This ICCV workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the version available on IEEE Xplore.

Depth Super-Resolution Meets Uncalibrated Photometric Stereo

Songyou Peng[1,2] Bjoern Haefner[1] Yvain Quéau[1] Daniel Cremers[1]
[1] Computer Vision Group, Technical University of Munich
[2] Erasmus Mundus Masters in Vision and Robotics (VIBOT)
{songyou.peng, bjoern.haefner, yvain.queau, cremers}@in.tum.de

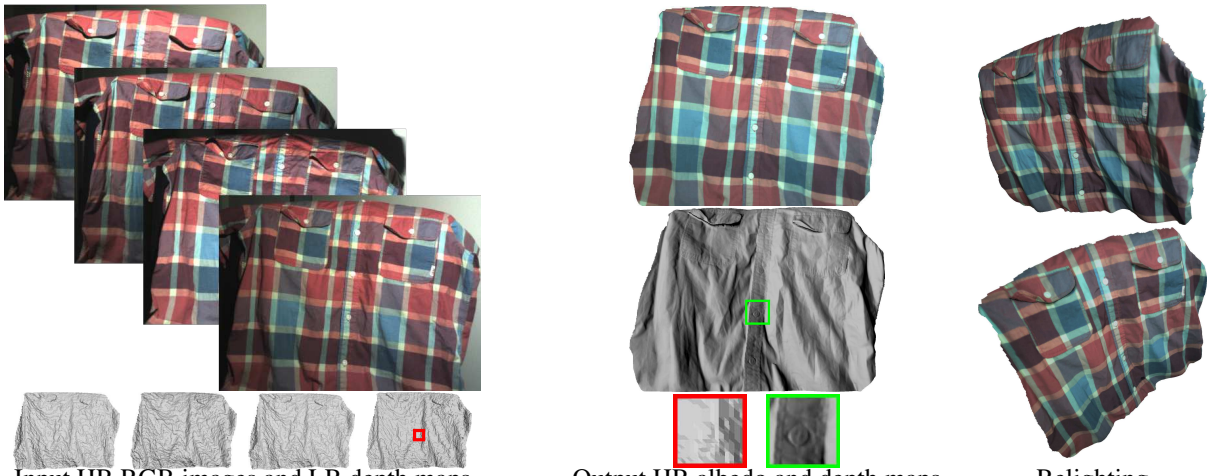


Figure 1: Given an RGB-D sequence of $n \geq 4$ low-resolution (320×240 px) depth maps and high-resolution (1280×1024 px) RGB images acquired from the same viewing angle but under varying, unknown lighting, high-resolution depth and reflectance maps are estimated by combining super-resolution and photometric stereo within a variational framework.

Abstract

A novel depth super-resolution approach for RGB-D sensors is presented. It disambiguates depth super-resolution through high-resolution photometric clues and, symmetrically, it disambiguates uncalibrated photometric stereo through low-resolution depth cues. To this end, an RGB-D sequence is acquired from the same viewing angle, while illuminating the scene from various uncalibrated directions. This sequence is handled by a variational framework which fits high-resolution shape and reflectance, as well as lighting, to both the low-resolution depth measurements and the high-resolution RGB ones. The key novelty consists in a new PDE-based photometric stereo regularizer which implicitly ensures surface regularity. This allows to carry out depth super-resolution in a purely data-driven manner, without the need for any ad-hoc prior or material calibration. Real-world experiments are carried out using an out-of-the-box RGB-D sensor and a hand-held LED light source.

1. Introduction

RGB-D sensors such as Microsoft Kinect or Asus Xtion Pro Live have become a popular way to acquire colored 3D-representations of the world at low-cost. Yet, the accuracy of such representations remains limited by two factors.

First, the depth channel is prone to quantization and noise, and it has a *coarser resolution* than the RGB one. For instance, the Asus Xtion Pro Live sensor provides QVGA (320×240 px) or VGA (640×480 px) depth resolution, while it offers SXGA (1280×1024 px) RGB resolution. Therefore, the RGB image is often downsampled to match the size of the depth channel. Some information is thus lost and the color may appear blurred or even aliased. Alternatively, the low-resolution (LR) depth map can be upsampled to the size of the high-resolution (HR) RGB image. This problem, known as super-resolution, is however ill-posed.

Second, the RGB image appears shaded due to ambient illumination. This may cause the 3D-reconstruction to look unrealistic in relighting or augmented reality applications. One would rather use *reflectance* in such applications, and not direct RGB (luminance) measurements.

This work simultaneously addresses both issues, by appropriately combining depth super-resolution and uncalibrated photometric stereo. It is shown that, by considering an RGB-D sequence acquired from the same viewing angle but under varying, unknown lighting, the LR depth measurements can be super-resolved without resorting to any ad-hoc prior or calibration. Reflectance and lighting are obtained as by-products. This is illustrated in Figure 1, and formalized as follows.

Problem Statement – Given a set of $n \geq 4$ HR RGB images $\mathbf{I}^i : \Omega_{\text{HR}} \subset \mathbb{R}^2 \to \mathbb{R}^3$, $i \in \{1, \dots, n\}$, and aligned LR depth maps $z_0^i : \Omega_{\text{LR}} \subset \Omega_{\text{HR}} \to \mathbb{R}$, $i \in \{1, \dots, n\}$, acquired from the same viewing angle but under varying, unknown lighting, estimate an HR depth map $z : \Omega_{\text{HR}} \to \mathbb{R}$, an HR reflectance map $\boldsymbol{\rho} : \Omega_{\text{HR}} \to \mathbb{R}^3$, and colored first-order spherical harmonics lighting $\{\mathbf{l}^i \in \mathbb{R}^{12}\}_i$.

Contribution and Organization of the Paper – After discussing related work in Section 2, we propose in Section 3 the new variational model (13) for joint depth super-resolution and uncalibrated photometric stereo. It combines a super-resolution fidelity term with a tailored PDE-based regularization term relying on photometric stereo. While the former ensures consistency between the sought HR depth map and the LR ones, the latter ensures that the sought HR depth map is both regular and consistent with the HR RGB images. Herein, low-resolution depth clues (resp., high-resolution photometric clues) act as natural disambiguation tools for uncalibrated photometric stereo (resp., depth super-resolution). This variational approach is evaluated in Section 4 against challenging synthetic and real-world datasets. Eventually, Section 5 summarizes our achievements and suggests future research directions.

2. Related Work

Depth Super-resolution – The most common way to achieve super-resolution consists of acquiring n LR measurements, and combine them into a single HR one. Starting from the seminal work of Tsai and Huang using Fourier analysis [25], various mathematical tools have been proposed for this task [27]. In the present work, we follow the variational approach.

The LR measurements $\{z_0^i\}_{i\in\{1,\dots,n\}}$ are assumed to result from downsampling and convolving an HR signal z, up to an additive, zero-mean homoskedastic Gaussian noise with standard deviation σ_z:

$$z_0^i = Kz + \varepsilon_z^i, \ \forall i \in \{1, \dots, m\}, \tag{1}$$

where K is the downsampling / convolution kernel, and $\varepsilon_z^i(\mathbf{p}) \sim \mathcal{N}(0, {\sigma_z}^2)$, $\mathbf{p} \in \Omega_{\text{HR}}$. In the present work, Kz can be described for each low-resolution pixel as a weighted sum over the corresponding super-resolution pixels, see [26] for a detailed explanation.

Estimating the HR signal z comes down to solving the inverse problem (1), which is ill-posed. A standard way to ensure well-posedness consists in introducing a prior on the HR signal and resorting to Bayesian inference. Such a strategy yields a variational problem of the form:

$$\min_{z:\,\Omega_{\text{HR}}\to\mathbb{R}} \mathcal{R}(z) + \frac{1}{2n}\sum_{i=1}^{n} \|Kz - z_0^i\|^2_{\ell^2(\Omega_{\text{LR}})}, \tag{2}$$

where $\mathcal{R}$ is a regularization term and $\|\cdot\|^2_{\ell^2(\Omega_{\text{LR}})}$ is the ℓ^2-norm over the LR domain Ω_{LR}. A typical choice for the regularizer is the total variation (TV) $\mathcal{R}(z) = \lambda\|\nabla z\|_{\ell^1(\Omega_{\text{HR}})}$ [19], with $\lambda > 0$ a tuning parameter and ∇ the gradient operator. This is essentially equivalent to assuming that the solution is piecewise-constant.

Super-resolution techniques have found numerous applications ranging from surveillance [7] to medical imaging [10], remote sensing [8] or, closer to our proposal, 3D-reconstruction using multi-view stereo [9] and RGB-D sensors [18]. In such applications where HR RGB measurements $\{\mathbf{I}^i\}_{i\in\{1,\dots,n\}}$ are available, they may be used as "guides" for depth super-resolution. For instance, the following anisotropic RGB image driven Huber-loss regularization term is advocated in [28]:

$$\mathcal{R}(z) = \int_{\Omega_{\text{HR}}} H_\varepsilon(z)\mathrm{d}\mathbf{p}, \ H_\varepsilon(z) := \begin{cases} \frac{\|D\nabla z\|^2}{2\varepsilon} & \text{if } \|D\nabla z\| \leq \varepsilon \\ \|D\nabla z\| - \frac{\varepsilon}{2} & \text{else,} \end{cases}, \tag{3}$$

where $D = \exp\left(\alpha\|\nabla\bar{\mathbf{I}}\|^\beta\right)\mathbf{v}\mathbf{v}^t + \mathbf{v}^\perp\left(\mathbf{v}^\perp\right)^t$ with $\mathbf{v} = \frac{\nabla\bar{\mathbf{I}}}{\|\nabla\bar{\mathbf{I}}\|}$, $\mathbf{v}^\perp$ a normal vector to $\mathbf{v}$, $\bar{\mathbf{I}} = \text{mean}\left(\{\mathbf{I}^i\}_{i\in\{1,\dots,n\}}\right)$, (α, β) some parameters and $\|\cdot\|$ is the standard (Euclidean) norm.x This regularizer tends to smooth z along, but not across edges and corners of the corresponding RGB image. Other image-based regularizers are also discussed in [22]. Employing the RGB measurements, which have a built-in higher resolution than the depth ones, indeed seems natural. However, this is not straightforward because image variations not only reflect shape variations, but also the interactions between light and matter. This is where photometric techniques come into play.

Uncalibrated Photometric Stereo – Inferring shape solely from image clues is an ill-posed problem, known as shape-from-shading [13]. It is impossible to unambiguously estimate shape from a single image, even when the reflectance is known. A natural way to disambiguate shape-from-shading is to consider not just one, but multiple images, obtained under varying lighting. This method is known as photometric stereo [29]. Assuming Lambertian reflectance with only additive, zero-mean homoskedastic Gaussian noise with standard deviation σ_I (no specularity or cast-shadow), and approximating lighting by first-order spherical harmonics [3], the following image forma-

tion model can be considered:

$$I^i_\star(\mathbf{p}) = \rho_\star(\mathbf{p})\, \mathbf{l}^i_\star \cdot \begin{bmatrix} \mathbf{n}(\mathbf{p}) \\ 1 \end{bmatrix} + \varepsilon^i_\star(\mathbf{p}), \tag{4}$$

with $(i, \star, \mathbf{p}) \in \{1, \dots, n\} \times \{R, G, B\} \times \Omega_{\text{HR}}$ the indices of the images, channel and pixel, $I^i_\star(\mathbf{p}) \in \mathbb{R}$ the i-th image value in channel $\star$ at pixel $\mathbf{p}$, $\rho_\star : \Omega_{\text{HR}} \to \mathbb{R}$ the albedo (Lambertian reflectance) map in channel $\star$, $\mathbf{l}^i_\star \in \mathbb{R}^4$ the i-th lighting vector in channel $\star$, $\mathbf{n}(\mathbf{p}) \in \mathbb{S}^2$ the unit-length outward normal at the surface point conjugate to pixel $\mathbf{p}$, and $\varepsilon^i_\star(\mathbf{p}) \sim \mathcal{N}(0, {\sigma_I}^2)$.

Uncalibrated photometric stereo aims at inferring reflectance, shape and lighting from the images, by solving the system of equations (4). Unfortunately, this problem is ill-posed: it can be solved only up to a linear ambiguity [12]. It is common to further enforce surface regularity [32], which reduces the ambiguity to a generalized bas-relief (GBR) one under directional lighting [4], and to a Lorenz one under spherical harmonics lighting [3]. Resolution of such ambiguities by resorting to additional priors [1, 21], and extensions to non-Lambertian reflectances [16], remain active research topics. It has also been shown recently in [23] that PDE-based approaches may be worthwhile for uncalibrated photometric stereo, because they implicitly enforce integrability and thus naturally reduce ambiguities.

Photometric RGB-D Sensing – Depth sensing improvement by shading analysis has been tackled in many recent works [11, 15, 20, 30, 31]. However, such methods do not actively control lighting, and thus they suffer from the same ambiguity as shape-from-shading. In particular, a smoothness prior on reflectance is always required. We will see in Section 4 that this considerably limits applicability. To unambiguously estimate reflectance using an RGB-D sensor, there is no other choice but to actively control lighting *i.e.*, to resort to photometric stereo [2, 5].

Photometric Super-Resolution – Super-resolution and photometric stereo have been widely studied, but rarely together. Some authors super-resolve the photometric stereo results [24], and others generate HR images using photometric stereo [6], but few employ LR depth clues. The only work in that direction is that in [17], where calibrated photometric stereo and structured light sensing are combined. However, this involves a non-standard setup and careful lighting calibration, and reflectance is assumed to be uniform. In contrast, we provide in the next section working tools for out-of-the-box RGB-D sensors and surfaces with spatially-varying reflectance. Therein, it is only assumed that the LR depth maps are aligned with the HR RGB images and that the RGB sensor's intrinsics are known (both, the warping function and the intrinsics can be accessed *e.g.*, using OpenNI 2 for ROS).

3. A Variational Framework for Photometric Stereo-Aware Depth Super-resolution

The main contribution of this work is now presented. It consists of the variational approach (13) to joint depth super-resolution and uncalibrated photometric stereo. This variational framework involves a regularizer built upon the PDE-based photometric stereo model described hereafter.

3.1. PDE-based Photometric Stereo with First-Order Spherical Harmonics Lighting

The super-resolution fidelity term in (2) is expressed in terms of depth, instead of normals. Therefore, we resort to a differential photometric stereo approach to design the regularization term. Let us first show how to express (4) as a system of nonlinear PDEs in z, $\boldsymbol{\rho}$ and $\{\mathbf{l}^i\}_{i \in \{1,\dots,n\}}$ over Ω_{HR} which have the following form:

$$\mathbf{A}^i(z, \boldsymbol{\rho}, \mathbf{l}^i)^\top \begin{bmatrix} \nabla z \\ z \end{bmatrix} = \mathbf{b}^i(\boldsymbol{\rho}, \mathbf{l}^i) + \boldsymbol{\varepsilon}^i,\ i \in \{1, \dots, n\}, \tag{5}$$

where $\mathbf{A}^i(z, \boldsymbol{\rho}, \mathbf{l}^i) : \Omega_{\text{HR}} \to \mathbb{R}^{3\times 3}$ and $\mathbf{b}^i(\boldsymbol{\rho}, \mathbf{l}^i) : \Omega_{\text{HR}} \to \mathbb{R}^3$ are fields which depend on the unknowns, and each $\boldsymbol{\varepsilon}^i$, $i \in \{1, \dots, n\}$ is a random $\Omega_{\text{HR}} \to \mathbb{R}^3$ homoskedastic Gaussian vector field with zero-mean and diagonal covariance matrix $\text{Diag}({\sigma_I}^2, {\sigma_I}^2, {\sigma_I}^2)$.

Under perspective projection, the normal in (4) reads:

$$\mathbf{n}(\mathbf{p}) = \frac{1}{d(z)(\mathbf{p})} \begin{bmatrix} f \nabla z(\mathbf{p}) \\ -z(\mathbf{p}) - \nabla z(\mathbf{p}) \cdot (\mathbf{p} - \mathbf{p}^0) \end{bmatrix}, \tag{6}$$

with $f > 0$ the focal length, $\mathbf{p}^0 \in \Omega_{\text{HR}}$ the principal point, and where $d(z)(\mathbf{p})$ is equal to the norm of the bracket (unit-length constraint). Plugging (6) into (4), the nonlinear system of PDEs (5) is obtained, with, $\forall \mathbf{p} \in \Omega_{\text{HR}}$:

$$\mathbf{A}^i(z, \boldsymbol{\rho}, \mathbf{l}^i)(\mathbf{p}) = \frac{1}{d(z)(\mathbf{p})} \left(f \begin{bmatrix} l^i_{R,1} & l^i_{G,1} & l^i_{B,1} \\ l^i_{R,2} & l^i_{G,2} & l^i_{B,2} \\ 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} \mathbf{p} - \mathbf{p}^0 \\ 1 \end{bmatrix} \left[l^i_{R,3}, l^i_{G,3}, l^i_{B,3} \right] \right) \text{Diag}(\boldsymbol{\rho}(\mathbf{p})), \tag{7}$$

$$\mathbf{b}^i(\boldsymbol{\rho}, \mathbf{l}^i)(\mathbf{p}) = \mathbf{I}^i(\mathbf{p}) - \begin{bmatrix} l^i_{R,4} & & \\ & l^i_{G,4} & \\ & & l^i_{B,4} \end{bmatrix} \boldsymbol{\rho}(\mathbf{p}), \tag{8}$$

where $\mathbf{I}^i(\mathbf{p}) = \left[I^i_R(\mathbf{p}), I^i_G(\mathbf{p}), I^i_B(\mathbf{p}) \right]^\top \in \mathbb{R}^3$, $\boldsymbol{\rho}(\mathbf{p}) = \left[\rho_R(\mathbf{p}), \rho_G(\mathbf{p}), \rho_B(\mathbf{p}) \right]^\top \in \mathbb{R}^3$, and $\mathbf{l}^i = \left[\left[\mathbf{l}^i_R \right]^\top, \left[\mathbf{l}^i_G \right]^\top, \left[\mathbf{l}^i_B \right]^\top \right]^\top \in \mathbb{R}^{12}$.

Let us remark that Model (5) is slightly more complex than previous PDE-based photometric stereo models such as the one in [23], because we consider first-order spherical harmonics lighting. In practice, this allows us to cope with much less restricted environments, for instance in the presence of strong ambient lighting.

3.2. Proposed Variational Framework

For the numerical solution, we follow a purely data-driven (maximum likelihood) variational approach. By independence of image and depth measurements, and of reflectance and lighting, the likelihood factorizes as follows:

$$\mathcal{P}(\{z_0^i, \mathbf{I}\}_i | z, \boldsymbol{\rho}, \{\mathbf{l}^i\}_i) = \mathcal{P}(\{\mathbf{I}\}_i | z, \boldsymbol{\rho}, \{\mathbf{l}^i\}_i)\, \mathcal{P}(\{z_0^i\}_i | z). \tag{9}$$

In addition, Equations (1) and (5) induce:

$$\mathcal{P}(\{\mathbf{I}\}_i | z, \boldsymbol{\rho}, \{\mathbf{l}^i\}_i) = \left(2\pi {\sigma_I}^2\right)^{-\frac{3n|\Omega_{\text{HR}}|}{2}} \exp\left\{ \left(-2{\sigma_I}^2\right)^{-1} \sum_{i=1}^{n} \left\| \mathbf{A}^i(z, \boldsymbol{\rho}, \mathbf{l}^i)^\top \begin{bmatrix} \nabla z \\ z \end{bmatrix} - \mathbf{b}^i(\boldsymbol{\rho}, \mathbf{l}^i) \right\|_{\ell^2(\Omega_{\text{HR}})} \right\}, \tag{10}$$

$$\mathcal{P}(\{z_0^i\}_i | z) = \left(2\pi {\sigma_z}^2\right)^{-\frac{n|\Omega_{\text{LR}}|}{2}} \exp\left\{ \left(2{\sigma_z}^2\right)^{-1} \sum_{i=1}^{n} \left\| Kz - z_0^i \right\|_{\ell^2(\Omega_{\text{LR}})} \right\}, \tag{11}$$

where $|\cdot|$ denotes cardinality. By further denoting:

$$\lambda = \frac{{\sigma_z}^2}{{\sigma_I}^2}, \tag{12}$$

and since maximizing likelihood (9) is equivalent to minimizing its negative logarithm, we obtain from Equations (9) to (12) the following variational model for joint depth super-resolution, reflectance and lighting estimation:

$$\min_{\substack{z:\,\Omega_{\text{HR}} \to \mathbb{R} \\ \boldsymbol{\rho}:\,\Omega_{\text{HR}} \to \mathbb{R}^3 \\ \{\mathbf{l}^i \in \mathbb{R}^{12}\}_i}} \left\{ \lambda \sum_{i=1}^{n} \left\| \mathbf{A}^i(z, \boldsymbol{\rho}, \mathbf{l}^i)^\top \begin{bmatrix} \nabla z \\ z \end{bmatrix} - \mathbf{b}^i(\boldsymbol{\rho}, \mathbf{l}^i) \right\|^2_{\ell^2(\Omega_{\text{HR}})} + \sum_{i=1}^{n} \|Kz - z_0^i\|^2_{\ell^2(\Omega_{\text{LR}})} \right\}. \tag{13}$$

Problem (13) yields the ill-posed uncalibrated photometric stereo one if $\lambda = +\infty$, and the ill-posed super-resolution one if $\lambda = 0$. We conjecture that any choice in between disambiguates both problems, but proving these conjectures is beyond the scope of this proof of concept work.

3.3. Alternating Optimization Strategy

The variational problem (13) is solved iteratively in terms of lighting, reflectance and depth, as illustrated in Figure 2. Reflectance and lighting updates come down to simple linear least-squares problems. During each depth update, we "freeze" the matrix fields $\mathbf{A}^i$ and $\mathbf{l}^i$ to their current values to obtain a linear (weighted) least-squares problem which is solved using conjugate gradient iterations. Initially, the reflectance is assumed uniformly white ($\boldsymbol{\rho} \equiv 1$) and the depth is obtained by meaning the LR measurements, filling missing values by biharmonic inpainting and eventually upsampling using bicubic interpolation. No initial lighting estimate is required, and the algorithm stops when the relative difference between two successive energy values falls below a threshold set to 0.01.

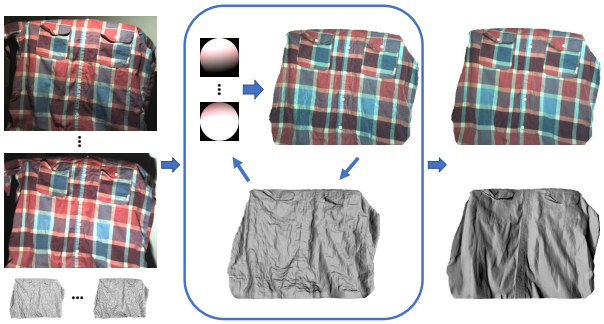


Figure 2: Sketch of our optimization framework for HR reflectance and depth estimation, given a series of HR RGB images and LR depth maps. Lighting, HR reflectance and HR depth are sequentially optimized until convergence.

4. Empirical Validation

4.1. Quantitative Evaluation on Synthetic Datasets

The public domain "Joyful Yell" 3D-shape is first considered. Depth maps with different scale factors are rendered, and noise (Gaussian, zero-mean, with standard deviation $\sigma_z = \alpha_z \|z\|_\infty$, $\alpha_z > 0$) is added to each LR depth map. The accuracy of the 3D-reconstruction is evaluated by comparing the HR 3D-reconstruction against the ground-truth. To create the RGB images, we proceed as follows. Using the ground truth depth map, normals are computed by finite differences. Then, random first-order spherical harmonics lighting vectors are generated, and an HR RGB image is taken as ground-truth albedo. All of these are eventually combined into an image generated according to Equation (4). Figure 3 summarizes this process.

Number of Images – We first evaluate in Figure 4 the impact of the number n of depth maps and photometric stereo images on the accuracy of the 3D-reconstruction. Quite obviously, the higher n, the more accurate the 3D-reconstruction. However, the runtime (evaluated on a Xeon processor at 3.50 GHz with 32 GB of RAM) of each iteration (convergence is reached within at most 15 iterations in all the experiments) increases linearly with n. Overall, the choice $n \in [10, 30]$ represents a good compromise between accuracy and speed. Besides, somewhat similar results are obtained with a scaling factor of 2 and 4, and only from a scaling factor of 8 the results start to significantly deteriorate. We believe this is not a problem because real-world RGB-D sensors such as the Asus Xtion Pro Live only provide depth maps with resolution $\frac{1}{2}$ or $\frac{1}{4}$ that of the HR RGB image.

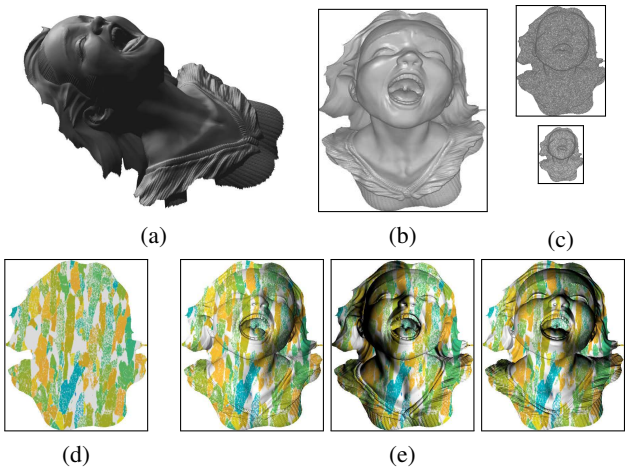


Figure 3: Synthetic dataset used in our quantitative experiments. (a) 3D-shape. (b) Ground truth HR (640×480 px) depth map. (c) LR noisy depth maps, for scaling factors of 2 (320×240 px) and 4 (160×120 px). (d) HR albedo map (source: https://mtex-toolbox.github.io/files/doc/EBSDSpatialPlots.html). (e) HR photometric stereo images.

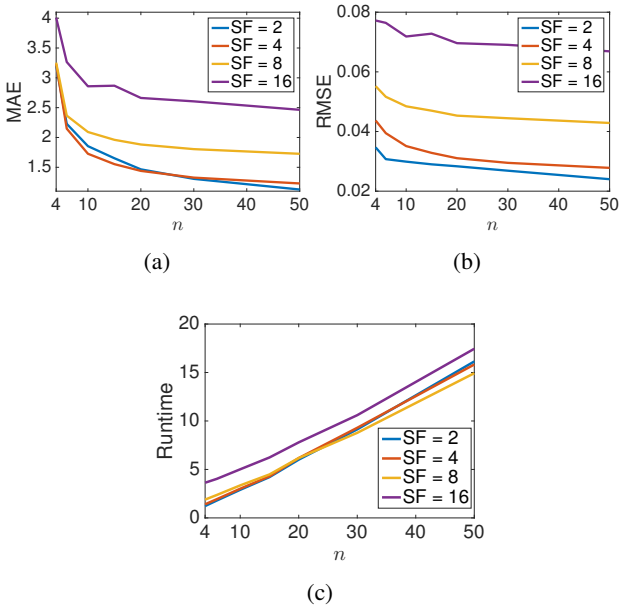


Figure 4: Impact of the number n of images on accuracy and computation time, for different scaling factors (SF). (a) Root Mean Square Error (RMSE, in arbitrary units) on depth. (b) Mean Angular Error (MAE, in degrees) on normals. 10 to 30 images are enough to obtain accurate results. (c) Runtime (in seconds) per each iteration as a function of the number n of images.

Parameter Tuning – The only parameter in Model (11) is λ, which controls the respective influence of the super-resolution and photometric terms. As expected, $\lambda \to 0$ yields a loss of fine-scale details (high mean angular error on normals due to the absence of photometric stereo-based estimation) while $\lambda \to \infty$ leads to a low-frequency bias (high root mean square error on depth due to the generalized bas-relief ambiguity). Figure 5 shows that the range $\lambda \in [10^{-2}, 10^{2}]$ provides satisfactory results. If not stated otherwise, the value $\lambda = 0.1$ is thus used in all our synthetic experiments.

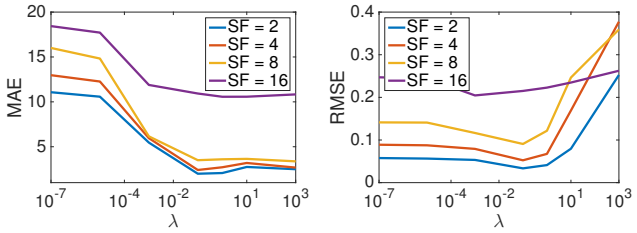


Figure 5: (a-b) Impact of the regularization parameter λ on accuracy. The interval $\lambda \in [10^{-2}, 10^{2}]$ constitutes an appropriate choice which both avoids the ambiguities of uncalibrated photometric stereo and the ill-posedness of super-resolution.

Robustness to Noise – Figure 6 evaluates robustness to noise in both the input LR depth maps and the HR RGB images. Unsurprisingly, accuracy severely deteriorates if noise is tremendous in both depth and RGB images. However, our approach is robust to a realistic amount of noise.

Comparison with Other Methods – Eventually, Figure 7 shows the advantage of our approach over standard image-driven depth super-resolution, pure uncalibrated photometric stereo, and shading-based refinement using a single LR RGB image. In this experiment, the depth super-resolution approach was implemented using (2) with (3) being the corresponding regularization term. Image-driven super-resolution interprets sharp image discontinuities as sharp depth features, because it is not able to estimate reflectance. For uncalibrated photometric stereo, we employed code from the authors of [21]. This method is able to estimate the albedo, but it still requires a prior in order to solve the generalized bas-relief ambiguity. It is thus not purely data-driven and subject to bias. As for shading-based refinement, the RGB-D fusion code provided in [20] was used. It assumes that both the RGB image and the depth map have the same resolution, thus it does not achieve super-resolution. Still, this experiment highlights the advantage of using a multiple-light setup: shape-from-shading methods have to

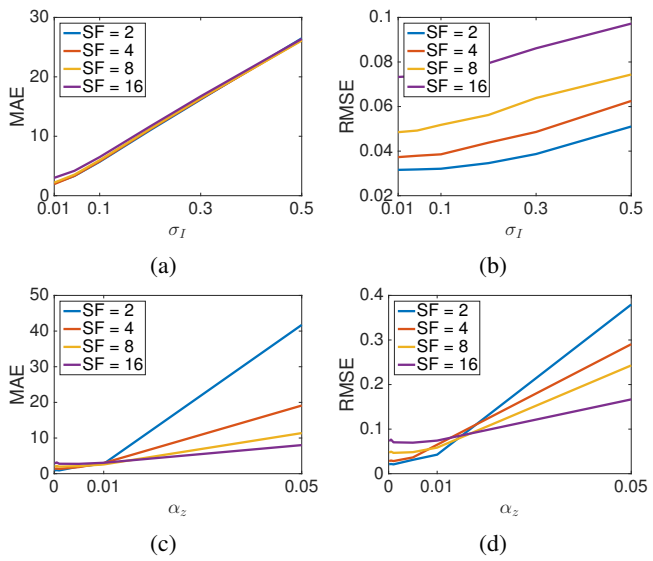


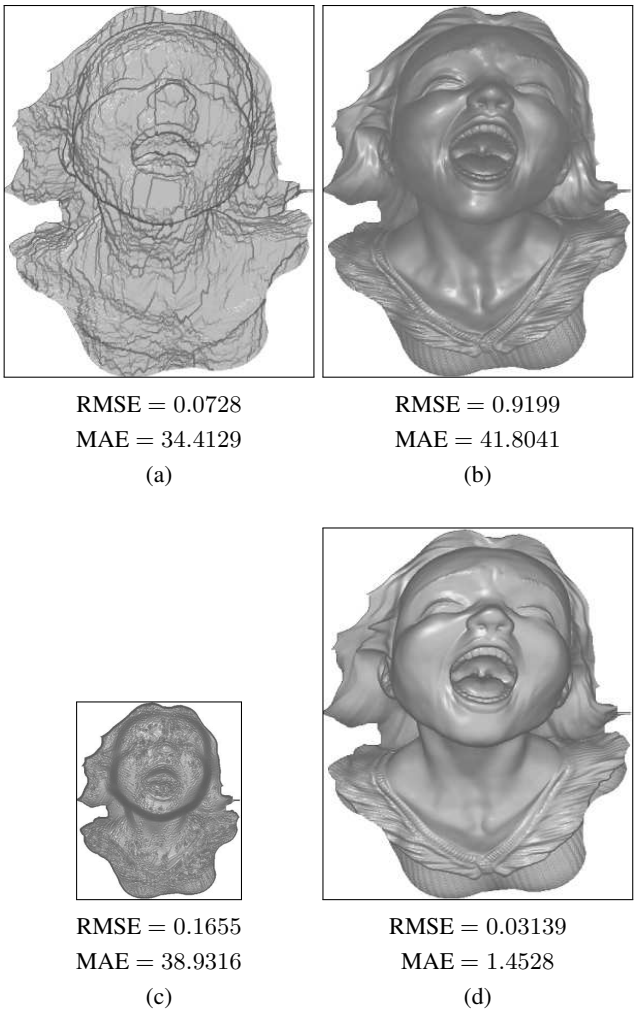
Figure 6: (a-b) Impact of the amount of zero-mean, Gaussian noise added to the RGB images on accuracy. (c-d) Same, with increasing amount of noise on the LR depth maps. Accuracy deteriorates linearly with respect to both noise levels. However, typical values for α_z are around 10^{-5} in real-world scenarios using a Microsoft Kinect v1 [14], so our method is more robust to this type of noise than required.

introduce a smoothness prior on the reflectance, which is often non-realistic and induces artefacts on the depth around reflectance discontinuities. Only by controlling the lighting this prior can be avoided, and overall the proposed method, which both estimates reflectance and solves the photometric ambiguities, provides the best results.

4.2. Qualitative Evaluation on Real-world Datasets

For real-world experiments, we use the Asus Xtion Pro Live, which has the same depth sensor as Microsoft's Kinect v1. The sensors provide a maximum RGB resolution of $1280 \times 1024\ px$ and QVGA ($320 \times 240\ px$) or VGA ($640 \times 480\ px$) depth resolution. Data is acquired in video mode, while moving a single white Luxeon Rebel LED in front of the object. Experiments are run in an office room with ambient lighting. From each sequence, we extract a series of $n = 20$ LR depth maps and HR RGB images. From the user perspective, acquisition of data is thus extremely simple, since no calibration is required.

We consider in Figures 1, 8 and 9 five different objects: a shirt with piecewise-constant reflectance, a tablet cover with piecewise-smooth reflectance and a fine wrinkle, a partly specular vase, a creased bag with some text painting, and a backpack with very thin geometric structures and piecewise-constant reflectance.



RMSE $= 0.0728$
MAE $= 34.4129$
(a)

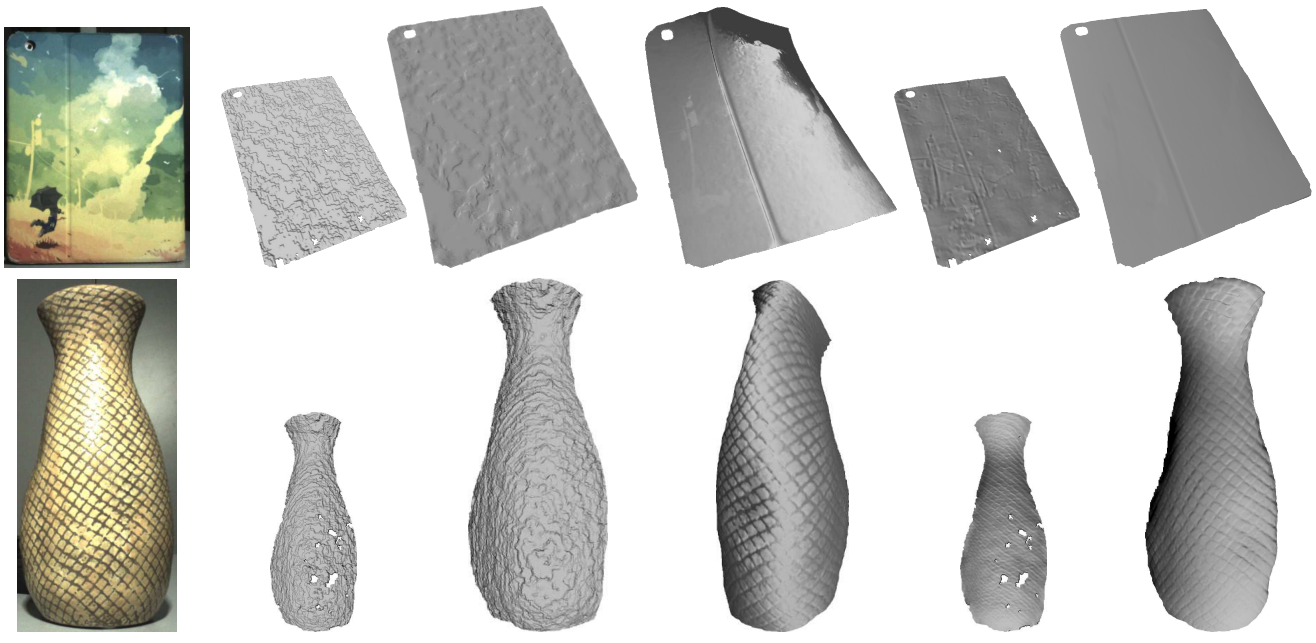
RMSE $= 0.9199$
MAE $= 41.8041$
(b)

RMSE $= 0.1655$
MAE $= 38.9316$
(c)

RMSE $= 0.03139$
MAE $= 1.4528$
(d)

Figure 7: Comparison between (a) image-driven depth super-resolution using (2) with (3), (b) uncalibrated photometric stereo [21], (c) single-image RGB-D fusion [20] and (d) the proposed photometric stereo-aware super-resolution method. Image-based depth super-resolution and RGB-D fusion are unable to appropriately handle spatially-varying reflectance. Uncalibrated photometric appropriately estimates reflectance and thin structures, but it is prone to a high low-frequency bias due to the generalized bas-relief ambiguity. The proposed photometric stereo-aware super-resolution circumvents both these issues. For the input depth map the RMSE is 0.0579 and MAE is 65.7150.

In all these cases, our method is able to successfully upsample the depth maps, while also recovering the fine geometric structures and separate reflectance from shading. Interestingly, robustness to specularities is enforced, although we only model Lambertian reflectance. This is probably due to having a rough prior on shape.



(a) (b) (c) (d) (e) (f)

Figure 8: Comparison between the proposed method and others, on real-world datasets. (a) One (out of $n = 20$) HR RGB image. (b) One of the LR depth maps (SF $= 2$). (c) Image-based depth super-resolution (Equations (2) and (3)). (d) Uncalibrated photometric stereo [21]. (e) RGB-D fusion [20]. (f) Proposed method ($\lambda = 1$). These results confirm the conclusion of the synthetic experiments in Figure 7.

(a) (b) (c) (d) (e)

Figure 9: Qualitative results on real-world datasets. (a) One of the HR RGB images. (b) One of the LR depth maps (SF $= 4$, but the LR depth maps are enlarged for the sake of illustration). (c) Estimated HR depth map (paper bag $\lambda = 40$, backpack $\lambda = 10$). (d) Estimated HR reflectance map. (e) Relighting of the HR 3D-model from new viewing and lighting angles.

5. Conclusion

We have presented a novel variational framework for depth super-resolution in RGB-D sensing, by resorting to the photometric stereo technique. For this task, it is enough to capture a sequence of low-resolution depth maps and high-resolution RGB images under uncalibrated, varying illumination. Then, the proposed variational framework is able to carry out unambiguous shape, reflectance and lighting estimation. The low-resolution depth measurements essentially disambiguate uncalibrated photometric stereo and, symmetrically, the photometric stereo-based regularization term disambiguates super-resolution. The proposed method can be used out-of-the-box using common devices, without any need for calibration. This is made possible by the tailored photometric stereo regularizer which implicitly ensures regularity of the super-resolved depth map.

For the future work, we will explore with more care the theoretical foundations of the proposed variational framework, and prove uniqueness of the solution by resorting to a continuous analysis of the problem.

References

[1] N. G. Alldrin, S. P. Mallick, and D. J. Kriegman. Resolving the generalized bas-relief ambiguity by entropy minimization. In *CVPR*, 2007.

[2] R. Anderson, B. Stenger, and R. Cipolla. Augmenting depth camera output using photometric stereo. In *MVA*, 2011.

[3] R. Basri, D. W. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *IJCV*, 72(3):239–257, 2007.

[4] P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille. The bas-relief ambiguity. *IJCV*, 35(1):33–44, 1999.

[5] A. Chatterjee and V. Madhav Govindu. Photometric refinement of depth maps for multi-albedo objects. In *CVPR*, 2015.

[6] S. Chaudhuri and M. V. Joshi. *Motion-free super-resolution*. Springer Verlag, 2005.

[7] M. Cristani, D. S. Cheng, V. Murino, and D. Pannullo. Distilling information with super-resolution for video surveillance. In *VSSN*, 2004.

[8] R. Fablet and F. Rousseau. Missing data super-resolution using non-local and statistical priors. In *ICIP*, 2015.

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

[10] H. Greenspan. Super-resolution in medical imaging. *The Computer Journal*, 52(1):43–63, 2008.

[11] Y. Han, J.-Y. Lee, and I. So Kweon. High quality shape from a single rgb-d image under uncalibrated natural illumination. In *ICCV*, 2013.

[12] H. Hayakawa. Photometric stereo under a light source with arbitrary motion. *JOSA A*, 11(11):3079–3089, 1994.

[13] B. K. P. Horn and M. J. Brooks, editors. *Shape from shading*. MIT Press, 1989.

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

[15] K. Kim, A. Torii, and M. Okutomi. Joint estimation of depth, reflectance and illumination for depth refinement. In *ICCVW*, 2015.

[16] F. Lu, X. Chen, I. Sato, and Y. Sato. SymPS: BRDF symmetry guided photometric stereo for shape and light source estimation. *PAMI*, (to appear), 2017.

[17] 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. *IJCV*, 102(1-3):18–32, 2013.

[18] R. Maier, J. Stückler, and D. Cremers. Super-resolution keyframe fusion for 3D modeling with high-quality textures. In *3DV*, 2015.

[19] A. Marquina and S. J. Osher. Image super-resolution by TV-regularization and Bregman iteration. *J. Sci. Comput.*, 37(3):367–382, 2008.

[20] R. Or-el, G. Rosman, A. Wetzler, R. Kimmel, and A. M. Bruckstein. RGBD-fusion: real-time high precision depth recovery. In *CVPR*, 2015.

[21] T. Papadhimitri and P. Favaro. A closed-form, consistent and robust solution to uncalibrated photometric stereo via local diffuse reflectance maxima. *IJCV*, 107(2):139–154, 2014.

[22] J. Park, H. Kim, Y. W. Tai, M. S. Brown, and I. S. Kweon. High-quality depth map upsampling and completion for rgb-d cameras. *TIP*, 23(12):5559–5572, 2014.

[23] Y. Quéau, T. Wu, F. Lauze, J.-D. Durou, and D. Cremers. A non-convex variational approach to photometric stereo under inaccurate lighting. In *CVPR*, 2017.

[24] P. Tan, S. Lin, and L. Quan. Subpixel photometric stereo. *PAMI*, 30(8):1460–1471, 2008.

[25] R. Y. Tsai and T. S. Huang. Multiframe image restoration and registration. *Advances in Computer Vision and Image Processing*, 1(2):317–339, 1984.

[26] M. Unger, T. Pock, M. Werlberger, and H. Bischof. A convex approach for variational super-resolution. In *DAGM*, 2010.

[27] J. D. Van Ouwerkerk. Image super-resolution survey. *Image and vision Computing*, 24(10):1039–1052, 2006.

[28] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof. Anisotropic huber-l1 optical flow. In *BMVC*, 2009.

[29] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Opt. Eng.*, 19(1):139–144, 1980.

[30] C. Wu, M. Zollhöfer, M. Nießner, M. Stamminger, S. Izadi, and C. Theobalt. Real-time shading-based refinement for consumer depth cameras. *TOG*, 33(6), 2014.

[31] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, and S. Lin. Shading-based shape refinement of rgb-d images. In *CVPR*, 2013.

[32] A. L. Yuille and D. Snow. Shape and albedo from multiple images using integrability. In *CVPR*, 1997.