution based on fixed point iterations. Considering (2) and (3), (25) can be rewritten as

$$\min_{z, \boldsymbol{\rho}, \{\mathbf{l}_i\}} \sum_i \|Kz - z_i^0\|_2^2 + \gamma \left\| \mathbf{l}_i^\top \begin{bmatrix} \tilde{\mathbf{n}}_{z, \nabla z} / d_{z, \nabla z} \\ 1 \end{bmatrix} \boldsymbol{\rho} - \mathbf{I}_i \right\|_2^2, \quad (26)$$

with $\mathbf{n}_{z, \nabla z} = \tilde{\mathbf{n}}_{z, \nabla z} / d_{z, \nabla z}$, where $d_{z, \nabla z}$ is a scalar field ensuring the unit-length constraint of the normals:

$$d_{z, \nabla z} = \sqrt{|f \nabla z|^2 + (-z - \mathbf{p}^\top \nabla z)^2}, \quad (27)$$

and $\tilde{\mathbf{n}}_{z, \nabla z}$ is a vector field encoding the normal direction:

$$\tilde{\mathbf{n}}_{z, \nabla z} = \begin{bmatrix} f \nabla z \\ -z - \mathbf{p}^\top \nabla z \end{bmatrix}. \quad (28)$$

In (26), only $d_{z, \nabla z}$ depends in a nonlinear way on the unknown depth z . Therefore, it seems natural to solve (26) iteratively, while freezing the nonlinearity (contrarily to the shape-from-shading case, in photometric stereo we experimentally found this fixed point strategy to be convergent, though we leave the convergence proof for future work). At iteration (k) and with the current estimates $(\boldsymbol{\rho}^{(k)}, \{\mathbf{l}_i^{(k)}\}, z^{(k)})$, one sweep of this scheme reads:

$$\boldsymbol{\rho}^{(k+1)} = \underset{\boldsymbol{\rho}}{\text{argmin}} \sum_i \left\| \mathbf{l}_i^{(k)\top} \begin{bmatrix} \tilde{\mathbf{n}}_{z^{(k)}, \nabla z^{(k)}} / d_{z^{(k)}, \nabla z^{(k)}} \\ 1 \end{bmatrix} \boldsymbol{\rho} - \mathbf{I}_i \right\|_2^2, \quad (29)$$

$$\mathbf{l}_i^{(k+1)} = \underset{\mathbf{l}_i}{\text{argmin}} \left\| \mathbf{l}_i^\top \begin{bmatrix} \tilde{\mathbf{n}}_{z^{(k)}, \nabla z^{(k)}} / d_{z^{(k)}, \nabla z^{(k)}} \\ 1 \end{bmatrix} \boldsymbol{\rho}^{(k+1)} - \mathbf{I}_i \right\|_2^2 \quad \forall i, \quad (30)$$

$$z^{(k+1)} = \underset{z}{\text{argmin}} \sum_i \|Kz - z_i^0\|_2^2 + \gamma \left\| \mathbf{l}_i^{(k+1)\top} \begin{bmatrix} \tilde{\mathbf{n}}_{z, \nabla z} / d_{z, \nabla z} \\ 1 \end{bmatrix} \boldsymbol{\rho}^{(k+1)} - \mathbf{I}_i \right\|_2^2. \quad (31)$$

All three problems (29), (30) and (31) are linear least-squares problems which we solve using the conjugate gradient method on the normal equations.

Our initial values for $(k) = (0)$ are chosen to be $\rho^{(0)} = \text{mean}(\{\mathbf{I}_i\})$, $\mathbf{I}_i^{(0)} = [0, 0, -1, 0]^\top \forall i$, and $z^{(0)}$ a smoothed version of $\text{mean}(\{z_i^0\})$ using the guided filter [78] followed by bicubic interpolation to upsample to the image domain Ω_{HR} . As in Section 3.2, to verify convergence we check if the relative residual r_{rel} falls below some threshold. In our experiments convergence was reached within at most 15 iterations, which corresponds to a few minutes in our Matlab implementation.

5.3 Experiments

We first considered synthetic datasets in order to experimentally determine appropriate values for the hyper-parameter γ and the number n of images. The values $\gamma = 0.01$ and $n \in [10, 30]$ were found to represent an appropriate compromise between accuracy and runtime (see Section 5.2 in the supplementary material). We then carried out qualitative and quantitative comparisons of our results against state-of-the-art uncalibrated photometric stereo [37], shading-based depth refinement using a low-resolution RGB image [51] and image-driven depth super-resolution using an anisotropic Huber-loss as regularisation term [1], [91]. Our approach was found to be the most effective on both synthetic and publicly available real-world datasets [41]. These experiments can be found in Sections 5.3 to 5.5 in the supplementary material.

Then, we carried out a qualitative evaluation of our results on data we captured ourselves using an Asus Xtion Pro Live (1280×1024 RGB and 320×240 depth) and an Intel Realsense D415 (1280×720 RGB and 640×480 depth). The setup is the same as in Section 3.3, just multiple images of the same static scene with static camera under varying lighting conditions are captured. Varying lighting was created by freely moving a handheld LED light source during the capturing process. From each image sequence, $n = 20$ high-resolution RGB images \mathbf{I}_i and low-resolution depth images z_i^0 were randomly extracted. Results are displayed in Figure 6. “Face 2” results are even more satisfactory compared to the deep learning-based approach in Figure 5, despite a small spike on the nose due to a small specular spot being present in every input image. Even the fine wrinkles and the buttons of the “Shirt” are recovered. The thin structures of the “Backpack” are appropriately recovered and the partly very low reflectance does not seem to deteriorate the depth estimate. The “Oven mitt” contains fine stitching structures which are successfully separated from the estimated albedo. The very fine geometric details of “Hat” are appropriately recovered in the depth, although some shading information remains visible in the reflectance. Interestingly, although our method is based on the Lambertian reflectance assumption, the high-quality shape of the reflective “Vase” can still be reconstructed and even where color is saturated at the specular regions, fine-scale geometric details are recovered. Eventually, among the three methods proposed in this article, only the uncalibrated photometric stereo-based approach can handle all three datasets in Figure 1, since reflectance is constrained neither to be piecewise-constant (“Rucksack”) nor to be that of a face (“Face 1”): the smoothly-varying albedo of the “Tabletcase” is appropriately estimated, and separated from the thin geometric wrinkle.



Fig. 6: Qualitative results of our uncalibrated photometric stereo-based method, on real-world data captured using a RealSense D415 (“Hat” and “Face 2”) or an Xtion Pro Live (five other datasets).

6 CONCLUSION

We investigated the use of photometric techniques for solving the depth super-resolution problem in RGB-D sensing. Three strategies were put forward: i) a shape-from-shading approach which requires a single RGB-D frame but is limited to objects exhibiting piecewise-constant reflectance, ii) a reflectance learning one which loosens this assumption by delegating reflectance estimation to a deep neural network trained on a specific class of objects such as faces, and iii) an uncalibrated photometric stereo setup which bypasses the need for albedo prior or training by acquiring additional data. These three approaches represent a continuum of solutions to photometric depth super-resolution with increasing level of accuracy, yet increasing amount of required resources.

This work may still be completed in several manners. First, the theoretical properties (proofs of convergence, existence and uniqueness of solutions, etc.) of the proposed numerical schemes may be explored. Second, all the methods presented here explicitly use the linear Lambertian image formation model: a natural line of future research would be to improve robustness to off-Lambertian effects such as specularities and cast-shadows, by resorting either to robust estimation techniques as in [42], or to non-Lambertian image formation models as in [92]. Eventually, the combination of deep learning and variational techniques might be further explored, for instance by devoting not only reflectance estimation to a deep neural network, but also lighting estimation as in [93]. Put together, these novelties could allow our approaches to handle more general surfaces as well as more general illumination conditions.

ACKNOWLEDGMENTS

The authors wish to thank Thomas Möllenhoff and Robert Maier for helpful discussions and comments.

REFERENCES

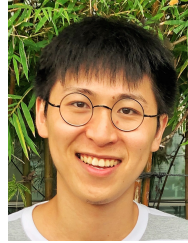
- [1] M. Unger, T. Pock, M. Werlberger, and H. Bischof, "A convex approach for variational super-resolution," in *Joint Pattern Recognition Symposium*, 2010, pp. 313–322.
- [2] J. Park, H. Kim, Y.-W. Tai, M. S. Brown, and I. S. Kweon, "High quality depth map upsampling for 3F-TOF cameras," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 1623–1630.
- [3] Y. Quéau, J.-D. Durou, and J.-F. Aujol, "Normal Integration: A Survey," *Journal of Mathematical Imaging and Vision*, vol. 60, no. 4, pp. 576–593, 2018.
- [4] R. Basri and D. P. Jacobs, "Lambertian reflectances and linear subspaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 218–233, 2003.
- [5] R. Ramamoorthi and P. Hanrahan, "An Efficient Representation for Irradiance Environment Maps," in *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*, 2001, pp. 497–500.
- [6] B. Goldlücke, M. Aubry, K. Kolev, and D. Cremers, "A super-resolution framework for high-accuracy multiview reconstruction," *International Journal of Computer Vision*, vol. 106, no. 2, pp. 172–191, 2014.
- [7] R. Maier, J. Stückler, and D. Cremers, "Super-resolution keyframe fusion for 3D modeling with high-quality textures," in *Proceedings of the International Conference on 3D Vision (3DV)*, 2015, pp. 536–544.
- [8] S. Schuon, C. Theobalt, J. Davis, and S. Thrun, "Lidarboost: Depth superresolution for TOF 3D shape scanning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 343–350.
- [9] O. Mac Aodha, N. D. F. Campbell, A. Nair, and G. J. Brostow, "Patch based synthesis for single depth image super-resolution," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2012, pp. 71–84.
- [10] J. Xie, R. S. Feris, and M.-T. Sun, "Edge-guided single depth image super resolution," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 428–438, 2016.
- [11] M. Hornáček, C. Rhemann, M. Gelautz, and C. Rother, "Depth super resolution by rigid body self-similarity in 3D," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 1123–1130.
- [12] J. Li, Z. Lu, G. Zeng, R. Gan, and H. Zha, "Similarity-aware patchwork assembly for depth image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 3374–3381.
- [13] J. Xie, R. S. Feris, S.-S. Yu, and M.-T. Sun, "Joint super resolution and denoising from a single depth image," *IEEE Transactions on Multimedia*, vol. 17, no. 9, pp. 1525–1537, 2015.
- [14] D. Ferstl, M. Rütter, and H. Bischof, "Variational depth super-resolution using example-based edge representations," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 513–521.
- [15] G. Riegler, M. Rütter, and H. Bischof, "ATGV-net: accurate depth super-resolution," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2016, pp. 268–284.
- [16] B. K. P. Horn, "Shape From Shading: A Method for Obtaining the Shape of a Smooth Opaque Object From One View," Ph.D. dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1970.
- [17] M. Breuß, E. Cristiani, J.-D. Durou, M. Falcone, and O. Vogel, "Perspective shape from shading: Ambiguity analysis and numerical approximations," *SIAM Journal on Imaging Sciences*, vol. 5, no. 1, pp. 311–342, 2012.
- [18] J.-D. Durou, M. Falcone, and M. Sagona, "Numerical Methods for Shape-from-shading: A New Survey with Benchmarks," *Computer Vision and Image Understanding*, vol. 109, no. 1, pp. 22–43, 2008.
- [19] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah, "Shape-from-shading: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 8, pp. 690–706, 1999.
- [20] B. K. P. Horn and M. J. Brooks, "The variational approach to shape from shading," *Computer Vision, Graphics, and Image Processing*, vol. 33, no. 2, pp. 174–208, 1986.
- [21] K. Ikeuchi and B. K. Horn, "Numerical shape from shading and occluding boundaries," *Artificial intelligence*, vol. 17, no. 1-3, pp. 141–184, 1981.
- [22] E. Cristiani and M. Falcone, "Fast semi-lagrangian schemes for the eikonal equation and applications," *SIAM Journal on Numerical Analysis*, vol. 45, no. 5, pp. 1979–2011, 2007.
- [23] M. Falcone and M. Sagona, "An algorithm for the global solution of the shape-from-shading model," in *Proceedings of the International Conference on Image Analysis and Processing (ICIAP)*, 1997, pp. 596–603.
- [24] P.-L. Lions, E. Rouy, and A. Tourin, "Shape-from-shading, viscosity solutions and edges," *Numerische Mathematik*, vol. 64, no. 1, pp. 323–353, 1993.
- [25] E. Rouy and A. Tourin, "A viscosity solutions approach to shape-from-shading," *SIAM Journal on Numerical Analysis*, vol. 29, no. 3, pp. 867–884, 1992.
- [26] E. H. Adelson and A. P. Pentland, *Perception as Bayesian inference*. Cambridge University Press, 1996, ch. The perception of shading and reflectance, pp. 409–423.
- [27] R. Huang and W. A. P. Smith, "Shape-from-shading under complex natural illumination," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2011, pp. 13–16.
- [28] M. K. Johnson and E. H. Adelson, "Shape estimation in natural illumination," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 2553–2560.
- [29] S. R. Richter and S. Roth, "Discriminative shape from shading in uncalibrated illumination," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1128–1136.
- [30] Y. Quéau, J. Mélou, F. Castan, D. Cremers, and J.-D. Durou, "A Variational Approach to Shape-from-shading Under Natural Illumination," in *Energy Minimization Methods for Computer Vision and Pattern Recognition (EMMCPVPR)*, 2017, pp. 342–357.

- [31] J. Barron and J. Malik, "Shape, illumination, and reflectance from shading," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 8, pp. 1670–1687, 2015.
- [32] R. J. Woodham, "Photometric Method for Determining Surface Orientation from Multiple Images," *Optical Engineering*, vol. 19, no. 1, pp. 139–144, 1980.
- [33] H. Hayakawa, "Photometric stereo under a light source with arbitrary motion," *Journal of the Optical Society of America A*, vol. 11, no. 11, pp. 3079–3089, 1994.
- [34] P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille, "The bas-relief ambiguity," *International Journal of Computer Vision*, vol. 35, no. 1, pp. 33–44, 1999.
- [35] R. Basri, D. W. Jacobs, and I. Kemelmacher, "Photometric stereo with general, unknown lighting," *International Journal of Computer Vision*, vol. 72, no. 3, pp. 239–257, 2007.
- [36] N. G. Aldrin, S. P. Mallick, and D. J. Kriegman, "Resolving the generalized bas-relief ambiguity by entropy minimization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.
- [37] T. Papadhimetri and P. Favaro, "A closed-form, consistent and robust solution to uncalibrated photometric stereo via local diffuse reflectance maxima," *International Journal of Computer Vision*, vol. 107, no. 2, pp. 139–154, 2014.
- [38] Y. Quéau, F. Lauze, and J.-D. Durou, "Solving Uncalibrated Photometric Stereo using Total Variation," *Journal of Mathematical Imaging and Vision*, vol. 52, no. 1, pp. 87–107, 2015.
- [39] F. Lu, X. Chen, I. Sato, and Y. Sato, "SympS: Brdf symmetry guided photometric stereo for shape and light source estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 1, pp. 221–234, 2018.
- [40] Z. Mo, B. Shi, F. Lu, S.-K. Yeung, and Y. Matsushita, "Uncalibrated photometric stereo under natural illumination," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 2936–2945.
- [41] B. Shi, Z. Mo, Z. Wu, D. Duan, S. K. Yeung, and P. Tan, "A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 271–284, 2019.
- [42] Y. Quéau, T. Wu, F. Lauze, J.-D. Durou, and D. Cremers, "A Non-Convex Variational Approach to Photometric Stereo under Inaccurate Lighting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 350–359.
- [43] S. Ikehata, "CNN-PS: CNN-based photometric stereo for general non-convex surfaces," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–18.
- [44] G. Chen, K. Han, B. Shi, Y. Matsushita, and K.-Y. K. Wong, "Self-calibrating deep photometric stereo networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [45] G. Choe, J. Park, Y.-W. Tai, and I. S. Kweon, "Refining geometry from depth sensors using IR shading images," *International Journal of Computer Vision*, vol. 122, no. 1, pp. 1–16, 2017.
- [46] R. Maier, K. Kim, D. Cremers, J. Kautz, and M. Nießner, "Intrinsic3d: High-quality 3D reconstruction by joint appearance and geometry optimization with spatially-varying lighting," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 3114–3122.
- [47] M. Zollhöfer, A. Dai, M. Innman, C. Wu, M. Stamminger, C. Theobalt, and M. Nießner, "Shading-based refinement on volumetric signed distance functions," *ACM Transactions on Graphics*, vol. 34, no. 4, pp. 96:1–96:14, 2015.
- [48] Y. Han, J.-Y. Lee, and I. S. Kweon, "High Quality Shape from a Single RGB-D Image under Uncalibrated Natural Illumination," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2013, pp. 1617–1624.
- [49] K. Kim, A. Torii, and M. Okutomi, "Joint estimation of depth, reflectance and illumination for depth refinement," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 199–207.
- [50] R. Or-El, R. Hershkovitz, A. Wetzler, G. Rosman, A. M. Bruckstein, and R. Kimmel, "Real-time depth refinement for specular objects," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 4378–4386.
- [51] R. Or-El, G. Rosman, A. Wetzler, R. Kimmel, and A. Bruckstein, "RGBD-Fusion: Real-Time High Precision Depth Recovery," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 5407–5416.
- [52] C. Wu, M. Zollhöfer, M. Nießner, M. Stamminger, S. Izadi, and C. Theobalt, "Real-time shading-based refinement for consumer depth cameras," *ACM Transactions on Graphics*, vol. 33, no. 6, pp. 200:1–200:10, 2014.
- [53] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, and S. Lin, "Shading-based shape refinement of RGB-D images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 1415–1422.
- [54] R. Anderson, B. Stenger, and R. Cipolla, "Augmenting depth camera output using photometric stereo," in *Proceedings of the IAPR Conference on Machine Vision Applications (MVA)*, 2011, pp. 369–372.
- [55] A. Chatterjee and V. Madhav Govindu, "Photometric refinement of depth maps for multi-albedo objects," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 933–941.
- [56] L. Xie, Y. Xu, X. Zhang, W. Bao, C. Tong, and B. Shi, "A self-calibrated photo-geometric depth camera," *The Visual Computer*, 2018.
- [57] Y. Zhang, Q. Zhang, and W. Feng, "High-Resolution Depth Refinement by Photometric and Multi-shading Constraints," in *PRICAI 2018: Trends in Artificial Intelligence*, 2018, pp. 201–209.
- [58] J. Diebel and S. Thrun, "An application of Markov random fields to range sensing," in *Advances in Neural Information Processing Systems*, 2006, pp. 291–298.
- [59] D. Ferstl, C. Reinbacher, R. Ranftl, M. Rütther, and H. Bischof, "Image guided depth upsampling using anisotropic total generalized variation," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2013, pp. 993–1000.
- [60] Q. Yang, R. Yang, J. Davis, and D. Nistér, "Spatial-depth super resolution for range images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.
- [61] B. Li, Y. Zhou, Y. Zhang, and A. Wang, "Depth image super-resolution based on joint sparse coding," *Pattern Recognition Letters*, 2019, (in press).
- [62] T.-W. Hui, C. C. Loy, and X. Tang, "Depth map super-resolution by deep multi-scale guidance," in *Proceedings of European Conference on Computer Vision (ECCV)*, 2016, pp. 353–369.
- [63] P. Tan, S. Lin, and L. Quan, "Subpixel photometric stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 8, pp. 1460–1471, 2008.
- [64] S. Chaudhuri and M. V. Joshi, *Motion-free super-resolution*. Springer Verlag, 2005.
- [65] Z. Lu, Y.-W. Tai, F. Deng, M. Ben-Ezra, and M. S. Brown, "A 3D imaging framework based on high-resolution photometric-stereo and low-resolution depth," *International Journal of Computer Vision*, vol. 102, no. 1-3, pp. 18–32, 2013.
- [66] B. Haefner, Y. Quéau, T. Möllenhoff, and D. Cremers, "Fight ill-posedness with ill-posedness: Single-shot variational depth super-resolution from shading," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 164–174.
- [67] S. Peng, B. Haefner, Y. Quéau, and D. Cremers, "Depth super-resolution meets uncalibrated photometric stereo," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, 2017, pp. 2961–2968.
- [68] D. Mumford, "Bayesian rationale for the variational formulation," in *Geometry-driven diffusion in computer vision*, 1994, pp. 135–146.
- [69] G. Graber, J. Balzer, S. Soatto, and T. Pock, "Efficient minimal-surface regularization of perspective depth maps in variational stereo," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 511–520.
- [70] E. H. Land, "The retinex theory of color vision," *Scientific American*, vol. 237, no. 6, pp. 108–120, 1977.
- [71] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [72] J. Eckstein and D. P. Bertsekas, "On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1, pp. 293–318, 1992.
- [73] R. Glowinski and A. Marroco, "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*, vol. 9, no. R2, pp. 41–76, 1975.

- [74] E. Strekalovskiy and D. Cremers, "Real-time minimization of the piecewise smooth Mumford-Shah functional," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2014, pp. 127–141.
- [75] M. Schmidt, "minFunc: unconstrained differentiable multivariate optimization in Matlab," 2005, <http://www.cs.ubc.ca/~schmidtm/Software/minFunc.html>.
- [76] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Mathematical programming*, vol. 45, no. 1, pp. 503–528, 1989.
- [77] "inpaint_nans," 2012, https://fr.mathworks.com/matlabcentral/fileexchange/4551-inpaint_nans.
- [78] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 6, pp. 1397–1409, 2013.
- [79] J. Shen, X. Yang, Y. Jia, and X. Li, "Intrinsic images using optimization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 3481–3487.
- [80] Q. Fan, J. Yang, G. Hua, B. Chen, and D. Wipf, "Revisiting deep intrinsic image decompositions," in *Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 8944–8952.
- [81] C. Li, K. Zhou, and S. Lin, "Intrinsic face image decomposition with human face priors," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2014, pp. 218–233.
- [82] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," in *Advances in Neural Information Processing Systems*, 2014, pp. 2366–2374.
- [83] G. Trigeorgis, P. Snape, I. Kokkinos, and S. Zafeiriou, "Face normals "in-the-wild" using fully convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 38–47.
- [84] Z. Shu, E. Yumer, S. Hadap, K. Sunkavalli, E. Shechtman, and D. Samaras, "Neural face editing with intrinsic image disentangling," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5444–5453.
- [85] N. Wang, Y. Zhang, Z. Li, Y. Fu, W. Liu, and Y.-G. Jiang, "Pixel2mesh: Generating 3d mesh models from single rgb images," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.
- [86] S. Sengupta, A. Kanazawa, C. D. Castillo, and D. W. Jacobs, "SfSNet: Learning Shape, Reflectance and Illuminance of Faces in the Wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 6296–6305.
- [87] J. Shi, Y. Dong, H. Su, and S. X. Yu, "Learning non-lambertian object intrinsics across shapenet categories," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5844–5853.
- [88] W.-C. Ma, T. Hawkins, P. Peers, C.-F. Chabert, M. Weiss, and P. Debevec, "Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination," in *Proceedings of the 18th Eurographics Conference on Rendering Techniques*, 2007, pp. 183–194.
- [89] G. Stratou, A. Ghosh, P. Debevec, and L. Morency, "Effect of illumination on automatic expression recognition: A novel 3d relightable facial database," in *Face and Gesture*, 2011, pp. 611–618.
- [90] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.
- [91] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof, "Anisotropic Huber-L1 Optical Flow," in *Proceedings of the British Machine Vision Conference*, 2009, pp. 108.1–108.11.
- [92] L. Chen, Y. Zheng, B. Shi, A. Subpa-Asa, and I. Sato, "A microfacet-based reflectance model for photometric stereo with highly specular surfaces," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 3162–3170.
- [93] M.-A. Gardner, K. Sunkavalli, E. Yumer, X. Shen, E. Gambaretto, C. Gagné, and J.-F. Lalonde, "Learning to predict indoor illumination from a single image," *ACM Transactions on Graphics*, vol. 36, no. 6, pp. 176:1–176:14, 2017.



Bjoern Haefner received his B.Sc. in Mathematics from the OTH Regensburg in 2013 and his M.Sc. in Mathematics in Science and Engineering from the Technical University of Munich in 2016. Since mid November 2016, he is a full-time PhD student in the Computer Vision and Artificial Intelligence chair at the Technical University of Munich. His research interests include RGB-D data processing for 3D reconstruction using variational methods.



Songyou Peng received the Erasmus Mundus M.Sc. in Computer Vision and Robotics in 2017. Between 2016 and 2017, he spent some time doing research at INRIA Grenoble and Technical University of Munich. Since 2018 he is a research engineer at Advanced Digital Sciences Center in Singapore. His research interests are computer vision and machine learning.



Alok Verma is pursuing a Master's degree in Biomedical Computing at the Technical University of Munich, Germany since 2017. Previously he worked as a senior electrical and software engineer at Philips Healthcare, Bangalore, India focusing on C-Arm X-ray Systems. His research interests are computer vision and deep learning for medical and non-medical images.



Yvain Quéau received his Ph.D from INP-ENSEEIH, Université de Toulouse, in 2015. From 2016 to 2018 he was a postdoctoral researcher in Technical University Munich, Germany, and then an associate professor with ISEN Brest, France. Since 2018 he is a CNRS researcher with the GREYC laboratory, Université de Caen, France. His research focuses on variational methods for solving inverse problems in computer vision.



Daniel Cremers received the PhD degree in computer science from the University of Mannheim, Germany. Subsequently, he spent two years as a postdoctoral researcher with UCLA and one year as a permanent researcher at Siemens Corporate Research, Princeton. From 2005 until 2009, he was associate professor with the University of Bonn. Since 2009 he holds the Chair of Computer Vision and Artificial Intelligence at the Technical University of Munich. He received numerous awards including the Gottfried-Wilhelm Leibniz Award 2016, the biggest award in German academia.