Fight ill-posedness with ill-posedness:
Single-shot variational depth super-resolution from shading

Bjoern Haefner Yvain Quéau Thomas Möllenhoff Daniel Cremers
Department of Informatics, Technical University of Munich, Germany
{bjoern.haefner,yvain.queau,thomas.moellenhoff,cremers}@tum.de

Abstract

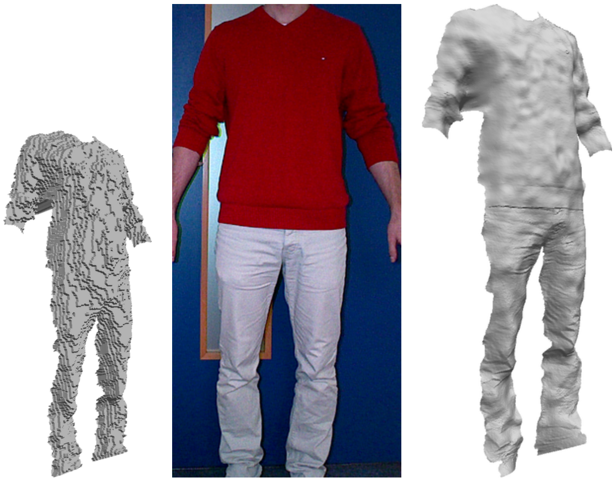
We put forward a principled variational approach for up-sampling a single depth map to the resolution of the companion color image provided by an RGB-D sensor. We combine heterogeneous depth and color data in order to jointly solve the ill-posed depth super-resolution and shape-from-shading problems. The low-frequency geometric information necessary to disambiguate shape-from-shading is extracted from the low-resolution depth measurements and, symmetrically, the high-resolution photometric clues in the RGB image provide the high-frequency information required to disambiguate depth super-resolution.

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. However, super-resolution of a solitary depth map without additional constraint is an ill-posed problem.

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 px^2 RGB images, but only up to 640×480 px^2 depth maps. Therefore, it seems natural to rely on color to refine depth. Yet, retrieving geometry from a single color image is another ill-posed problem, called shape-from-shading. Besides, combining it with depth clues requires the RGB and depth images to have the same resolution.

The resolution of the depth map thus remains a limiting factor in single-shot RGB-D sensing. This work aims at breaking this barrier by jointly refining and upsampling the depth map using shape-from-shading. In other words, **we fight the ill-posedness of single depth image super-resolution using shape-from shading, and vice-versa**.



Input: one low-resolution depth map and its companion high-resolution image

Our result

Figure 1: We carry out single-shot depth super-resolution for commodity RGB-D sensors, using shape-from-shading. By combining low-resolution depth (left) and high-resolution color clues (middle), detail-preserving super-resolution is achieved (right). All figures best viewed in the electronic version.

The rest of this paper is organized as follows. Section 2 reviews the single depth image super-resolution and shape-from-shading problems, in order to motivate their joint solving in the context of RGB-D sensing. Section 3 then introduces a principled Bayesian approach to joint depth super-resolution and shape-from shading. This yields a nonconvex variational problem which is solved using a dedicated ADMM algorithm. Our approach is evaluated against a broad variety of real-world datasets in Section 4, and our conclusions are eventually drawn in Section 5.

2. Motivation and related work

Let us first recall the ambiguities arising in single depth image super-resolution and in shape-from-shading, and how they have been handled in the literature.

2.1. Ill-posedness in single depth image super-resolution

A depth map is a function which associates to each 2D point of the image plane, the third component of its conjugate 3D-point, relatively to the camera coordinate system. Depth sensors provide out-of-the-box samples of the depth map over a discrete low-resolution rectangular 2D grid $\Omega_{\text{LR}} \subset \mathbb{R}^2$. We will denote by $z_0 : \Omega_{\text{LR}} \to \mathbb{R}$, $p \mapsto z_0(p)$ such a mapping between a pixel p and the measured depth value $z_0(p)$.

Due to hardware constraints, the depth observations z_0 are limited by the resolution of the sensor (*i.e.*, the number of pixels in Ω_{LR}). The single depth image super-resolution problem consists in estimating a high-resolution depth map $z : \Omega_{\text{HR}} \to \mathbb{R}$ over a larger domain $\Omega_{\text{HR}} \supset \Omega_{\text{LR}}$, which coincides with the low-resolution observations z_0 over Ω_{LR} once it is downsampled. Following [14], this can be formally written as

$$z_0 = Kz + \eta_z. \tag{1}$$

In (1), $K : \mathbb{R}^{\Omega_{\text{HR}}} \to \mathbb{R}^{\Omega_{\text{LR}}}$ is a linear operator combining warping, blurring and downsampling [55]. It can be calibrated beforehand, hence assumed to be known, see for instance [44]. As for η_z, it stands for the realisation of some stochastic process representing measurement errors, quantisation, etc.

Single depth image super-resolution requires solving Equation (1) in terms of the high-resolution depth map z. However, K in (1) maps from a high-dimensional space Ω_{HR} to a low-dimensional one Ω_{LR}, hence it cannot be inverted. Single depth image (blind) super-resolution is thus an ill-posed problem, as there exist infinitely many choices for interpolating between observations, as sketched in Figure 2. Therefore, one must find a way to constrain the problem, as well as to handle noise. This can be achieved by adding observations obtained from different viewing angles [20, 40, 53], but in this work we rather target single-shot applications.

When the input consists in a solitary depth map, disambiguation can be carried out by introducing a smoothness prior on the high-resolution depth map, a strategy which has led to a number of variational approaches, see for instance [55]. More recently, several machine learning approaches have been put forward, which essentially rely on a dictionary of low- and high-resolution depth or edge patches [38, 58]. To avoid resorting to a database, such a dictionary can be constructed from a single depth image by looking for self-similarities [27, 34]. Nevertheless, learning-based depth super-resolution methods remain prone to over-fitting, an issue which has been specifically tackled in [59]. Over-fitting can also be avoided by combining the respective benefits of machine learning and variational approaches [17, 50].

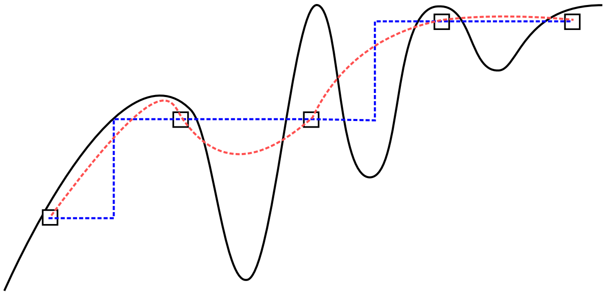


Figure 2: There exist infinitely many ways (dashed lines) to interpolate between low-resolution depth samples (rectangles). Our disambiguation strategy builds upon shape-from-shading applied to the companion high-resolution color image (*cf.* Figure 3), in order to resurrect the fine-scale geometric details of the genuine surface (solid line).

In the RGB-D framework, a high-resolution color image is also available. It can be used as a "guide" to interpolate missing depth values. Several methods were thus proposed to coalign the depth edges in the super-resolved map with edges of the given high-resolution color image [11, 16, 44, 60]. Yet, such approaches only consider sparse features in the high-resolution data, although the whole color image actually conveys shape clues. Indeed, brightness is directly related to the local orientation, hence a photometric approach to depth super-resolution for RGB-D sensors should be feasible and permit to recover fine-scale geometric details. There is, however, surprisingly few works in that direction: to the best of our knowledge, this has been achieved only in [37, 45], but these methods rely on a sequence of images acquired under varying lighting, hence they do not tackle the single-shot problem.

2.2. Ill-posedness in shape-from-shading

Shape-from-shading [25] is another classical inverse problem which aims at inferring shape from a single graylevel or color image of a scene. It consists in inverting an image formation model relating the image irradiance I to the scene radiance $\mathcal{R}$, which depends on the surface shape (represented here by the depth map z), the incident lighting l and the surface reflectance ρ:

$$I = \mathcal{R}(z|l, \rho) + \eta_I, \tag{2}$$

with η_I the realisation of a stochastic process standing for noise, quantisation and outliers.

Assuming frontal lighting, uniform Lambertian reflectance, Lipschitz-continuous depth and orthographic projection, solving (2) in terms of the depth map z comes down to solving the eikonal equation [7]

$$|\nabla z| = \sqrt{\frac{1}{I^2} - 1}. \tag{3}$$

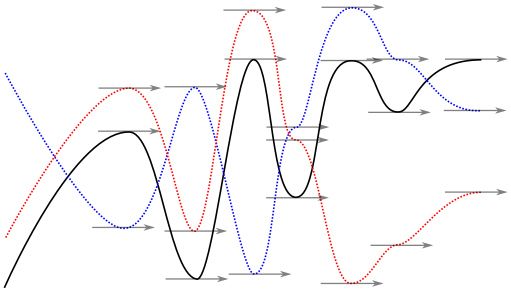


Figure 3: Shape-from-shading suffers from the concave / convex ambiguity: the genuine surface (solid line) and both the surfaces depicted by dashed lines produce the same image, if lit and viewed from above. We put forward low-resolution depth clues (*cf*. Figure 2) for disambiguation.

It is noteworthy that (3) only provides the magnitude of the depth gradient, and not its direction. The local shape is thus unambiguous in singular points (the tangent vectors in Figure 3), but two singular points may either be connected by "going up" or by "going down". This is the well-celebrated concave / convex ambiguity. One out of the infinitely many solutions of (3) can be numerically computed by variational methods [26, 29] or by resorting to the viscosity solution theory [10, 15, 35, 51]. See [6, 12, 62] for further details about numerical shape-from-shading.

Even under the unrealistic assumptions yielding the eikonal shape-from-shading model (3), shape inference is ill-posed. Hence, one may expect that more realistic lighting and reflectance assumptions will add more ambiguities. Several steps in the direction of handling natural lighting have been achieved [28, 31, 49], but they still require the reflectance to be uniform. However, in general both the lighting and the reflectance may be arbitrarily complex. This is nicely visualized in the "workshop metaphor" of Adelson and Pentland [1]: 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. To solve this series of ambiguities, additional constraints must be introduced. Barron *et al*. proposed for this purpose appropriate priors for reflectance (sparsity of the gradients), lighting (spherical harmonics model [4, 48]) and shape (smoothness), and combined them in order to achieve shape, reflectance and illumination from shading [3].

Recently, the shape-from-shading problem has 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 the multi-view setup [39, 63], but also in the single-shot case [9, 22, 42, 43, 57, 61] we tackle in this paper. Still, these methods require the resolutions of the input image and of the depth map to be the same.

2.3. Intuitive justification of our proposal

In view of this brief discussion on single depth image super-resolution and shape-from-shading, we conclude that, in the context of RGB-D sensing, the high-frequency information necessary to achieve detail-preserving depth super-resolution could be provided by the photometric data. Similarly, the low-frequency information necessary to disambiguate shape-from-shading could be conveyed by the geometric data. Compare Figures 2 and 3, and see Figure 4. It should thus be possible to achieve joint depth map refinement and super-resolution in a single shot, without resorting to additional data (new viewing angles or illumination conditions, learnt dictionary, etc.). In the next section, we formulate this task as a principled variational problem, by resorting to Bayesian inference.

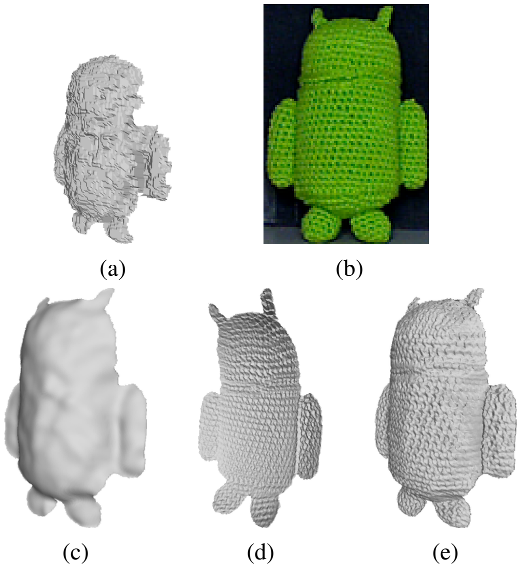


Figure 4: (a-b) Input low-resolution depth and high-resolution color images. (c) Blind super-resolution (achieved by disabling the shape-from-shading term in (18)) cannot hallucinate high-frequency geometric details from (a). (d) Shape-from-shading (achieved by setting $\mu = 0$ in (18)) applied to (b) appropriately recover such thin structures, but it is prone to low-frequency errors. (e) The combination of both techniques yields appropriate restoration of both high- and low-frequency components.

3. A variational approach to joint depth super-resolution and shape-from-shading

We formulate shading-based depth super-resolution as the joint solving of (1) (super-resolution) and (2) (shape-from-shading) in terms of the high-resolution depth map $z : \Omega_{\text{HR}} \to \mathbb{R}$, given a low-resolution depth map $z_0 : \Omega_{\text{LR}} \to \mathbb{R}$ and a high-resolution RGB image $I : \Omega_{\text{HR}} \to \mathbb{R}^3$.

We aim at recovering not only a high-resolution depth map which is consistent both with the low-resolution depth measurements and with the high-resolution color data, but also the hidden parameters of the image formation model (2) *i.e.*, the reflectance ρ and the lighting l. This can be achieved by maximizing the posterior distribution of the input data which, according to Bayes rule, is given by

$$\mathcal{P}(z,\rho,l|z_0,I) = \frac{\mathcal{P}(z_0,I|z,\rho,l)\,\mathcal{P}(z,\rho,l)}{\mathcal{P}(z_0,I)}, \tag{4}$$

where the numerator is the product of the likelihood with the prior, and the denominator is the evidence, which can be discarded since it plays no role in maximum a posteriori (MAP) estimation. In order to make the independency assumptions as transparent as possible and to motivate the final energy we aim at minimizing (see (18)), we follow in the next subsections David Mumford's approach [41] to derive a variational model from the posterior distribution (4).

3.1. Likelihood

Let us start with the first term in the numerator of (4) *i.e.*, the likelihood. By construction of RGB-D sensors, depth and color observations are independent, hence $\mathcal{P}(z_0,I|z,\rho,l) = \mathcal{P}(z_0|z,\rho,l)\,\mathcal{P}(I|z,\rho,l)$. We further assume that the depth observations are independent from the surface reflectance and from the lighting, hence $\mathcal{P}(z_0|z,\rho,l) = \mathcal{P}(z_0|z)$ and thus:

$$\mathcal{P}(z_0,I|z,\rho,l) = \mathcal{P}(z_0|z)\,\mathcal{P}(I|z,\rho,l). \tag{5}$$

Assuming homoskedastic, zero-mean Gaussian noise η_z with variance σ_z^2 in (1), the first factor in (5) writes

$$\mathcal{P}(z_0|z) \propto \exp\left\{-\frac{\|Kz - z_0\|^2_{\ell^2(\Omega_{\text{LR}})}}{2\sigma_z^2}\right\}. \tag{6}$$

Next, we discuss the second factor in (5), by making Equation (2) explicit. In general, the irradiance in channel $\star \in \{R,G,B\}$ writes

$$I_\star = \int_\lambda \int_\omega c_\star(\lambda)\rho(\lambda)\phi(\lambda,\omega)\max\{0,\mathbf{s}(\omega)\cdot\mathbf{n}_z\}\,\mathrm{d}\omega\,\mathrm{d}\lambda + \eta_I, \tag{7}$$

where integration is carried out over all wavelengths λ (ρ is the spectral reflectance of the surface and $c_\star$ is the transmission spectrum of the camera in channel $\star$) and all incident lighting directions ω ($\mathbf{s}(\omega)$ is the unit-length vector pointing towards the light source located in direction ω, and $\phi(\cdot,\omega)$ is the spectrum of this source), and $\mathbf{n}_z$ is the unit-length surface normal (which depends on the underlying depth map z). Assuming achromatic lighting i.e., $\phi(\cdot,\omega) := \phi(\omega)$, and using a first-order[1] spherical harmonics approximation of the inner integral, we obtain

$$I = \underbrace{\begin{bmatrix} \int_\lambda c_R(\lambda)\rho(\lambda)\mathrm{d}\lambda \\ \int_\lambda c_G(\lambda)\rho(\lambda)\mathrm{d}\lambda \\ \int_\lambda c_B(\lambda)\rho(\lambda)\mathrm{d}\lambda \end{bmatrix}}_{:=\rho} l \cdot \begin{bmatrix} \mathbf{n}_z \\ 1 \end{bmatrix} + \eta_I, \tag{8}$$

with $l \in \mathbb{R}^4$ the achromatic "light vector", $\rho : \Omega_{\text{HR}} \to \mathbb{R}^3$ the albedo (Lambertian reflectance) map, relatively to the camera transmission spectra $\{c_\star\}_{\star\in\{R,G,B\}}$, and $\mathbf{n}_z : \Omega_{\text{HR}} \to \mathbb{S}^2 \subset \mathbb{R}^3$ the field of unit-length surface normals. Assuming perspective projection with focal length $f > 0$ and $\mathbf{p} : \Omega_{\text{HR}} \to \mathbb{R}^2$ the field of pixel coordinates with respect to the principal point, the normal field is given by

$$\mathbf{n}_z = \frac{1}{\sqrt{|f\,\nabla z|^2 + (-z - \mathbf{p}\cdot\nabla z)^2}} \begin{bmatrix} f\,\nabla z \\ -z - \mathbf{p}\cdot\nabla z \end{bmatrix} \tag{9}$$

(see, for instance, [46]).

Assuming that the image noise is homoskedastically Gaussian-distributed with zero-mean and covariance matrix $\text{Diag}(\sigma_I^2,\sigma_I^2,\sigma_I^2)$, we obtain

$$\mathcal{P}(I|z,\rho,l) \propto \exp\left\{-\frac{\|(l\cdot\mathbf{m}_{z,\nabla z})\,\rho - I\|^2_{\ell^2(\Omega_{\text{HR}})}}{2\sigma_I^2}\right\}, \tag{10}$$

where, according to (8) and (9), $\mathbf{m}_{z,\nabla z}$ is a $\Omega_{\text{HR}} \to \mathbb{R}^4$ vector field defined as

$$\mathbf{m}_{z,\nabla z} = \begin{bmatrix} \dfrac{f\,\nabla z}{\sqrt{|f\,\nabla z|^2 + (-z - \mathbf{p}\cdot\nabla z)^2}} \\ \dfrac{-z - \mathbf{p}\cdot\nabla z}{\sqrt{|f\,\nabla z|^2 + (-z - \mathbf{p}\cdot\nabla z)^2}} \\ 1 \end{bmatrix}. \tag{11}$$

3.2. Priors

We now consider the second factor in the numerator of (4) *i.e.*, the prior distribution. We assume that depth, reflectance and lighting are independent (independence of reflectance from depth and lighting follows from the Lambertian assumption, and independence of lighting from depth follows from the distant-light assumption required to derive the spherical harmonics model (8), see [4, 48]). This implies

$$\mathcal{P}(z,\rho,l) = \mathcal{P}(z)\,\mathcal{P}(\rho)\,\mathcal{P}(l). \tag{12}$$

Since lighting has already been modeled as a low-frequency phenomenon for the sake of expliciting the image formation model (8), we do not need to introduce any other prior $\mathcal{P}(l)$ and thus we use an improper prior

$$\mathcal{P}(l) = \text{constant}. \tag{13}$$

[1] The whole proposed method is straightforward to extend to second-order spherical harmonics. However we did not observe substantial improvement with this extension, hence we discuss only the first-order case, which can capture more than 85% of natural illumination [18].

Regarding the depth map z, we follow the recent work [21] and opt for a minimal surface prior. Remark that

$$\mathrm{d}\mathcal{A}_{z,\nabla z} = \frac{z}{f^2}\sqrt{|f\,\nabla z|^2 + (-z - \mathbf{p}\cdot\nabla z)^2} \tag{14}$$

is a $\Omega_{\text{HR}} \rightarrow \mathbb{R}$ scalar field which maps each pixel to the area of the corresponding surface element. Thus $\|\mathrm{d}\mathcal{A}_{z,\nabla z}\|_{\ell^1(\Omega_{\text{HR}})}$ is the total surface area and the minimal surface prior writes

$$\mathcal{P}(z) \propto \exp\left\{-\frac{\|\mathrm{d}\mathcal{A}_{z,\nabla z}\|_{\ell^1(\Omega_{\text{HR}})}}{\alpha}\right\}, \tag{15}$$

with $\alpha > 0$ a free parameter controlling smoothness.

According to the Retinex theory [33], the reflectance ρ can be assumed piecewise constant. This yields a Potts prior

$$\mathcal{P}(\rho) \propto \exp\left\{-\frac{\|\nabla\rho\|_{\ell^0(\Omega_{\text{HR}})}}{\beta}\right\}, \tag{16}$$

with $\beta > 0$ a scale parameter, and $\|\cdot\|_{\ell^0}$ an abusive notation for the length of the discontinuity set:

$$\|\nabla\rho\|_{\ell^0(\Omega_{\text{HR}})} = \sum_{p\in\Omega_{\text{HR}}} \begin{cases} 0, & \text{if } |\nabla\rho(p)|_2 = 0, \\ 1, & \text{otherwise,} \end{cases} \tag{17}$$

where $|\cdot|_2$ is the Euclidean norm in $\mathbb{R}^6$.

3.3. Variational formulation

Replacing the maximisation of the posterior distribution (4) by the minimisation of its negative logarithm, combining Equations (4)–(6), (10), (12)–(16), and neglecting the additive constants, we end up with the variational model

$$\min_{\substack{\rho\colon \Omega_{\text{HR}}\rightarrow\mathbb{R}^3 \\ l\in\mathbb{R}^4 \\ z\colon\Omega_{\text{HR}}\rightarrow\mathbb{R}}} \|(l\cdot\mathbf{m}_{z,\nabla z})\,\rho - I\|^2_{\ell^2(\Omega_{\text{HR}})} + \mu\,\|Kz - z_0\|^2_{\ell^2(\Omega_{\text{LR}})} + \nu\,\|\mathrm{d}\mathcal{A}_{z,\nabla z}\|_{\ell^1(\Omega_{\text{HR}})} + \lambda\,\|\nabla\rho\|_{\ell^0(\Omega_{\text{HR}})}\,, \tag{18}$$

with the following definitions of the weights:

$$\mu = \frac{\sigma_I^2}{\sigma_z^2}, \quad \nu = \frac{2\sigma_I^2}{\alpha} \quad \text{and } \lambda = \frac{2\sigma_I^2}{\beta}. \tag{19}$$

3.4. Numerical solution

We now describe an algorithm for effectively solving the variational problem (18), which is both nonsmooth and nonconvex. In order to tackle the nonlinear dependency upon the depth and its gradient arising from shape-from-shading and minimal surface regularisation, we follow [47] and introduce an auxiliary variable $\theta := (z,\nabla z)$, then rewrite (18) as a constrained optimisation problem:

$$\min_{\substack{\rho\colon \Omega_{\text{HR}}\rightarrow\mathbb{R}^3 \\ l\in\mathbb{R}^4 \\ z\colon\Omega_{\text{HR}}\rightarrow\mathbb{R} \\ \theta\colon\Omega_{\text{HR}}\rightarrow\mathbb{R}^3}} \|(l\cdot\mathbf{m}_{\theta})\,\rho - I\|^2_{\ell^2(\Omega_{\text{HR}})} + \mu\,\|Kz - z_0\|^2_{\ell^2(\Omega_{\text{LR}})} + \nu\,\|\mathrm{d}\mathcal{A}_{\theta}\|_{\ell^1(\Omega_{\text{HR}})} + \lambda\,\|\nabla\rho\|_{\ell^0(\Omega_{\text{HR}})}$$

$$\text{s.t. } \theta = (z,\nabla z). \tag{20}$$

We then use a multi-block variant of ADMM [5, 13, 19] to solve (20)[2]. Given the current estimates $(\rho^{(k)}, l^{(k)}, \theta^{(k)}, z^{(k)})$ at iteration (k), the variables are updated according to the following sweep:

$$\rho^{(k+1)} = \operatorname*{argmin}_{\rho} \left\|\left(l^{(k)}\cdot\mathbf{m}_{\theta^{(k)}}\right)\rho - I\right\|^2_{\ell^2(\Omega_{\text{HR}})} + \lambda\,\|\nabla\rho\|_{\ell^0(\Omega_{\text{HR}})}\,, \tag{21}$$

$$l^{(k+1)} = \operatorname*{argmin}_{l} \left\|(l\cdot\mathbf{m}_{\theta^{(k)}})\,\rho^{(k+1)} - I\right\|^2_{\ell^2(\Omega_{\text{HR}})}, \tag{22}$$

$$\theta^{(k+1)} = \operatorname*{argmin}_{\theta} \left\|\left(l^{(k+1)}\cdot\mathbf{m}_{\theta}\right)\rho^{(k+1)} - I\right\|^2_{\ell^2(\Omega_{\text{HR}})} + \nu\|\mathrm{d}\mathcal{A}_{\theta}\|_{\ell^1(\Omega_{\text{HR}})} + \frac{\kappa}{2}\left\|\theta - (z,\nabla z)^{(k)} + u^{(k)}\right\|^2_{\ell^2(\Omega_{\text{HR}})}, \tag{23}$$

$$z^{(k+1)} = \operatorname*{argmin}_{z} \mu\,\|Kz - z_0\|^2_{\ell^2(\Omega_{\text{LR}})} + \frac{\kappa}{2}\left\|\theta^{(k+1)} - (z,\nabla z) + u^{(k)}\right\|^2_{\ell^2(\Omega_{\text{HR}})}, \tag{24}$$

$$u^{(k+1)} = u^{(k)} + \theta^{(k+1)} - (z^{(k+1)}, \nabla z^{(k+1)}), \tag{25}$$

where u and κ are a Lagrange multiplier and a step size, respectively. In our implementation κ is determined automatically using the varying penalty procedure [23].

To solve the albedo sub-problem (21) we resort to primal-dual iterations [54]. The lighting update (22) is solved using pseudo-inverse. The θ-update (23) comes down to a series of independent (there is no coupling between neighboring pixels, thanks to the ADMM strategy) nonlinear optimisation problems, which we solve using the implementation [52] of the L-BFGS method [36], using the Moreau envelope of the ℓ^1 norm to ensure differentiability. The depth update (24) requires solving a large sparse linear least-squares problem, which we tackle using conjugate gradient on the normal equations.

Although the overall optimisation problem (18) is nonconvex, recent works [24, 30, 56] have demonstrated that under mild assumptions on the cost function and small enough step size κ, nonconvex ADMM converges to a critical point. In practice, we found the proposed ADMM scheme to be stable and always observed convergence. In our experiments we use as initial guess: $\rho^{(0)} = I$, $l^{(0)} = [0,0,-1,0]^\top$, $z^{(0)}$ a smoothed (using bilinear filtering) version of a linear interpolation of the low-resolution input z_0, $\theta^{(0)} = (z^0, \nabla z^{(0)})$, $u^{(0)} \equiv 0$ and $\kappa^{(0)} = 10^{-4}$. In all our experiments, 10 to 20 global iterations (k) were sufficient to reach convergence, which is evaluated through the relative residual between two successive depth estimates $z^{(k+1)}$ and $z^{(k)}$. On a recent laptop computer with $i7$ processor, such a process requires around one minute (code is implemented in Matlab except the albedo update, which is implemented in CUDA).

[2] Code and dataset is available at https://github.com/BjoernHaefner/DepthSRfromShading.

4. Experimental validation

In this section we evaluate our variational approach to joint depth super-resolution and shape-from-shading against challenging synthetic and real-world datasets.

4.1. Synthetic data

We first discuss the choice of the parameters involved in the variational problem (18). Although their optimal values can be deduced from the data statistics (see (19)), it can be difficult to estimate such statistics in practice and thus we rather consider μ, ν and λ as tunable hyper-parameters. The formulae in (19) remain however insightful regarding the way these parameters should be tuned.

To select an appropriate set of parameters, we consider a synthetic dataset (the publicly available "Joyful Yell" 3D-shape) which we render under first-order spherical harmonics lighting ($l = [0, 0, -1, 0.2]^\top$) with three different reflectance maps as depicted in Figure 5. Additive zero-mean Gaussian noise with standard deviation 1% that of the original images is added to the high resolution (640×480 px^2) images. Ground-truth high resolution and input low-resolution (320×240 px^2) depth maps are rendered from the 3D-model. Non-uniform zero-mean Gaussian noise with standard deviation 10^{-3} times the squared original depth value (consistently with the real-world measurements from [32]) is then added to the low-resolution depth map. Quantitative evaluation is carried out by evaluating the root mean squared error (RMSE) between the estimated depth and albedo maps and the ground-truth ones.

Initially, we chose $\mu = \frac{1}{12}$, $\nu = 2$ and $\lambda = 1$. Then, we evaluated the impact of varying each parameter, keeping the others fixed to these values found empirically. Results are shown in Figure 6. Quite logically, μ should not be set too high otherwise the resulting depth map is as noisy as the input. Low values always allow a good albedo estimation, but the range $\mu \in [10^{-2}, 1]$ seems to provide the most accurate depth maps. Regarding λ, larger values should be chosen if the reflectance is uniform, but they induce high errors whenever it is not. On the other hand, low values systematically yield high errors since the reflectance estimate absorbs all the shading information (this is the "painter's explanation" in the "workshop metaphor" [1]). In between, the range $\lambda \in [10^{-1}, 10]$ seems to always give reasonable results. Eventually, high values of ν should be avoided in order to prevent over-smoothing.

Since we chose to disambiguate shape-from-shading by assuming piecewise-constant reflectance, the minimal surface prior plays no role in disambiguation. This explains why low values of ν should be preferred. Depth regularisation matters only when color cannot be exploited, for instance due to shadows, black reflectance or saturation. This will be better visualised in the real-world experiments.

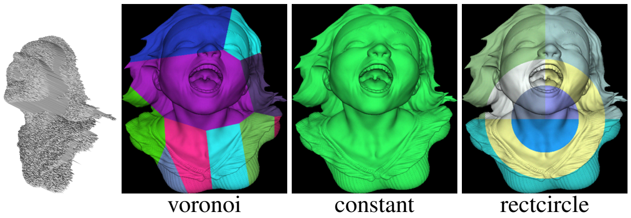


Figure 5: Synthetic dataset used for quantitative evaluation. Left: low-resolution depth map. Right: high-resolution RGB images, rendered using three different albedo maps.

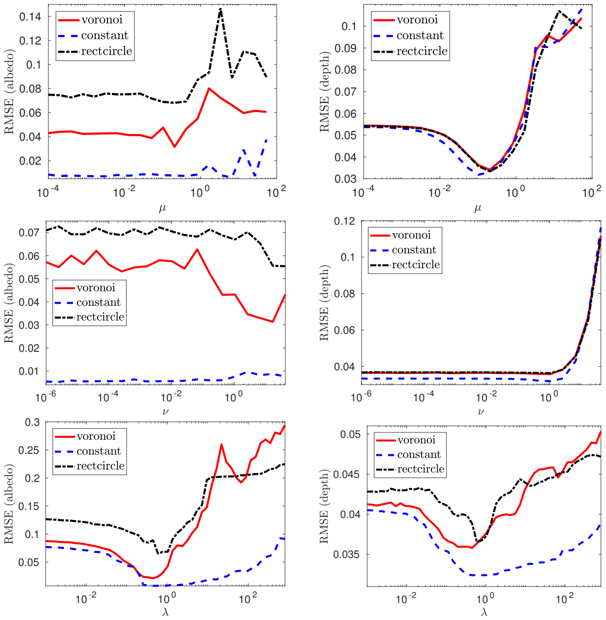


Figure 6: Impact of the parameters μ, ν and λ on the accuracy of the albedo and depth estimates. Based on those experiments, we select the set of parameters $(\mu, \nu, \lambda) = (10^{-1}, 10^{-1}, 2)$ for our experiments.

In Figure 7, we compare our method with two other single-shot ones: a learning-based approach [58] and an image-based one [60]. To emphasise the interest of joint shape-from-shading and super-resolution over shading-based depth refinement using the downsampled image, we also show the results of [43]. For fair comparison with [58], this time we use a scaling factor of 4 for all methods *i.e.*, the depth maps are rendered at 120×160 px^2. To evaluate the recovery of thin structures, we provide the mean angular error with respect to surface normals. The learning-based method can obviously not hallucinate surface details since it does not use the color image. The image-based method does a much better job, but it is largely overcome by shading-based super-resolution.

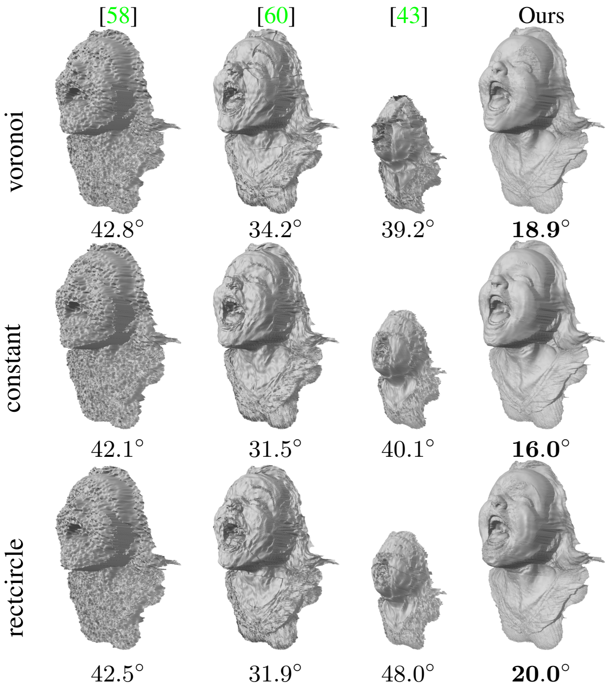


Figure 7: Comparison between learning-based [58], image-based [60] and shading-based (ours) depth super-resolution, as well as shading-based refinement using low-resolution images [43]. Our method systematically outperforms the others (numbers are the mean angular errors on normals).

4.2. Real-world data

For real-world experiments, we use the Asus Xtion Pro Live sensor, which delivers 1280 × 1024 px^2 RGB and 640 × 480 px^2 depth images at 30 fps. Data are acquired in an indoor office with ambient lighting, and objects are manually segmented from background before processing.

Figures 1, 4, 8, 9, 10 and 13 present real-world results. Combining depth super-resolution and shape-from-shading apparently resolves the low-frequency and high-frequency ambiguities arising in either of the inverse problems. Over-segmentation of reflectance may happen, but this does not seem to impact depth recovery. Whenever color gets saturated or too low, then minimal surface drives super-resolution, which adds robustness. Additional results using depth maps with lower resolution (320 × 240 px^2) are presented in Figure 11. Our method only fails when reflectance does not fit the Potts prior, as shown in Figure 12 for an object with smoothly-varying reflectance. It induces bias in the estimated depth such that reflectance based artifacts appear. Handling such cases would require using another prior for the reflectance, or actively controlling lighting. This has already been achieved in RGB-D sensing [2, 8, 45], but it is not compatible with single-shot applications.

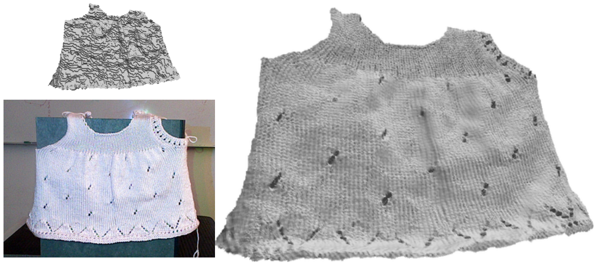


Figure 8: Super-resolution of a "dress". The estimated reflectance map is uniform, hence it is not displayed here.

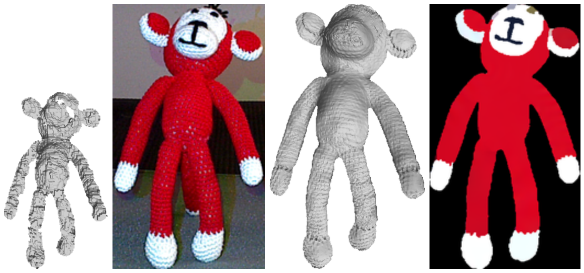


Figure 9: Super-resolution of a "monkey doll". Fine-scale shape and reflectance structures are nicely recovered.

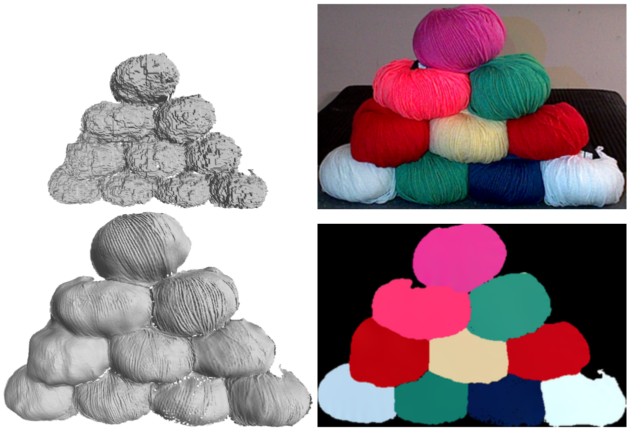


Figure 10: Super-resolution of "wool balls". Minimal surface drives super-resolution when color gets saturated.

5. Conclusion

A variational approach to single-shot depth super-resolution for RGB-D sensors is proposed. It fully exploits the color information in order to guide super-resolution, by resorting to the shape-from-shading technique. Low-resolution depth cues resolve the ambiguities arising in shape-from-shading and, symmetrically, high-resolution photometric clues resolve those of depth super-resolution.

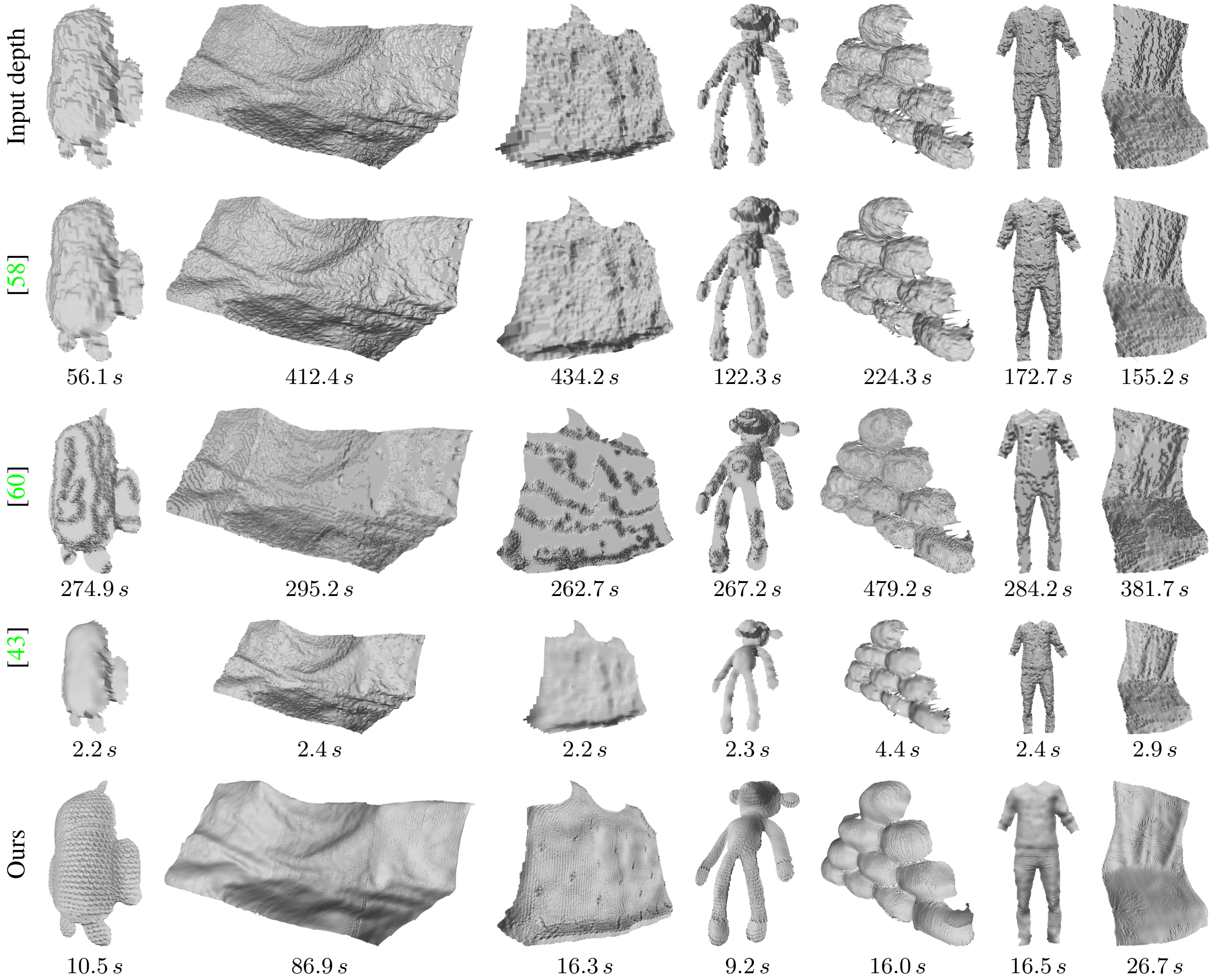


Figure 11: Comparison between our super-resolution method, two others [58, 60] and shading-based depth refinement on the low-resolution images [43]. Our shading-based super-resolution restores the complex geometry the best. Numbers represent runtime in seconds.

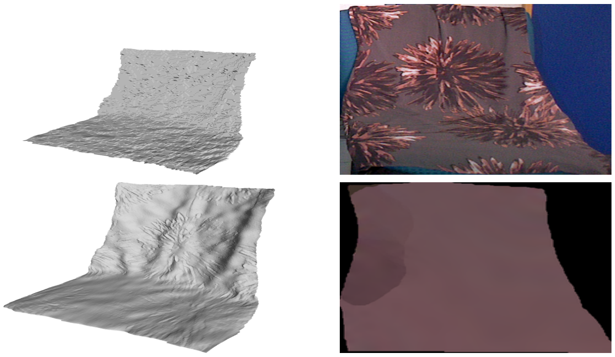


Figure 12: If the pictured object does not match our Potts prior for the reflectance, artifacts appear.

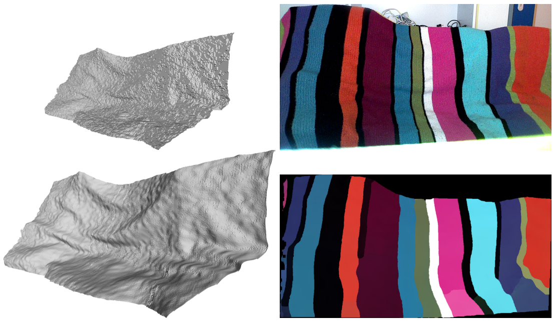


Figure 13: Super-resolution of a "blanket". Despite over-segmentation of the reflectance, thin structures are recovered. Even in black areas without shading information, results remain satisfactory thanks to the minimal surface prior.

References

[1] E. H. Adelson and A. P. Pentland. *Perception as Bayesian inference*, chapter The perception of shading and reflectance, pages 409–423. Cambridge University Press, 1996. 3, 6

[2] R. Anderson, B. Stenger, and R. Cipolla. Augmenting depth camera output using photometric stereo. In *Proceedings of the IAPR Conference on Machine Vision Applications*, 2011. 7

[3] J. Barron and J. Malik. Shape, illumination, and reflectance from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(8):1670–1687, 2015. 3

[4] R. Basri and D. P. Jacobs. Lambertian reflectances and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):218–233, 2003. 3, 4

[5] 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*, 3(1):1–122, 2011. 5

[6] 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*, 5(1):311–342, 2012. 3

[7] A. R. Bruss. The eikonal equation: Some results applicable to computer vision. *Journal of Mathematical Physics*, 23(5):890–896, 1982. 2

[8] 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*, pages 933–941, 2015. 7

[9] 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*, 122(1):1–16, 2017. 3

[10] E. Cristiani and M. Falcone. Fast semi-lagrangian schemes for the eikonal equation and applications. *SIAM Journal on Numerical Analysis*, 45(5):1979–2011, 2007. 3

[11] J. Diebel and S. Thrun. An application of Markov random fields to range sensing. In *Advances in Neural Information Processing Systems*, pages 291–298, 2006. 2

[12] J.-D. Durou, M. Falcone, and M. Sagona. Numerical Methods for Shape-from-shading: A New Survey with Benchmarks. *Computer Vision and Image Understanding*, 109(1):22–43, 2008. 3

[13] J. Eckstein and D. P. Bertsekas. On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1):293–318, 1992. 5

[14] M. Elad and A. Feuer. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Transactions on Image Processing*, 6(12):1646–1658, 1997. 2

[15] 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*, pages 596–603, 1997. 3

[16] D. Ferstl, C. Reinbacher, R. Ranftl, M. Rüther, and H. Bischof. Image guided depth upsampling using anisotropic total generalized variation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 993–1000, 2013. 2

[17] D. Ferstl, M. Ruther, and H. Bischof. Variational depth superresolution using example-based edge representations. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 513–521, 2015. 2

[18] D. Frolova, D. Simakov, and R. Basri. Accuracy of spherical harmonic approximations for images of Lambertian objects under far and near lighting. In *Proceedings of the European Conference on Computer Vision*, pages 574–587, 2004. 4

[19] 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*, 9(R2):41–76, 1975. 5

[20] B. Goldlücke, M. Aubry, K. Kolev, and D. Cremers. A super-resolution framework for high-accuracy multiview reconstruction. *International Journal of Computer Vision*, 106(2):172–191, 2014. 2

[21] 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*, pages 511–520, 2015. 5

[22] 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*, pages 1617–1624, 2013. 3

[23] B. S. He, H. Yang, and S. L. Wang. Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities. *Journal of Optimization Theory and Applications*, 106(2):337–356, 2000. 5

[24] M. Hong, Z.-Q. Luo, and M. Razaviyayn. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. *SIAM Journal on Optimization*, 26(1):337–364, 2016. 5

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

[26] B. K. P. Horn and M. J. Brooks. The variational approach to shape from shading. *Computer Vision, Graphics, and Image Processing*, 33(2):174–208, 1986. 3

[27] M. Hornácek, 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*, pages 1123–1130, 2013. 2

[28] R. Huang and W. A. P. Smith. Shape-from-shading under complex natural illumination. In *Proceedings of the IEEE International Conference on Image Processing*, pages 13–16, 2011. 3

[29] K. Ikeuchi and B. K. Horn. Numerical shape from shading and occluding boundaries. *Artificial intelligence*, 17(1-3):141–184, 1981. 3

[30] B. Jiang, T. Lin, S. Ma, and S. Zhang. Structured nonconvex and nonsmooth optimization: algorithms and iteration complexity analysis. *arXiv preprint arXiv:1605.02408*, 2016. 5

[31] M. K. Johnson and E. H. Adelson. Shape estimation in natural illumination. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2553–2560, 2011. 3

[32] K. Khoshelham and S. O. Elberink. Accuracy and resolution of Kinect depth data for indoor mapping applications. *Sensors*, 12(2):1437–1454, 2012. 6

[33] E. H. Land. The retinex theory of color vision. *Scientific American*, 237(6):108–120, 1977. 5

[34] 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*, pages 3374–3381, 2014. 2

[35] P.-L. Lions, E. Rouy, and A. Tourin. Shape-from-shading, viscosity solutions and edges. *Numerische Mathematik*, 64(1):323–353, 1993. 3

[36] D. C. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical programming*, 45(1):503–528, 1989. 5

[37] 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*, 102(1-3):18–32, 2013. 2

[38] 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*, pages 71–84, 2012. 2

[39] 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*, 2017. 3

[40] 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*, pages 536–544, 2015. 2

[41] D. Mumford. Bayesian rationale for the variational formulation. In *Geometry-driven diffusion in computer vision*, pages 135–146. 1994. 4

[42] 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*, pages 4378–4386, 2016. 3

[43] 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*, pages 5407–5416, 2015. 3, 6, 7, 8

[44] 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*, pages 1623–1630, 2011. 2

[45] 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 Workshops*, 2017. 2, 7

[46] Y. Quéau, J.-D. Durou, and J.-F. Aujol. Normal Integration: A Survey. *Journal of Mathematical Imaging and Vision*, 2017. 4

[47] 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 in Computer Vision and Pattern Recognition (EMMCVPR)*, 2017. 5

[48] R. Ramamoorthi and P. Hanrahan. An Efficient Representation for Irradiance Environment Maps. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*, pages 497–500, 2001. 3, 4

[49] 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*, pages 1128–1136, 2015. 3

[50] G. Riegler, M. Rüther, and H. Bischof. ATGV-net: accurate depth super-resolution. In *Proceedings of the European Conference on Computer Vision*, pages 268–284, 2016. 2

[51] E. Rouy and A. Tourin. A viscosity solutions approach to shape-from-shading. *SIAM Journal on Numerical Analysis*, 29(3):867–884, 1992. 3

[52] M. Schmidt. minFunc: unconstrained differentiable multivariate optimization in Matlab. `http://www.cs.ubc.ca/~schmidtm/Software/minFunc.html`, 2005. 5

[53] 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*, pages 343–350, 2009. 2

[54] E. Strekalovskiy and D. Cremers. Real-time minimization of the piecewise smooth Mumford-Shah functional. In *Proceedings of the European Conference on Computer Vision*, pages 127–141, 2014. 5

[55] M. Unger, T. Pock, M. Werlberger, and H. Bischof. A convex approach for variational super-resolution. In *DAGM Symposium*, pages 313–322, 2010. 2

[56] Y. Wang, W. Yin, and J. Zeng. Global convergence of ADMM in nonconvex nonsmooth optimization. *arXiv preprint arXiv:1511.06324*, 2015. 5

[57] 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*, 33(6):200:1–200:10, 2014. 3

[58] J. Xie, R. S. Feris, and M.-T. Sun. Edge-guided single depth image super resolution. *IEEE Transactions on Image Processing*, 25(1):428–438, 2016. 2, 6, 7, 8

[59] 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*, 17(9):1525–1537, 2015. 2

[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*, 2007. 2, 6, 7, 8

[61] 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*, pages 1415–1422, 2013. 3

[62] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah. Shape-from-shading: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8):690–706, 1999. 3

[63] 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*, 34(4):96:1–96:14, 2015. 3