RGBD-Fusion: Real-Time High Precision Depth Recovery

Roy Or - El[1] Guy Rosman[2] Aaron Wetzler[1] Ron Kimmel[1] Alfred M. Bruckstein[1]
[1]Technion, Israel Institute of Technology
[2]Computer Science and Artificial Intelligence Lab, MIT
royorel@tx.technion.ac.il rosman@csail.mit.edu twerd@cs.technion.ac.il
ron@cs.technion.ac.il freddy@cs.technion.ac.il

Abstract

The popularity of low-cost RGB-D scanners is increasing on a daily basis. Nevertheless, existing scanners often cannot capture subtle details in the environment. We present a novel method to enhance the depth map by fusing the intensity and depth information to create more detailed range profiles. The lighting model we use can handle natural scene illumination. It is integrated in a shape from shading like technique to improve the visual fidelity of the reconstructed object. Unlike previous efforts in this domain, the detailed geometry is calculated directly, without the need to explicitly find and integrate surface normals. In addition, the proposed method operates four orders of magnitude faster than the state of the art. Qualitative and quantitative visual and statistical evidence support the improvement in the depth obtained by the suggested method.

1. Introduction

The availability of affordable depth scanners has sparked a revolution in many applications of computer vision, such as robotics, human motion capture, and scene modeling and analysis. The increased availability of such scanners naturally raises the question of whether it is possible to exploit the associated intensity image to improve their lack of accuracy. To obtain fine details such as facial features, one must compensate for the measurement errors inherent in the depth scanners.

Our goal is to fuse the captured data from the RGB-D scanner in order to enhance the accuracy of the acquired depth maps. For this purpose, we must accurately align and combine both depth and scene color or intensity cues. Assuming that the scanner is stationary and its calibration parameters are known, aligning the intensity and depth data is a relatively straightforward task. Recently, scanners that allow access to both infra-red scene illumination and depth maps, have become available enabling the possibility of even richer RGB-D-I fusion.

Reconstructing a shape from color or intensity images, known as shape from shading [16, 5, 22], is a well researched area in computer vision. These shape estimation problems usually suffer from ambiguities since there can be several possible surfaces that can explain a given image. Recently, attempts have been made to eliminate some of these ambiguities by using more elaborated lighting models, and richer, natural illumination environments [19, 20]. Moreover, it was observed that data from depth sensors combined with shape from shading methods can be used to eliminate ambiguities and improve the depth maps [42, 41, 15].

We introduce a novel real-time method to directly enhance surface recovery that achieves state of the art accuracy. We apply a lighting model that uses normals estimated from the depth profile, and eliminates the need for calibration of the scene lighting. The lighting model accounts for light sources, multiple albedos, and local lighting effects such as specularities, shadows and interreflections.

Assuming that the lighting model explains the smooth nature of the intensity image, and that high frequency data in the image is related to the surface geometry, we reconstruct a high quality surface without first finding and integrating its normals. Instead, we use the relation between the surface gradient, its normals, and a smoothed version of the input depth map to define a surface dependent cost functional. In order to achieve fast convergence, we relinearize the variational problem.

The main contributions of this paper are:

1. Presenting a novel robust depth enhancement method that operates under natural illumination and handles multiple albedo objects.

2. Showing that depth accuracy can be enhanced in real-time by efficiently fusing the RGB-D inputs.

3. Showing that improved depth maps can be acquired directly using shape from shading technique that avoids

the need to first find the surface normals and then integrate them.

The paper outline is as follows: we overview previous efforts in Section 2. The proposed algorithm is presented in Section 3. Results are shown in Section 4, with disscusions in Section 5.

2. Related Work

Here, we briefly review some of the research done in depth enhancement and shape from shading. We refer to just a few representative papers that capture the major development and the state of the art in these fields.

2.1. Depth Enhancement

Depth enhancement algorithms mostly rely on one of the following strategies: using multiple depth maps, employing pre-learned depth priors and combining depth and intensity maps.

Multiple depth maps. Chen, and Medioni laid the foundation to this paradigm in [7] by registering overlapping depth maps to create an accurate and complete 3D models of objects. Digne *et al*. [10] decomposed laser scans to low and high frequency components using the intrinsic heat equation. They fuse together the low frequency components of the scans and keep the high frequency data untouched to produce a higher resolution model of an object. Merrel *et al*. [30] generated depth images from intensity videos which were later fused to create a high resolution 3D model of objects. Schuon, *et al*. [36] aligned multiple slightly translated depth maps to enhance depth resolution. They later extended this in [8] to shape reconstruction from global alignment of several super-resolved depth maps. Tong, *et al*. [38] used a non-rigid registration technique to combine depth videos from three Kinects to produce a high resolution scan of a human body. Probably the most popular effort in this area is the KinectFusion algorithm [32], in which a real time depth stream is fused on the GPU into a truncated signed distance function to accurately describe a 3D model of a scanned volume.

Pre-learned depth priors. Oisin *et al*. [28] use a dictionary of synthetic depth patches to build a high resolution depth map. Hornáček *et al*. [18] extended to 3D both the self similarity method introduced in [37] and the PatchMatch algorithm by Barnes *et al*. [1] and showed how they can be coupled to increase spatial and depth resolution. Li, *et al*. [25] extract features from a training set of high resolution color image and low resolution depth map patches, then, they learn a mapping function between the color and low resolution depth patches to the high resolution depth patches. Finally, depth resolution is enhanced by a sparse coding algorithm. In the context of 3D scanning, Rosman *et al*. [35] demonstrated the use of a sparse dictionary for range images for 3D structured-light reconstruction.

Depth and intensity maps. The basic assumption behind these methods is that depth discontinuities are strongly related to intensity discontinuities. In [26, 40] a joint bilateral upsampling of intensity images was used to enhance the depth resolution. Park *et al*. [33] combined a non-local means regularization term with an edge weighting neighborhood smoothness term and a data fidelity term to define an energy function whose minimization recovers a high quality depth map. In a more recent paper, Lee and Lee [24] used an optical flow like algorithm to simultaneously increase the resolution of both intensity and depth images. This was achieved using a single depth image and multiple intensity images from a video camera. Lu *et al*. [27] assemble similar RGBD patches into a matrix and use its low-rank estimation to enhance the depth map.

2.2. Shape from Shading

Shape from shading. The shape from shading problem under the assumption of a Lambertian surface and uniform illumination was first introduced by Horn in [16]. The surface is recovered using the characteristic strip expansion method. In 1986, Horn and Brooks [17] explored variational approaches for solving the shape from shading problem. Later, Bruckstein [5] developed a direct method of recovering the surface by level sets evolution assuming that the light source is directly above the surface. This method was later generalized by Kimmel and Bruckstein [21] to handle general cases of uniform lighting from any direction. In [22], Kimmel and Sethian show that a fast solution to the shape from shading problem can be obtained from a modification to the fast marching algorithm. The work of Mecca *et al*. [29] has recently shown a PDE formulation for direct surface reconstruction using photometric stereo. Two surveys in [43, 11] comprehensively cover the shape from shading problem as studied over the last few decades.

Recently, attempts were made to solve the shape from shading problem under uncalibrated natural illumination. Forsyth [12] modeled the shading by using a spatially slow varying light source to reconstruct the surface. Huang and Smith [19] use first order spherical harmonics to approximate the surface reflectance map. The shape is then recovered by using an edge preserving smoothing constraint and by minimizing the local brightness error. Johnson and Adelson [20] modeled the shading as a quadratic function of the surface normals. They showed, counter-intuitively, that natural illumination reduces the surface normals ambiguity and thus making the shape from shading problem simpler to solve.

Bohme *et al*. [4] imposed a shading constraint on a probalistic image formation model to find a MAP estimate of an enhanced range map. In their paper, Han *et al*. [14]

showed that shape from shading can be used in order to improve the quality of a shape obtained from a depth map under natural illumination. The initial surface normals obtained from a Kinect depth map are refined and then fused with depth information using a fusion algorithm from Nehab *et al*. [31]. Yu *et al*. [41] recover albedos and lighting of segmented intensity image and combine them to recover an accurate depth map using shape from shading. Zhang *et al*. [42] fuse depth maps and color images captured under different illumination conditions and use photometric stereo to improve the shape quality. Haque *et al*. [15] used a photometric stereo approach to reconstruct the shape normals from a given depth map and then integrate the normals to recover the object. They also suggested a method to fuse multiple reconstructions. Wu *et al*. [39] used spherical harmonics to recover the scene shading from normals obtained from a consumer depth scanners. The shape is then refined in real-time by finding the surface that minimizes the difference between the shading and intensity image gradients.

3. Shape Refinement Framework

We now propose a framework for depth refinement. The input is a depth map and a corresponding intensity image. We assume that the input depth and intensity images were taken from a calibrated fixed system. The intrinsic matrices and the extrinsic parameters of the depth and color sensors are assumed to be known.

We first wish to obtain a rough version of the input surface. However, due to measurement inaccuracies, the given depth profile is fairly noisy. For a smooth estimate, we apply a bilateral filter on the input depth map.

Next, we estimate initial surface normals corresponding to the smoothed surface. A lighting model can now be evaluated. We start by recovering the shading from the initial normals and the intensity. The subsequent step accounts for different albedos and shadows. Finally, the last step estimates varying illumination that better explains local lighting effects which only affect portions of the image. Once the lighting model is determined, we move on to enhance the surface. Like modern shape from shading methods, we take advantage of the fact that an initial depth map is given. With a depth map input, a high quality surface can be directly reconstructed without first refining the estimated normals. This is done by a variational process designed to minimize a depth based cost functional. Finally, we show how to speed up the reconstruction process.

3.1. Lighting Estimation

The shading function relates a surface geometry to its intensity image. The image is taken under natural illumination where there is no single point light source. Thus, the correct scene lighting cannot be recovered with a Lambertian model. A more complex lighting model is needed. Grosse *et al*. [13] introduced an extended intrinsic image decomposition model that has been widely used for recovering intrinsic images. We show how we can efficiently incorporate this model for our problem in order to get state of the art surface reconstruction. Define,

$$L(i,j,\vec{n}) = \rho(i,j)S(\vec{n}) + \beta(i,j), \tag{1}$$

where $L(i,j,\vec{n})$ is the image lighting at each pixel, $S(\vec{n})$ is the shading, $\rho(i,j)$ accounts for multiple scene albedos and shadowed areas since it adjusts the shading intensity. $\beta(i,j)$ is added as an independent, spatially changing, near light source, that accounts for local lighting variations such as interreflections and specularities. We note that the (i,j) indexing is sometimes omitted for convenience throughout the paper.

Clearly, since we only have a single input image, without any prior knowledge, recovering S, ρ and β for each pixel is an ill-posed problem. However, the given depth map helps us recover all three components for each pixel.

3.1.1 Shading Computation

First, we assume a Lambertian scene and recover the shading S, associated with light sources that have a uniform effect on the image. Once the shading is computed, we move on to find ρ and β, to better explain the intensity image given the object geometry. During the shading recovery process, we set ρ to 1 and β to 0.

Basri and Jacobs [3] and Ramamoorthi and Hanrahan [34] found that the irradiance of diffuse objects in natural illumination scenes can be well described by the low order spherical harmonics components. Thus, a smooth function is sufficient to recover the shading image. For the sake of simple and efficient modelling, we opt to use zero and first order spherical harmonics, which are a linear polynomial of the surface normals and are independent on the pixel's location. Therefore, they are given by

$$S(\vec{n}) = \vec{m}^T\tilde{n}, \tag{2}$$

where $\vec{n}$ is the surface normal, $S(\vec{n})$ is the shading function, $\vec{m}$ is a vector of the four first order spherical harmonics coefficients, and $\tilde{n} = (\vec{n},1)^T$.

Every valid pixel in the aligned intensity image I can be used to recover the shading. Hence, we have an overdetermined least squares parameter estimation problem

$$\underset{\vec{m}}{\operatorname{argmin}} \|\vec{m}^T\tilde{n} - I\|_2^2. \tag{3}$$

The rough normals we obtained from the initial depth map eliminate the need for assumptions and constraints on the shape or using several images. This produces a straightforward parameter fitting problem unlike the classical shape from shading and photometric stereo approaches. Despite

having only the normals of the smoothed surface we can still obtain an accurate shading model since the least square process is not sensitive to high frequency changes and subtle shape details. In addition, the estimated surface normals eliminate the need for pre-calibrating the system lighting and we can handle dynamic lighting environments.

Background normals obviously affect the shading model outcome since they are related to different materials with different albedos, hence, their irradiance is different. Nonetheless, unlike similar methods, our method is robust to such outliers as our lighting model and surface refinement scheme were designed to handle that case.

3.1.2 Multiple Albedo Recovery

The shading alone gives us only a rough assessment of the lighting, as it explains mostly distant and ambient light sources and only holds for diffuse surfaces with uniform albedo. Specularities, shadows and nearby light sources remain unaccounted for. In addition, multiple scene albedos, unbalanced lighting or shadowed areas affect the shading model by biasing its parameters. An additional cause for the errors is the rough geometry used to recover the shading model in (2). In order to handle these problems, ρ and β should be computed.

Finding ρ and β is essential to enhance the surface geometry, without them, lighting variations will be incorrectly compensated for by adjusting the shape structure. Since we now have the shading S, we can move on to recover ρ.

Now, we freeze S to the shading image we just found and optimize ρ to distinguish between the scene albedos and account for shadows (β is still set to 0). We set a fidelity term to minimize the ℓ_2 error between the proposed model and the input image. However, without regularization, $\rho(i,j)$ is prone to overfitting since one can simply set $\rho = I/S$ and get an exact explanation for the image pixels. To avoid overfitting, a prior term that prevents ρ from changing rapidly is used. Thereby, the model explains only lighting changes and not geometry changes. We follow the retinex theory [23] and like other intrinsic images recovery algorithms, we assume that the albedo map is piecewise smooth and that there is a low number of albedos in the image. Unlike many intrinsic image recovery frameworks like [2, 6], who use a Gaussian mixture model for albedo recovery we use a weighted Laplacian to distinguish between materials and albedos on the scene while maintaining the smooth changing nature of light. This penalty term is defined as

$$\left\| \sum_{k \in \mathcal{N}} \omega_k^c \omega_k^d (\rho - \rho_k) \right\|_2^2, \tag{4}$$

where $\mathcal{N}$ is the neighborhood of the pixel. ω_k^c is an intensity weighting term suggested in Equation (6) in [14],

$$\omega_k^c = \begin{cases} 0, & \|I_k - I\|_2^2 > \tau \\ \exp\left(-\dfrac{\|I_k - I(i,j)\|_2^2}{2\sigma_c^2}\right), & \text{otherwise,} \end{cases} \tag{5}$$

and ω_k^d is the following depth weighting term

$$\omega_k^d = \exp\left(-\frac{\|z_k - z(i,j)\|_2^2}{2\sigma_d^2}\right). \tag{6}$$

Here, σ_d is a parameter responsible for the allowed depth discontinuity and $z(i,j)$ represents the depth value of the respected pixel. This regularization term basically performs a three dimensional segmentation of the scene, dividing it into piecewise smooth parts. Therefore, material and albedo changes are accounted for but subtle changes in the surface are smoothed. To summarize, we have the following regularized linear least squares problem with respect to ρ

$$\min_{\rho} \|\rho S(\vec{n}) - I\|_2^2 + \lambda_\rho \| \sum_{k \in \mathcal{N}} \omega_k^c \omega_k^d (\rho - \rho_k) \|_2^2. \tag{7}$$

3.1.3 Lighting Variations Recovery

Finaly, after $\rho(i,j)$ is found we move on to find $\beta(i,j)$. A similar functional to the one used for $\rho(i,j)$ can also be used to recover $\beta(i,j)$, since specularities still maintain smooth variations. Despite that, we need to keep in mind the observation of [3, 34], first order spherical harmonics account for 87.5% of the scene lighting. Hence, we also limit the energy of $\beta(i,j)$ in order to be consistent with the shading model. Therefore, $\beta(i,j)$ is found by solving

$$\min_{\beta} \|\beta - (I - \rho S(\vec{n}))\|_2^2 + \lambda_\beta^1 \| \sum_{k \in \mathcal{N}} \omega_k^c \omega_k^d (\beta - \beta_k) \|_2^2 + \lambda_\beta^2 \|\beta\|_2^2. \tag{8}$$

3.2. Refining the Surface

At this point our complete lighting model is set to explain the scene's lighting. Now, in order to complete the recovery process, fine geometry details need to be restored. A typical SFS method would now adjust the surface normals, trying to minimize

$$\|L(i,j,\vec{n}) - I\|_2^2 \tag{9}$$

along with some regularization terms or constraints. The resulting cost function will usually be minimized in the $(p-q)$ gradient space.

However, according to [12], in order to minimize (9) schemes that use the $(p-q)$ gradient space can yield surfaces that tilt away from the viewing direction. Moreover, an error in the lighting model which can be caused by normal outliers such as background normals would aggravate this artifact. Therefore, to avoid this phenomena, we take

further advantage of the given depth map. We write the problem as a functional of z, and force the surface to change only in the viewing direction, limiting the surface distortion and increasing the robustness of the method to lighting model errors.

We use the geometric relation between surface normals and the surface gradient given by

$$\vec{n} = \frac{(z_x, z_y, -1)}{\sqrt{1 + \|\nabla z\|^2}}, \tag{10}$$

where

$$z_x = \frac{dz}{dx}, \quad z_y = \frac{dz}{dy}, \tag{11}$$

to directly enhance the depth map. The surface gradient, represented as a function of z, connects between the intensity image and the lighting model. Therefore, by fixing the lighting model parameters and allowing the surface gradient to vary, subtle details in the surface geometry can be recovered by minimizing the difference between the measured intensity image and our shading model,

$$\|L(\nabla z) - I\|_2^2. \tag{12}$$

Formulating the shape from shading term as function of z simplifies the numerical scheme and reduces ambiguities. Since we already have the rough surface geometry, only simple fidelity and smoothness terms are needed to regularize the shading. Therefore, our objective function for surface refinement is

$$f(z) = \|L(\nabla z) - I\|_2^2 + \lambda_z^1 \|z - z_0\|_2^2 + \lambda_z^2 \|\Delta z\|_2^2, \tag{13}$$

Where z_0 is the initial depth map and Δ represents the Laplacian of the surface. Using a depth based numerical scheme instead of $(p-q)$ gradient space scheme, makes our algorithm less sensitive to noise and more robust to lighting model errors caused by normal outliers. This has a great implication in handling real-world scenarios where the desired shape cannot be easily distinguished from its background.

The functional introduced is non-linear due to the shading term since the dependency between the surface normals and it's gradient requires geometric normalization. A solution to (13) can be found using the Levenberg-Marquadt algorithm or various Trust-Region methods, however, their convergence is slow and not suitable for real-time applications.

In order to accelerate the performance of the algorithm, we reformulate the problem in a similar way to IRLS optimization scheme. We do so by freezing non-linear terms inside the shading model. This allows us to solve a linear system at each iteration, and update the non-linear terms at the end of each iteration. First, we recall Equation (10) and

Algorithm 1: Accelerated Surface Enhancement

Input: $z_0, \vec{m}, \rho, \beta$ - initial surface, lighting parameters
while $f(z^{k-1}) - f(z^k) > 0$ **do**
  Update $\tilde{n}^k = (\vec{n}^k, 1)^T$
  Update $L(\nabla z^k) = \rho(\vec{m}^T \tilde{n}^k) + \beta$
  Update z^k to be the minimizer of $f(z^k)$
end

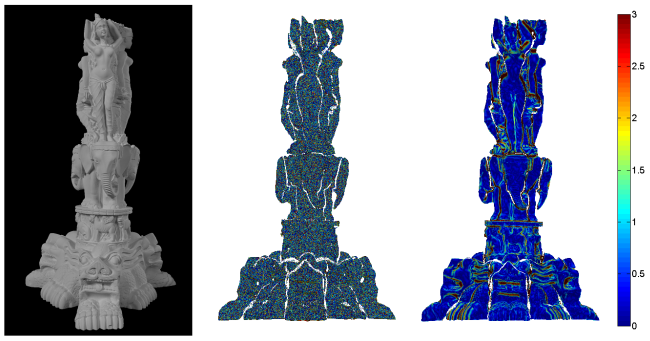


Figure 1: Thai Statue error analysis. From left to right: Input color image. Error image of the raw depth map. Error image of the final result. Note how the algorithm reduces the initial surface errors.

eliminate the denumerator using the auxiliary variables

$$\begin{aligned} \vec{n}^k &= w^k (z_x^k, z_y^k, -1)^T, \\ w^k &= (1 + \|\nabla z^{k-1}\|^2)^{-\frac{1}{2}} \end{aligned} \tag{14}$$

The new lighting linearized model reads

$$L(i, j, \nabla z) = \rho(i, j) \cdot (\vec{m}^T \tilde{n}^k) + \beta(i, j). \tag{15}$$

This results an updated shading term. Now, at each iteration we need to solve the following functional for z^k,

$$\begin{aligned} f(z^k) =& \|\rho(\vec{m}^T \tilde{n}^k) - (I - \beta)\|_2^2 \\ &+ \lambda_z^1 \|z^k - z_0\|_2^2 + \lambda_z^2 \|\Delta z^k\|_2^2. \end{aligned} \tag{16}$$

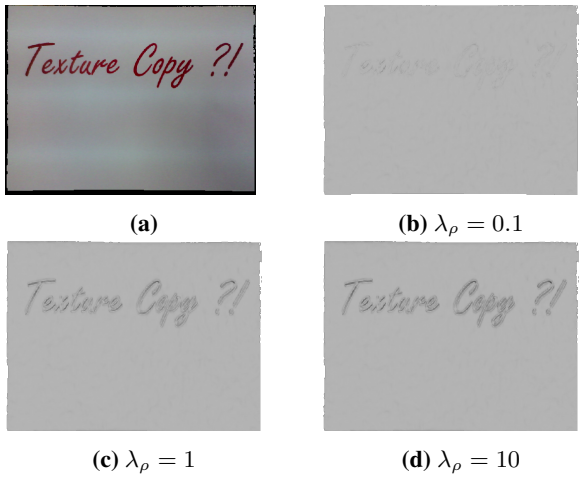
This process is repeated as long as the objective function $f(z)$ decreases. A detailed explanation of the update rule can be found in Algorithm 1.

4. Results

In order to test the proposed algorithm we performed a series of experiments to validate its efficiency and accuracy. We show that our results are quantitatively and visually state of the art, using both synthetic and real data. In addition, we display the ability of our algorithm to avoid texture copy artifacts, handle multiple albedo objects, demonstrate the robustness of our algorithm to background normals outliers, and present a runtime profile of the proposed method.

	Median				90^{th} %			
	Initial	Han *et al.*	Wu *et al.*	Proposed	Initial	Han *et al.*	Wu *et al.*	Proposed
Thai Statue	1.014	0.506	0.341	**0.291**	2.463	2.298	1.831	**1.585**
Lincoln	1.012	0.386	0.198	**0.195**	2.461	1.430	0.873	**0.866**
Coffee	1.013	0.470	0.268	**0.253**	2.473	2.681	2.454	**1.309**
C-3PO	1.013	0.344	**0.164**	0.199	2.474	1.314	**0.899**	0.923
Cheeseburger	1.014	0.283	**0.189**	0.208	2.466	1.561	1.160	**1.147**

Table 1: Quantitative comparison of depth accuracy on simulated models.

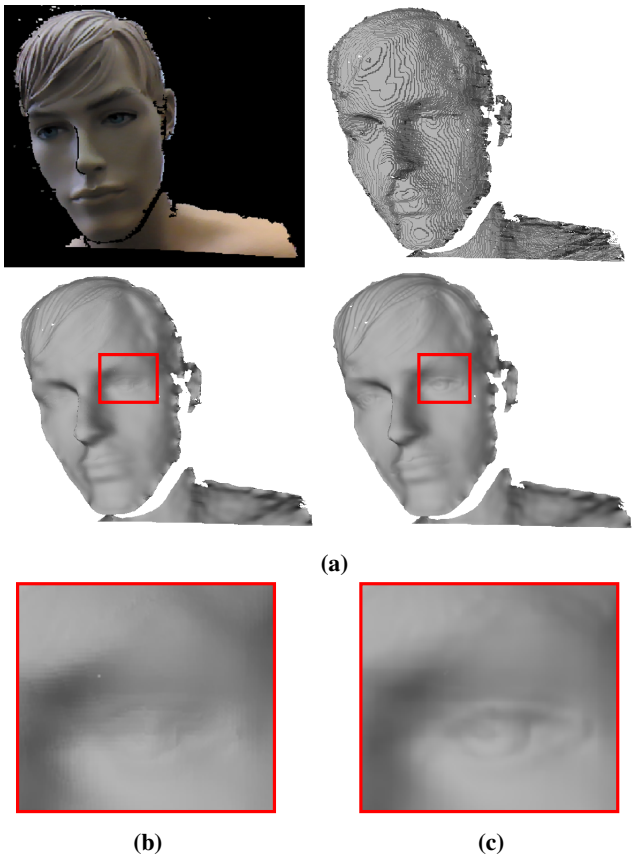


(a) **(b)** $\lambda_\rho = 0.1$ **(c)** $\lambda_\rho = 1$ **(d)** $\lambda_\rho = 10$

Figure 2: Texture copy. A correct albedo recovery model (b) mitigates texture copy artifact which the input figure (a) is prone to. The implications of poorly chosen albedo model can be easily seen in reconstructions (c) and (d). We note that $\lambda_\rho = 0.1$ was used throughout section 4.

First, we start by performing a quantitative comparison between our method and our implementation of the methods proposed by [14] and [39] which will be referred to as HLK and WZNSIT respectively. In this experiment we use synthetic data in order to have a reference model. We took objects from the Stanford 3D repository [9] and the Smithsonian 3D archive and simulated a complex lighting environment using Blender. In addition, we also used complete 3D models and scenes from the public Blendswap repository. Each model is used to test a different scenario, "Thai Statue"[1] tests a Lambertian object in a three-point lighting environment with minimal shadows. "Lincoln"[2] tests a Lambertian object in a complex lighting environment with multiple casted shadows. "Coffee"[3] involves a complex scene with a coffee mug and splashed liquid. "C3PO"[4] is a non-Lambertian object with a point light source. "Cheeseburger"[5] is a non-Lambertian, multiple albedo object with three-point lighting.

[1] http://graphics.stanford.edu/data/3Dscanrep
[2] http://3d.si.edu/downloads/27
[3] http://www.blendswap.com/blends/view/56136
[4] http://www.blendswap.com/blends/view/48372
[5] http://www.blendswap.com/blends/view/68651

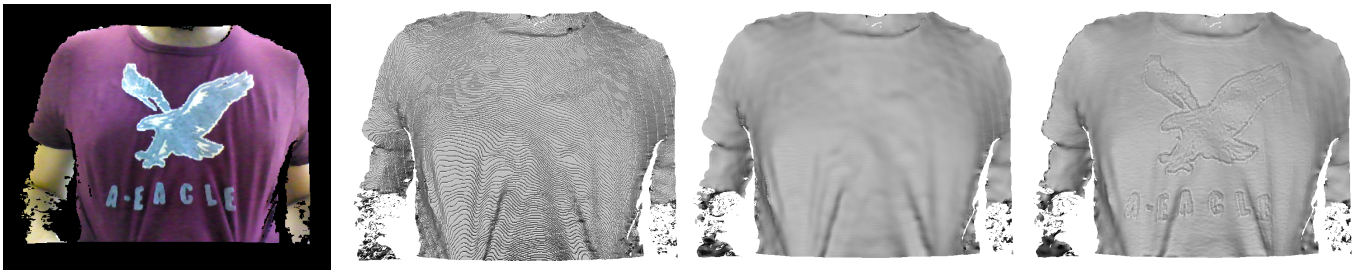


(a) **(b)** **(c)**

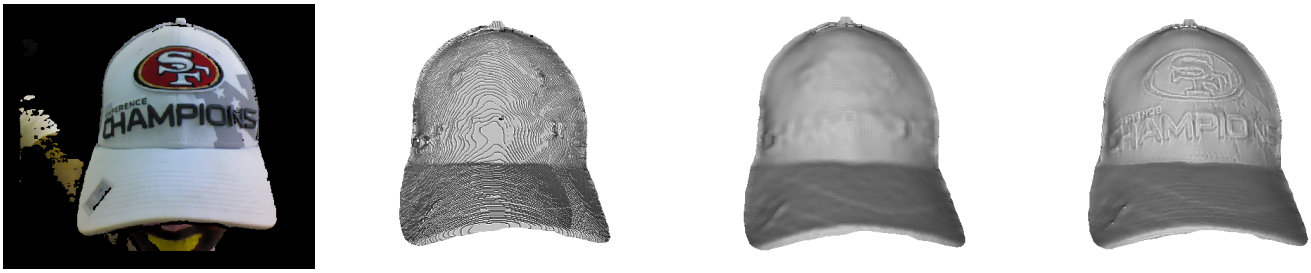
Figure 3: Mannequin. (a) Upper Left: Color Image. Upper Right: Raw Depth. Bottom Left: Result of Wu *et al.* Bottom Right: Our Result. (b), (c) Magnifications of the mannequin eye. The mannequin's hair and facial features can be easily recognized in our reconstruction.

All models were rendered with Cycles renderer of Blender[6]. We added Gaussian noise with zero mean and standard deviation of 1.5 to the depth maps to simulate a depth sensor noise. The algorithm parameters were set to $\lambda_\rho = 0.1$, $\lambda_\beta^1 = 1$, $\lambda_\beta^2 = 1$, $\tau = 0.05$, $\sigma_c = \sqrt{0.05}$, $\sigma_d = \sqrt{50}$, $\lambda_z^1 = 0.004$, $\lambda_z^2 = 0.0075$, these values were carried throughout all our experiments. We evaluate the performance of each method by measuring the median of the depth error, and the 90^{th} percentile of the depth error com-

[6] www.blender.org



(a) Shirt



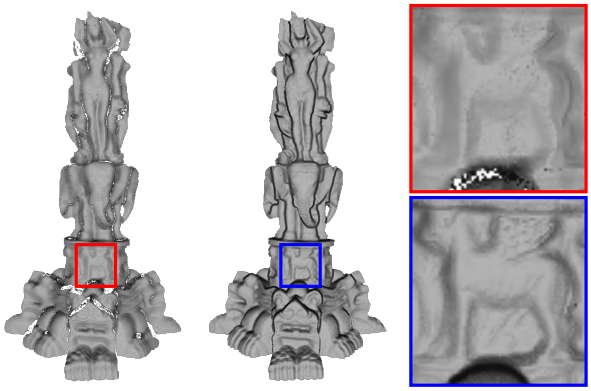
(b) Baseball Cap

Figure 4: Results of shape enhancement of real world multiple albedo objects. Left to right: Color Image, Raw Depth, Bilateral Filtering and the Proposed Method. Note how surface wrinkles and small surface protrusions are now visible.

pared to the ground truth. The results are summarized in Table 1. An example of the accuracy improvement of our method can be easily seen in Figure 1, which compares between the Thai Statue input errors and the output errors with respect to the ground truth.

We now show the qualitative results of the proposed framework from real data, captured by Intel's Real-Sense RGB-D sensor. First, we show how the proposed method handles texture, which usually lead to artifacts in shape from shading methods. In our experiment, we printed a text on a white page and captured it with our RGB-D scanner. Figure 2 shows how the texture copy artifact is mitigated by correctly modeling the scene albedos using λ_ρ. Figure 3 compares between the reconstruction results of WZN-SIT and the proposed framework in a real world scenario of a mannequin captured under natural lighting. The proposed reconstruction procedure better captures the fine details. Figure 4 illustrates how our algorithm handles real world shapes with multiple albedos. The algorithm successfully reveals the letters and eagle on the shirt along with the "SF" logo, "champions" and even the stitches on the baseball cap. One should also notice that the algorithm was slightly confused by the grey "N" which is printed on the cap but do not stick out of it like the rest of the writing. We expect such results to be improved with stronger priors, however, incorporating such priors into real time systems is beyond the scope of this paper.

Next, we show the robustness of our method to normal outliers. Such robustness is important for real-time performance, where segmenting the shape may cost precious time. In turn, background normals distort the shading recovery



(a) **(b)** **(c)**

Figure 5: Robustness to normal outliers: Left to right: HLK reconstruction with the entire depth map normals (a). Our method reconstruction with the entire depth map normals (b). Magnification of the results is presented in (c). The proposed method yield accurate reconstruction despite the distorted shading.

process, which might degrade the surface reconstruction. In this experiment, we run the method using the normals of the entire depth map. Thus, we deliberately distort the shading estimation process and examine how it affects the algorithm output. We ran this test on our method and on HLK. The results are presented in Figure 5. We see that the proposed framework can gracefully handle a large amount of normal outliers, hence, it can be used in real-time scenarios with no need to separate the object from its background.

In Figure 6 we can see how our method produces high quality reconstruction of a multiple albedo object without any prior knowledge of the shape or albedos. This is also

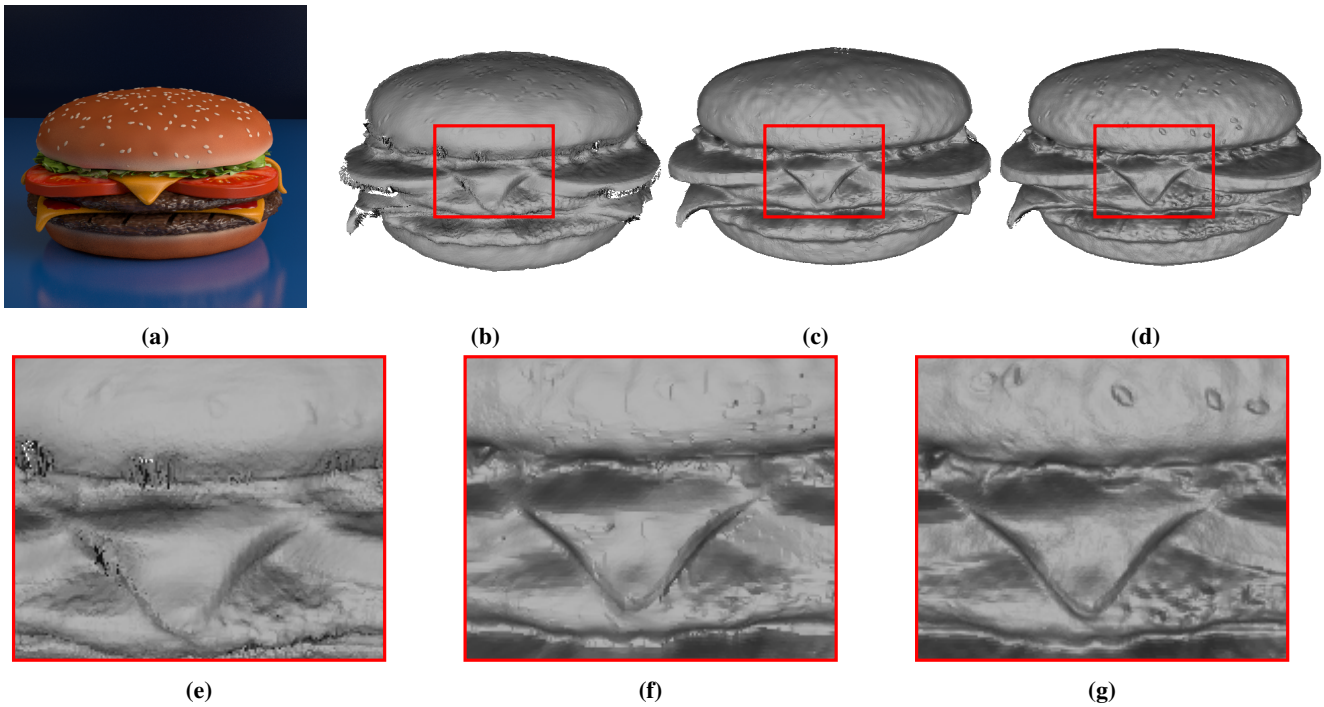


Figure 6: Handling a multiple albedo object. (a) Color Image. (b) HLK Reconstruction. (c) WZNSIT Reconstruction. (d) Our Reconstruction. (e) - (g) Magnifications of HLK, WZNSIT and Our Method respectively. Note how the proposed framework sharply distinguish albedo changes.

Section	Time
Bilateral Filter	3.8ms
Image alignment	31.1ms
Normal Estimation	5.3ms
Lighting Recovery	40.3ms
Surface Refinement	22.6ms
Total Runtime	103.1ms

Table 2: Algorithm's profiling. Please see supplementary material for a video of the real-time application.

crucial aspect for real-time performance and everyday use in dynamic scenes.

Finally, we introduce a profiling for the shape from RGBD method. An unoptimized implementation of the algorithm was tested on an Intel i7 3.4GHz processor with 16GB RAM and an Nvidia Geforce GTX TITAN GPU. The entire process runs at about 10 fps for a 640×480 depth profiles. The time breakdown of our algorithm is given in Table 2. A demo of the real-time algorithm implementation can be found in the supplementary material.

5. Conclusions

We introduced a novel computational approach that recovers details of a given rough surface using its intensity image. Our method is the first to reconstruct explicitly the surface profiles without integrating normals. Furthermore, thanks to an efficient optimization scheme the algorithm runs at about 10 frames per second. The proposed framework is more accurate than reported state of the art and runs approximately 20000 times faster.

Acknowledgements

We wish to thank the anonymous reviewers for their comments which helped us refine our paper and David Dovrat for his help with the implementation. This research was supported by European Community's FP7-ERC program grant agreement no. 267414 and by the Broadcom Foundation. G.R is partially funded by VITALITE Army Research Office Multidisciplinary Research Initiative program, award W911NF-11-1-0391.

References

[1] C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman. PatchMatch: a randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics-TOG*, 28(3):24, 2009. 2

[2] J. T. Barron, , and J. Malik. Shape, albedo, and illumination from a single image of an unknown object. In *Computer Vision and Pattern Recognition*, pages 334–341, Washington, DC, USA, 2012. IEEE Computer Society. 4

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

[4] M. Böhme, M. Haker, T. Martinetz, and E. Barth. Shading constraint improves accuracy of time-of-flight measurements. *Computer vision and image understanding*, 114(12):1329–1335, 2010. 2

[5] A. M. Bruckstein. On shape from shading. *Computer Vision, Graphics, and Image Processing*, 44(2):139–154, 1988. 1, 2

[6] J. Chang, R. Cabezas, and J. W. Fisher III. Bayesian nonparametric intrinsic image decomposition. In *European Conference on Computer Vision 2014*, pages 704–719. Springer. 4

[7] Y. Chen and G. Medioni. Object modelling by registration of multiple range images. *Image and vision computing*, 10(3):145–155, 1992. 2

[8] Y. Cui, S. Schuon, D. Chan, S. Thrun, and C. Theobalt. 3D shape scanning with a time-of-flight camera. In *IEEE Conference on Computer Vision and Pattern Recognition, 2010*, pages 1173–1180. 2

[9] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *Proceedings of the ACM conference on Computer graphics and interactive techniques, SIGGRAPH*, pages 303–312, 1996. 6

[10] J. Digne, J.-M. Morel, N. Audfray, and C. Lartigue. High fidelity scan merging. In *Computer Graphics Forum*, volume 29, pages 1643–1651. Wiley Online Library, 2010. 2

[11] 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. 2

[12] D. A. Forsyth. Variable-source shading analysis. *International Journal of Computer Vision*, 91(3):280–302, 2011. 2, 4

[13] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman. Ground-truth dataset and baseline evaluations for intrinsic image algorithms. In *International Conference on Computer Vision*, pages 2335–2342, 2009. 3

[14] Y. Han, J. Y. Lee, and I. S. Kweon. High quality shape from a single RGB-D image under uncalibrated natural illumination. In *IEEE International Conference on Computer Vision*, pages 1617–1624, 2013. 2, 4, 6

[15] S. Haque, A. Chatterjee, V. M. Govindu, et al. High quality photometric reconstruction using a depth camera. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2283–2290, 2014. 1, 3

[16] B. K. Horn. *Shape from shading: A method for obtaining the shape of a smooth opaque object from one view*. PhD thesis, 1970. 1, 2

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

[18] M. Hornáček, C. Rhemann, M. Gelautz, and C. Rother. Depth super resolution by rigid body self-similarity in 3D. In *IEEE Conference on Computer Vision and Pattern Recognition, 2013*, pages 1123–1130. 2

[19] R. Huang and W. A. Smith. Shape-from-shading under complex natural illumination. In *18th IEEE International Conference on Image Processing, 2011*, pages 13–16, 2011. 1, 2

[20] M. K. Johnson and E. H. Adelson. Shape estimation in natural illumination. In *IEEE Conference on Computer Vision and Pattern Recognition, 2011*, pages 2553–2560. 1, 2

[21] R. Kimmel and A. M. Bruckstein. Tracking level sets by level sets: a method for solving the shape from shading problem. *Computer Vision and Image Understanding*, 62(1):47–58, 1995. 2

[22] R. Kimmel and J. A. Sethian. Optimal algorithm for shape from shading and path planning. *Journal of Mathematical Imaging and Vision*, 14(3):237–244, 2001. 1, 2

[23] E. H. Land and J. J. McCann. Lightness and retinex theory. *J. Opt. Soc. Am.*, 61(1):1–11, Jan 1971. 4

[24] H. S. Lee and K. M. Lee. Simultaneous super-resolution of depth and images using a single camera. In *IEEE Conference on Computer Vision and Pattern Recognition, 2013*, pages 281–288. 2

[25] Y. Li, T. Xue, L. Sun, and J. Liu. Joint example-based depth map super-resolution. In *IEEE International Conference on Multimedia and Expo, 2012*, pages 152–157. 2

[26] M.-Y. Liu, O. Tuzel, and Y. Taguchi. Joint geodesic upsampling of depth images. In *IEEE Conference on Computer Vision and Pattern Recognition, 2013*, pages 169–176. 2

[27] S. Lu, X. Ren, and F. Liu. Depth enhancement via low-rank matrix completion. pages 3390–3397, 2014. 2

[28] O. Mac Aodha, N. D. Campbell, A. Nair, and G. J. Brostow. Patch based synthesis for single depth image super-resolution. In *European Conference on Computer Vision, 2012*, pages 71–84. Springer, 2012. 2

[29] R. Mecca, A. Wetzler, R. Kimmel, and A. M. Bruckstein. Direct shape recovery from photometric stereo with shadows. In *3DTV-Conference, 2013 International Conference on*, pages 382–389. IEEE, 2013. 2

[30] P. Merrell, A. Akbarzadeh, L. Wang, P. Mordohai, J.-M. Frahm, R. Yang, D. Nistér, and M. Pollefeys. Real-time visibility-based fusion of depth maps. In *IEEE 11th International Conference on, Computer Vision, 2007*, pages 1–8. 2

[31] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3D geometry. In *ACM Transactions on Graphics*, volume 24, pages 536–543, 2005. 3

[32] R. A. Newcombe, A. J. Davison, S. Izadi, P. Kohli, O. Hilliges, J. Shotton, D. Molyneaux, S. Hodges, D. Kim, and A. Fitzgibbon. KinectFusion: Real-time dense surface mapping and tracking. In *IEEE international symposium on Mixed and augmented reality*, pages 127–136, 2011. 2

[33] J. Park, H. Kim, Y.-W. Tai, M. S. Brown, and I. Kweon. High quality depth map upsampling for 3D-TOF cameras. In *IEEE International Conference on Computer Vision*, pages 1623–1630, 2011. 2

[34] R. Ramamoorthi and P. Hanrahan. An efficient representation for irradiance environment maps. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 497–500. ACM, 2001. 3, 4

[35] G. Rosman, A. Dubrovina, and R. Kimmel. Sparse modeling of shape from structured light. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIM-*

PVT), 2012 Second International Conference on, pages 456–463. IEEE, 2012. 2

[36] S. Schuon, C. Theobalt, J. Davis, and S. Thrun. Lidarboost: Depth superresolution for tof 3D shape scanning. In *IEEE Conference on Computer Vision and Pattern Recognition, 2009*, pages 343–350. 2

[37] E. Shechtman and M. Irani. Matching local self-similarities across images and videos. In *IEEE Conference on Computer Vision and Pattern Recognition, 2007.*, pages 1–8. 2

[38] J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan. Scanning 3D full human bodies using kinects. *IEEE Transactions on Visualization and Computer Graphics*, 18(4):643–650, 2012. 2

[39] C. Wu, M. Zollhfer, M. Niener, M. Stamminger, S. Izadi, and C. Theobalt. Real-time shading-based refinement for consumer depth cameras. In *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2014)*, volume 33, December 2014. 3, 6

[40] Q. Yang, R. Yang, J. Davis, and D. Nistér. Spatial-depth super resolution for range images. In *IEEE Conference on Computer Vision and Pattern Recognition, 2007*, pages 1–8. 2

[41] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, and S. Lin. Shading-based shape refinement of RGB-D images. In *IEEE Conference on Computer Vision and Pattern Recognition, 2013*, pages 1415–1422. 1, 3

[42] Q. Zhang, M. Ye, R. Yang, Y. Matsushita, B. Wilburn, and H. Yu. Edge-preserving photometric stereo via depth fusion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2472–2479, 2012. 1, 3

[43] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah. Shape-from-shading: a survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(8):690–706, 1999. 2