Real-Time Depth Refinement for Specular Objects

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

Abstract

The introduction of consumer RGB-D scanners set off a major boost in 3D computer vision research. Yet, the precision of existing depth scanners is not accurate enough to recover fine details of a scanned object. While modern shading based depth refinement methods have been proven to work well with Lambertian objects, they break down in the presence of specularities. We present a novel shape from shading framework that addresses this issue and enhances both diffuse and specular objects' depth profiles. We take advantage of the built-in monochromatic IR projector and IR images of the RGB-D scanners and present a lighting model that accounts for the specular regions in the input image. Using this model, we reconstruct the depth map in real-time. Both quantitative tests and visual evaluations prove that the proposed method produces state of the art depth reconstruction results.

1. Introduction

The introduction of commodity RGB-D scanners marked the beginning of a new age for computer vision and computer graphics. Despite their popularity, such scanners can obtain only the rough geometry of scanned surfaces due to limited depth sensing accuracy. One way to mitigate this limitation is to refine the depth output of these scanners using the available RGB and IR images.

A popular approach to surface reconstruction from image shading cues is the Shape from Shading (SfS). Shape reconstruction from a single image is an ill-posed problem since beside the surface geometry, the observed image also depends on properties like the surface reflectance, the lighting conditions and the viewing direction. Incorporating data from depth sensors has proved to be successful in eliminating some of these ambiguities [7, 22, 12]. However, many of these efforts are based on the assumption that the scanned surfaces are fully Lambertian, which limits the variety of objects they can be applied to. Directly applying such methods to specular objects introduces artifacts to the surface in highly specular regions due to the model's inability to account for sudden changes in image intensity.

Here, we propose a novel real-time framework for depth enhancement of non-diffuse surfaces. To that end, we use the IR image supplied by the depth scanners. The narrowband nature of the IR projector and IR camera provides a controlled lighting environment. Unlike previous approaches, we exploit this friendly environment to introduce a new lighting model for depth refinement that accounts for specular reflections as well as multiple albedos. To enable our real-time method we directly enhance the depth map by using an efficient optimization scheme which avoids the traditional normals refinemet step.

The paper outline is as follows: Section 2 reviews previous efforts in the field. An overview of the problem is presented in Section 3. The new method is introduced in Section 4, with results in Section 5. Section 6 concludes the paper.

2. Related Efforts

The classical SfS framework assumes a Lambertian object with constant albedo and a single, distant, lighting source with known direction. There are several notable methods which solve the classical SfS problem. These can be divided into two groups: propagation methods and variational ones. Both frameworks were extensively researched during the last four decades. Representative papers from each school of thought are covered in [24, 6].

The main practical drawback about classical shape from shading, is that although a diffusive single albedo setup can be easily designed in a laboratory, it can be rarely found in more realistic environments. As such, modern SfS approaches attempt to reconstruct the surface without any assumptions about the scene lighting and/or the object albedos. In order to account for the unknown scene conditions, these algorithms either use learning techniques to construct priors for the shape and scene parameters, or acquire a rough depth map from a 3D scanner to initialize the surface.

Learning based methods. Barron and Malik [1] constructed priors from statistical data of multiple images to recover the shape, albedo and illumination of a given input image. Kar *et al*. [10] learn 3D deformable models from 2D annotations in order to recover detailed shapes. Richter and Roth [15] extract color, textons and silhouette features from a test image to estimate a reflectance map from which patches of objects from a database are rendered and used in a learning framework for regression of surface normals. Although these methods produce excellent results, they depend on the quality and size of their training data, whereas the proposed axiomatic approach does not require a training stage and is therefore applicable in more general settings.

Depth map based methods. Bohme *et al*. [3] find a MAP estimate of an enhanced range map by imposing a shading constraint on a probalistic image formation model. Yu *et al*. [23] use mean shift clustering and second order spherical harmonics to estimate the fdepth map scene albedos and lighting from a color image. These estimations are then combined together to improve the given depth map accuracy. Han *et al*. [7] propose a quadratic global lighting model along with a spatially varying local lighting model to enhance the quality of the depth profile. Kadambi *et al*. [9] fuse normals obtained from polarization cues with rough depth maps to obtain accurate reconstructions. Even though this method can handle specular surfaces, it requires at least three photos to reconstruct the normals and it does not run in real-time. Several IR based methods were introduced in [8, 5, 4, 19]. The authors of [8, 4] suggest a multi-shot photometric stereo approach to reconstruct the object normals. Choe *et al*. [5] refine 3D meshes from Kinect Fusion [11] using IR images captured during the fusion pipeline. Although this method can handle uncalibrated lighting, it is niether one-shot nor real-time since a mesh must first be acquired before the refinement process begins. Ti *et al*. [19] propose a simultaneous time-of flight and photometric stereo algorithm that utilizes several light sources to produce accurate surface and surface normals. Although this method can be implemented in real time, it requires four shots per frame for reconstruction as opposed to our single shot approach. More inline with our approach, Wu *et al*. [22] use second order spherical harmonics to estimate the global scene lighting, which is then followed by efficient scheme to reconstruct the object. In [12] Or - El *et al*. introduced a real-time framework for direct depth refinement that handles natural lighting and multiple albedo objects. Both algorithms rely on shading cues from an RGB image taken under uncalibrated illumination with possibly multiple light sources. Correctly modeling image specularities under such conditions is difficult. We propose to overcome the light source ambiguity issue by using the availability of a single IR source with known configuration.

3. Overview

Shape from Shading (SfS) tries to relate an object's geometry to its image irradiance. Like many other inverse problems, SfS is also an ill-posed one because the per-pixel image intensity is determined by several elements: the surface geometry, its albedo, scene lighting, the camera parameters and the viewing direction.

When using depth maps from RGB-D scanners one could recover the camera parameters and viewing direction, yet, in order to obtain the correct surface, we first need to account for the scene lighting and the surface albedos. Failing to do so would cause the algorithm to change the surface geometry and introduce undesired deformations. Using cues from an RGB image under uncalibrated illumination like [7, 22, 12] requires an estimation of global lighting parameters. Although such estimations work well for diffuse objects, they usually fail when dealing with specular ones and result in a distorted geometry. The reason is that specularities are sparse outliers that are not accounted for by classical lighting models. Furthermore, trying to use estimated lighting directions to model specularities is prone to fail when there are multiple light sources in the scene.

In our scenario, the main lighting in the IR image comes from the scanner's projector, which can be treated as a point light source. Observe that in this setting, we do not need to estimate a global lighting direction, instead, we use a near light field model to describe the per-pixel lighting direction. Subsequently, we can also account for specularities and non-uniform albedo map.

In our setting, an initial depth estimation is given by the scanner. We avoid the process of computing a refined normal field and then fusing depth with normal estimates, which is common to SfS methods, and solve directly for the depth. This eliminates the need to enforce integrability and reduces the problem size by half. We deal with the nonlinear part by calculating a first order approximation of the cost functional and thereby achieve real-time performance.

4. Proposed Framework

A novel IR based real-time framework for depth enhancement is proposed. The suggested algorithm requires a depth map and an IR image as inputs. We assume that the IR camera and the depth camera have the same intrinsic parameters, as is usually the case with common depth scanners. In addition, we also assume that the whole system is calibrated and that the translation vector between the scanner's IR projector and IR camera is known.

Unfortunately, the raw depth map is usually quantized and the surface geometry is highly distorted. Therefore, we first smooth the raw depth map and estimate the surface nor-

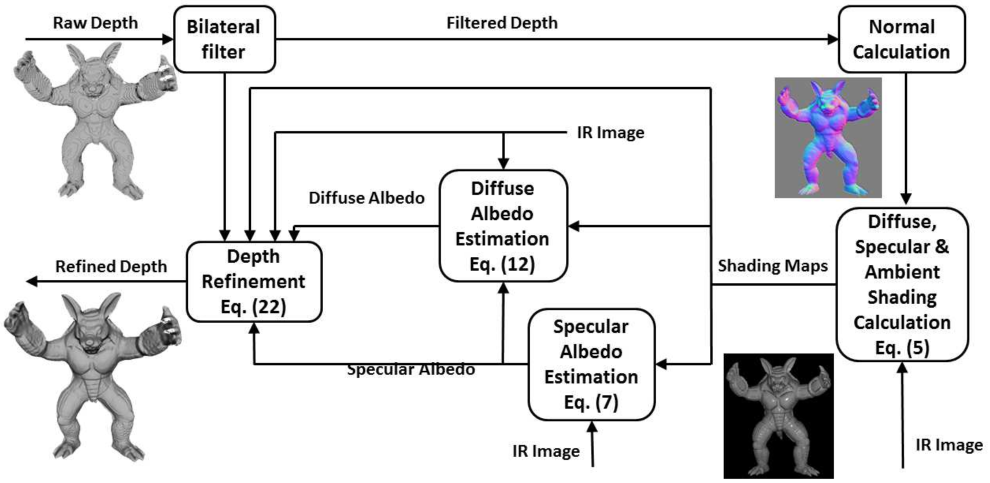
Raw Depth
Bilateral filter
Filtered Depth
Normal Calculation
IR Image
Diffuse Albedo
Diffuse Albedo Estimation Eq. (12)
Refined Depth
Depth Refinement Eq. (22)
Shading Maps
Diffuse, Specular & Ambient Shading Calculation Eq. (5)
Specular Albedo
Specular Albedo Estimation Eq. (7)
IR Image
IR Image

Figure 1: Algorithm's flowchart

mals. We then move on to recover the scene lighting using a near-field lighting model which explicitly accounts for object albedos and specularities.

After we find the scene lighting along with albedo and specular maps, we can directly update the surface geometry by designing a cost functional that relates the depth and IR intensity values at each pixel. We also show how the reconstruction process can be accelerated in order to obtain real-time performance. Figure 1 shows a flowchart of the proposed algorithm.

4.1. Near Field Lighting Model

Using an IR image as an input provides several advantages to the reconstruction process. Unlike other methods which require alignment between RGB and depth images, in our case, the depth map and IR image are already aligned as they were captured by the same camera. Moreover, the narrowband nature of the IR camera means that the main light source in the image is the scanner's own IR projector whose location relative to the camera is known. Therefore, we can model the IR projector as a point light source and use a near field lighting model to describe the given IR image intensity at each pixel,

$$I = \frac{a\rho_d}{d_p^2} S_{diff} + \rho_d S_{amb} + \frac{a\rho_s}{d_p^2} S_{spec}. \tag{1}$$

Here, a is the projector intensity which is assumed to be constant throughout the image. d_p is the distance of the surface point from the projector. ρ_d and ρ_s are the diffuse and specular albedos. S_{amb} is the ambient lighting in the scene, which is also assumed to be constant over the image. S_{diff} is the diffuse shading function of the image which is given by the Lambertian reflectance model

$$S_{diff} = \vec{N} \cdot \vec{l}_p. \tag{2}$$

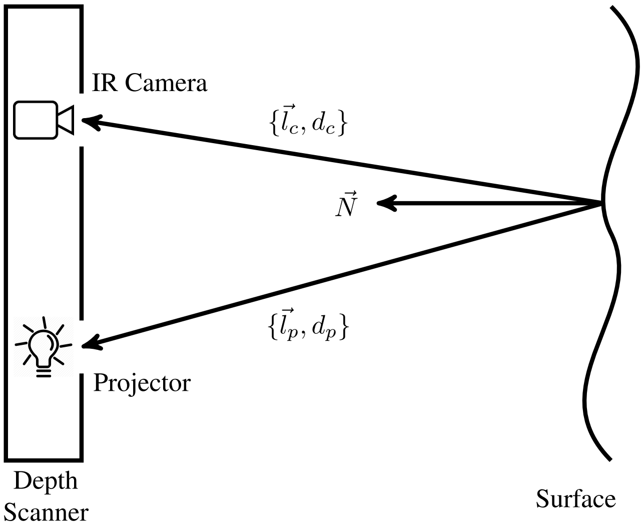

Figure 2: Scene lighting model

The specular shading function S_{spec} is set according to the Phong reflectance model

$$S_{spec} = \left(\left(2(\vec{l}_p \cdot \vec{N})\vec{N} - \vec{l}_p\right) \cdot \vec{l}_c\right)^{\alpha}, \tag{3}$$

where $\vec{N}$ is the surface normal, $\vec{l}_p, \vec{l}_c$ are the directions from the surface point to the projector and camera respectively and α is the shininess constant which we set to $\alpha = 2$. Figure 2 describes the scene lighting model. For ease of notation, we define

$$\tilde{S}_{diff} = \frac{a}{d_p^2} S_{diff}, \quad \tilde{S}_{spec} = \frac{a}{d_p^2} S_{spec}. \tag{4}$$

The intrinsic camera matrix and the relative location of the projector with respect to camera are known. In addition, the initial surface normals can be easily calculated from the given rough surface. Therefore, $\vec{l}_c, \vec{l}_p, d_p, S_{diff}$ and S_{spec}

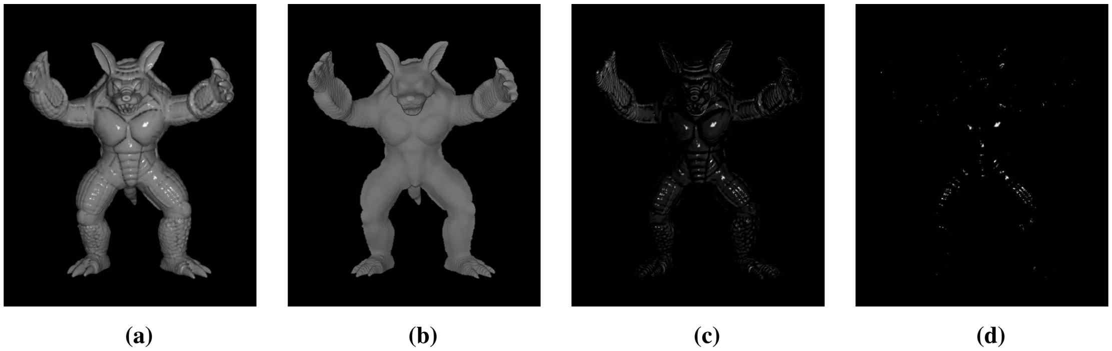

Figure 3: (a) Simulated IR image of the Armadillo mesh. (b) Recovered image of the diffuse and ambient shading $\tilde{S}_{\mathit{diff}}+S_{\mathit{amb}}$. (c) Residual image for specular albedo estimation I^s_{res}. (d) Ground Truth specularity map of (a). Note that specularities in (d) are basically the sparse representation of the residual image (c).

can be found directly whereas $a, S_{\mathit{amb}}, \rho_d$ and ρ_s need to be recovered. Although we are using a rough depth normal field to compute $\vec{l}_c, \vec{l}_p, d_p, S_{\mathit{diff}}$ and S_{spec} we still get accurate shading maps since the lighting is not sensitive to minor changes in the depth or normal field as shown in [2, 14]. Decomposing the IR image into its Lambertian and Specular lighting components along with their respective albedo maps has no unique solution. To achieve accurate results while maintaining real-time performance we choose a greedy approach which first assumes Lambertian lighting and gradually accounts for the lighting model from Eq. 1. Every pixel in the IR image which has an assigned normal can be used to recover a and S_{amb}. Generally, most of the light reflected back to the camera is related to the diffuse component of the object whereas highly specular areas usually have a more sparse nature. Thus, the specular areas can be treated as outliers in a parameter fitting scheme as they have minimal effect on the outcome. This allows us to assume that the object is fully Lambertian (i.e $\rho_d = 1, \rho_s = 0$), which in turn, gives us the following overdetermined linear system for n valid pixels ($n \gg 2$),

$$\begin{pmatrix} \frac{S^1_{\mathit{diff}}}{(d^1_p)^2} & 1 \\ \vdots & \vdots \\ \frac{S^n_{\mathit{diff}}}{(d^n_p)^2} & 1 \end{pmatrix} \begin{pmatrix} a \\ S_{\mathit{amb}} \end{pmatrix} = \begin{pmatrix} I_1 \\ \vdots \\ I_n \end{pmatrix}. \tag{5}$$

4.1.1 Specular Albedo Map

The specular shading map is important since it reveals the object areas which are likely to produce specular reflections in the IR image. Without it, bright diffuse objects can be mistaken for specularities. Yet, since $\tilde{S}_{\mathit{spec}}$ was calculated as if the object is purely specular, using it by itself will fail to correctly represent the specular irradiance, as it would falsely brighten non-specular areas. In order to obtain an accurate representation of the specularities it is essential to find the specular albedo map to attenuate the non-specular areas of $\tilde{S}_{\mathit{spec}}$.

We now show how we can take advantage of the sparse nature of the specularities to recover ρ_s and get the correct specular scene lighting. We will define a residual image I^s_{res} as being a difference between the original image I and our current diffuse approximation together with the ambient lighting. Formally, we write this as

$$I^s_{\mathit{res}} = I - (\tilde{S}_{\mathit{diff}} + S_{\mathit{amb}}). \tag{6}$$

As can be seen in Figure 3 (c), the sparse bright areas of I^s_{res} are attributable to the true specularities in I. Specular areas have finite local support, therefore we choose to model the residual image I^s_{res} as $\rho_s \tilde{S}_{\mathit{spec}}$ such that ρ_s will be a sparse specular albedo map. This will yield an image that contains just the bright areas of I^s_{res}. In addition, in order to preserve the smooth nature of specularities we add a smoothness term that minimizes the L1 Total-Variation of ρ_s. To summarize, the energy minimization problem to estimate ρ_s can be written as

$$\min_{\rho_s} \lambda^s_1 \|\rho_s \tilde{S}_{\mathit{spec}} - I^s_{\mathit{res}}\|^2_2 + \lambda^s_2 \|\rho_s\|_1 + \lambda^s_3 \|\nabla \rho_s\|_1, \tag{7}$$

where $\lambda^s_1, \lambda^s_2, \lambda^s_3$ are weighting terms for the fidelity, sparsity and smoothness terms, respectively. To minimize the cost functional, we use a variation of the Augmented Lagrangian method suggested in [21] where we substitute the frequency domain solution with a Gauss-Seidel scheme on the GPU. We refer the reader to the above paper for additional details on the optimization procedure.

4.1.2 Recovering the Diffuse Albedo

As was the case with specular shading, the diffuse shading map alone does not sufficiently explain the diffuse lighting. This is due to the fact that the diffuse shading is calculated as if there was only a single object with uniform albedo. In reality however, most objects are composed of multiple different materials with different reflectance properties that need to be accounted for.

Using the estimated specular lighting from section 4.1.1 we can now compute a residual image between the original image I and the specular scene lighting which we write as

$$I^d_{res} = I - \rho_s \tilde{S}_{spec}. \tag{8}$$

I^d_{res} should now contain only the diffuse and ambient irradiance of the original image I. This can be used in a data fidelity term for a cost functional designed to find the diffuse albedo map ρ_d.

We also wish to preserve the piecewise-smoothness of the diffuse albedo map. Otherwise, geometry distortions will be mistaken for albedos and we will not be able to recover the correct surface. The IR image and the rough depth map provide us several cues that will help us to enforce piecewise smoothness. Sharp changes in the intensity of the IR image imply a change in the material reflectance. Moreover, depth discontinuities can also signal possible changes in the albedo.

We now wish to fuse the cues from the initial depth profile and the IR image together with the piecewise-smooth albedo requirement. Past papers [7, 12] have used bilateral smoothing. Here, instead, we base our scheme on the geomtric Beltrami framework such as in [18, 17, 20] which has the advantage of promoting alignment of the embedding space channels. Let,

$$\mathcal{M}(x,y) = \{x, y, \beta_I I^d_{res}(x,y), \beta_z z(x,y), \beta_\rho \rho_d(x,y)\} \tag{9}$$

be a two dimensional manifold embedded in a $5D$ space with the metric

$$G = \begin{pmatrix} \langle \mathcal{M}_x, \mathcal{M}_x \rangle & \langle \mathcal{M}_x, \mathcal{M}_y \rangle \\ \langle \mathcal{M}_x, \mathcal{M}_y \rangle & \langle \mathcal{M}_y, \mathcal{M}_y \rangle \end{pmatrix}. \tag{10}$$

The gradient of ρ_d with respect to the $5D$ manifold is

$$\nabla_G \rho_d = G^{-1} \cdot \nabla \rho_d, \tag{11}$$

By choosing large enough values of β_I, β_z and β_ρ and minimizing the L1 Total-Variation of ρ_d with respect to the manifold metric, we basically perform selective smoothing according to the "feature" space (I^d_{res}, z, ρ_d). For instance, if $\beta_I \gg \beta_z, \beta_\rho, 1$, the manifold gradient would get small values when sharp edges are present in I^d_{res} since G^{-1} would decrease the weight of the gradient at such locations.

To conclude, the minimization problem we should solve in order to find the diffuse albedo map is

$$\min_{\rho_d} \lambda^d_1 \left\| \rho_d \left(\tilde{S}_{diff} + S_{amb} \right) - I^d_{res} \right\|_2^2 + \lambda^d_2 \|\nabla_G \rho_d\|_1. \tag{12}$$

Here, λ^d_1, λ^d_2 are weighting terms for the fidelity and piecewise-smooth penalties. We can minimize this functional using the Augmented Lagrangian method proposed in [16]. The metric is calculated separately for each pixel, therefore, it can be implemented very efficiently on a GPU with limited effect on the algorithm's runtime.

4.2. Surface Reconstruction

Once we account for the scene lighting, any differences between the IR image and the image rendered with our lighting model are attributed to geometry errors of the depth profile. Usually, shading based reconstruction algorithms opt to use the dual stage process of finding the correct surface normals and then integrating them in order to obtain the refined depth. Although this approach is widely used, it has some significant shortcomings. Calculating the normal field is an ill-posed problem with $2n$ unknowns if n is the number of pixels. The abundance of variables can result in distorted surfaces that are tilted away from the camera. In addition, since the normal field is an implicit surface representation, further regularization such as the integrability constraint is needed to ensure that the resulting normals would represent a valid surface. This additional energy minimization functional can impact the performance of the algorithm.

Instead, we use the strategy suggested in [12, 22] and take advantage of the rough depth profile acquired by the scanner. Using the explicit depth values forces the surface to move only in the direction of the camera rays, avoids unwanted distortions, eliminates the need to use an integrability constraint and saves computation time and memory by reducing the number of variables.

In order to directly refine the surface, we relate the depth values to the image intensity through the surface normals. Assuming that the perspective camera intrinsic parameters are known, the $3D$ position $P(i,j)$ of each pixel is given by

$$P\left(z(i,j)\right) = \left(\frac{j - c_x}{f_x} z(i,j), \frac{i - c_y}{f_y} z(i,j), z(i,j) \right)^T, \tag{13}$$

where f_x, f_y are the focal lengths of the camera and (c_x, c_y) is the camera's principal point. The surface normal $\vec{N}$ at each $3D$ point is then calculated by

$$\vec{N}\left(z(i,j)\right) = \frac{P_x \times P_y}{\|P_x \times P_y\|}. \tag{14}$$

We can use Eqs. (1), (2) and (14) to write down a depth based shading term written directly in terms of z,

$$E_{sh}(z) = \left\| \frac{a \rho_d}{d_p^2} (\vec{N}(z) \cdot \vec{l}_p) + \rho_d S_{amb} + \rho_s \tilde{S}_{spec} - I \right\|_2^2. \tag{15}$$

This allows us to refine z by penalizing shading mismatch with the original image I. We also use a fidelity term that penalizes the distance from the initial 3D points

$$E_f(z) = \|w(z - z_0)\|_2^2, \\ w = \sqrt{1 + \left(\frac{j - c_x}{f_x}\right)^2 + \left(\frac{i - c_y}{f_y}\right)^2}, \tag{16}$$

and a smoothness term that minimizes the second order TV-L1 of the surface

$$E_{sm}(z) = \|Hz\|_1, \quad H = \begin{pmatrix} D_{xx} \\ D_{yy} \end{pmatrix}. \tag{17}$$

Here, D_{xx}, D_{yy} are the second derivatives of the surface.

Combining Eqs. (15), (16) and (17) into a cost functional

Model	IR	NL - SH1	NL - SH2
Armadillo	**2.018**	12.813	11.631
Dragon	**3.569**	10.422	10.660
Greek Statue	**2.960**	7.241	9.067
Stone Lion	**4.428**	7.8294	8.640
Cheeseburger	**9.517**	17.881	19.346
Pumpkin	**10.006**	13.716	16.088

Table 1: Quantitative comparison of RMSE of the specular lighting estimation in IR and natural lighting scenarios. IR refers to the lighting scenario described in Section 4.1, NL - SH1/2 represents a natural lighting scenario with first/second order spherical harmonics used to recover the diffuse and ambient shading as well as $\vec{l}_p$. All values are in gray intensity units $[0, 255]$.

results in a non-linear optimization problem

$$\min_z \lambda_1^z E_{sh}(z) + \lambda_2^z E_f(z) + \lambda_3^z E_{sm}(z), \qquad (18)$$

where $\lambda_1^z, \lambda_2^z, \lambda_3^z$ are the weights for the shading, fidelity and smoothness terms, respectively. Although there are several possible methods to solve this problem, a fast scheme is required for real-time performance. To accurately and efficiently refine the surface we base our approach on the iterative scheme suggested in [13]. Rewriting Eq. (15) as a function of the discrete depth map z, and using forward derivatives we have

$$\begin{aligned} I_{i,j} - \rho_d S_{amb} - \rho_s \tilde{S}_{spec} &= \frac{a\rho_d}{d_p^2}(\vec{N}(z)\cdot\vec{l}_p) \\ &= f(z_{i,j}, z_{i+1,j}, z_{i,j+1}). \end{aligned} \qquad (19)$$

At each iteration k we can approximate f using the first order Taylor expansion about $(z_{i,j}^{k-1}, z_{i+1,j}^{k-1}, z_{i,j+1}^{k-1})$, such that

$$\begin{aligned} &I_{i,j} - \rho_d S_{amb} - \rho_s \tilde{S}_{spec} = f(z_{i,j}^k, z_{i+1,j}^k, z_{i,j+1}^k) \\ &\approx f(z_{i,j}^{k-1}, z_{i+1,j}^{k-1}, z_{i,j+1}^{k-1}) + \frac{\partial f}{\partial z_{i,j}^{k-1}}(z_{i,j}^k - z_{i,j}^{k-1}) \\ &+ \frac{\partial f}{\partial z_{i+1,j}^{k-1}}(z_{i+1,j}^k - z_{i+1,j}^{k-1}) + \frac{\partial f}{\partial z_{i,j+1}^{k-1}}(z_{i,j+1}^k - z_{i,j+1}^{k-1}). \end{aligned} \qquad (20)$$

Rearranging terms to isolate terms including z from the current iteration, we can define

$$\begin{aligned} I_{res}^{z^k} &= I_{i,j} - \rho_d S_{amb} - \rho_s \tilde{S}_{spec} \\ &\quad - f(z_{i,j}^{k-1}, z_{i+1,j}^{k-1}, z_{i,j+1}^{k-1}) + \frac{\partial f}{\partial z_{i,j}^{k-1}} z_{i,j}^{k-1} \\ &\quad + \frac{\partial f}{\partial z_{i+1,j}^{k-1}} z_{i+1,j}^{k-1} + \frac{\partial f}{\partial z_{i,j+1}^{k-1}} z_{i,j+1}^{k-1} \end{aligned}, \qquad (21)$$

and therefore minimize

$$\min_{z^k} \lambda_1^z \|Az^k - I_{res}^{z^k}\|_2^2 + \lambda_2^z \|w(z^k - z_0)\|_2^2 + \lambda_3^z \|Hz^k\|_1 \qquad (22)$$

at each iteration with the Augmented Lagrangian method of [21]. Here, A is a matrix that represents the linear operations performed on the vector z^k. Finally, we note that this pipeline was implemented on an Intel i7 3.4GHz processor with 16GB of RAM and an NVIDIA GeForce GTX650 GPU. The runtime for a 640 × 480 image is approximately 80 milliseconds.

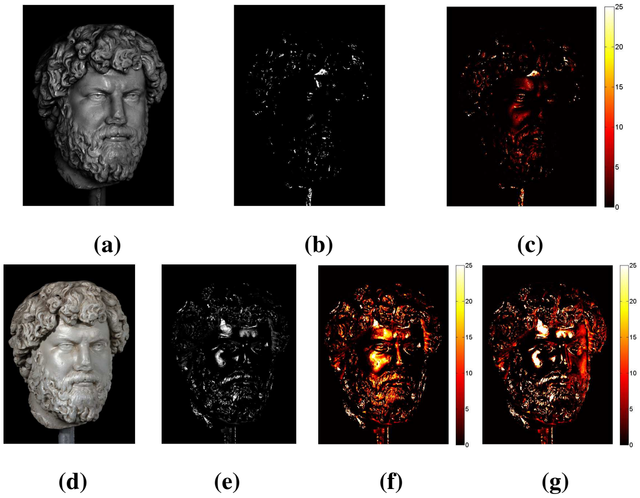

Figure 4: Greek Statue: (a) Single light source IR image. (b) Ground truth specular irradiance map for (a). (c) Specular irradiance estimation error map. This is the absolute difference map between our predicted specular irradiance and the ground truth. (d) Multiple light source natural lighting (NL) image. (e) Specular lighting ground truth of (d). (f,g) Specular irradiance error maps of (d) as estimated using first (SH1) and second (SH2) order spherical harmonics respectively. Note the reduced errors when using a single known light source (c) as opposed to estimating multiple unknown light sources using spherical harmonics lighting models (f,g).

5. Results

We preformed several tests in order to evaluate the quality and accuracy of the proposed algorithm. We show the algorithm's accuracy in recovering the specular lighting of the scene and why it is vital to use an IR image instead of an RGB image. In addition, we demonstrate that the proposed framework is state of the art, both visually and qualitatively.

In order to test the specular lighting framework, we took 3D objects from the Stanford $3D$[1], $123D$ Gallery[2] and Blendswap[3] repositories. For each model we assigned a mix of diffuse and specular shaders and rendered them under an IR lighting scenario described in Section 4.1 (single light source) and natural lighting scenarios (multiple light sources) using the Cycles renderer in Blender. To get a ground truth specularity map for each lighting scenario, we also captured each model without its specular shaders and subtracted the resulting images.

We tested the accuracy of our model in recovering specularities for each lighting setup. We used Eqs. (2) and (5) to

[1] http://graphics.stanford.edu/data/3Dscanrep/

[2] http://www.123dapp.com/Gallery/content/all

[3] http://www.blendswap.com/

Model	Median Error (mm)			90th % (mm)		
	Wu *et al.*	Or-El *et al.*	Proposed	Wu *et al.*	Or-El *et al.*	Proposed
Armadillo	0.335	0.318	**0.294**	1.005	0.821	**0.655**
Dragon	0.337	0.344	**0.324**	0.971	0.917	**0.870**
Greek Statue	0.306	0.281	**0.265**	0.988	0.806	**0.737**
Stone Lion	0.375	0.376	**0.355**	**0.874**	0.966	0.949
Cheeseburger	0.191	0.186	**0.168**	0.894	**0.756**	0.783
Pumpkin	0.299	0.272	**0.242**	0.942	0.700	**0.671**

Table 2: Quantitative comparison of depth accuracy in specular areas. All values are in millimeters.

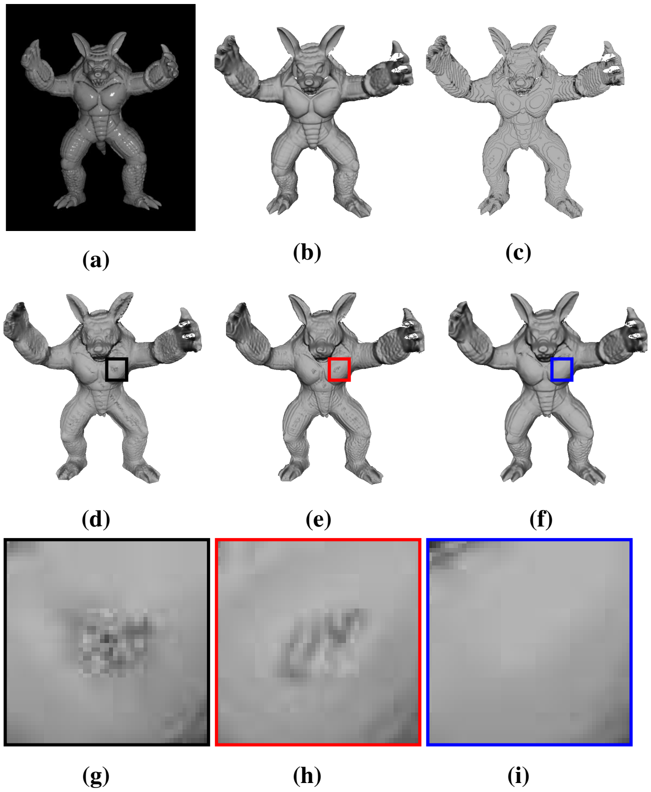

Figure 5: Results for the simulated Armadillo scene, (a) Input IR image. (b) Ground truth model. (c) Initial Depth. (d)-(f) Reconstructions of Wu *et al.*, Or - El *et al.* and our proposed method respectively. (g)-(i) Magnifications of a specular area. Note how our surface is free from distortions in specular areas unlike the other methods.

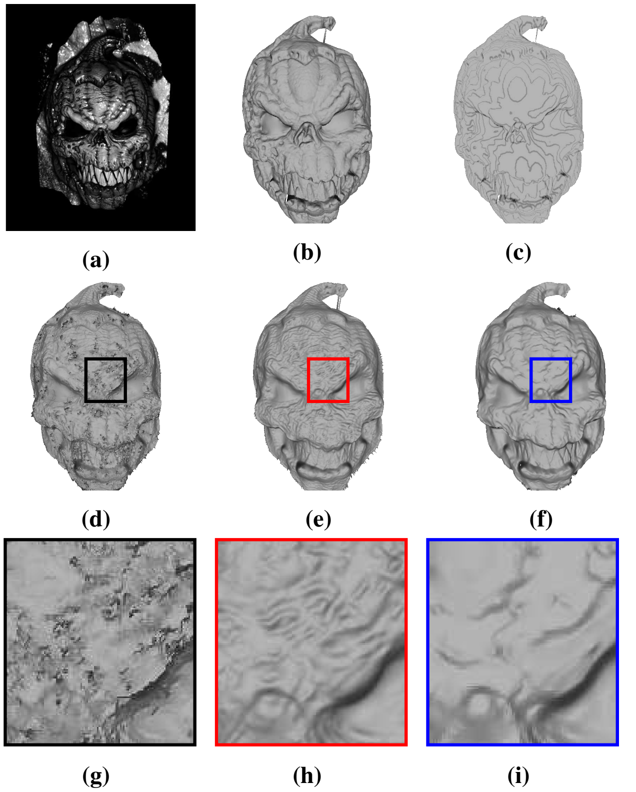

Figure 6: Results for the simulated Pumpkin scene, (a) Input IR image. (b) Ground truth model. (c) Initial Depth. (d)-(f) Reconstructions of Wu *et al.*, Or - El *et al.* and our proposed method respectively. (g)-(i) Magnifications of a specular area. Note the lack of hallucinated features in our method.

get the diffuse and ambient shading maps under IR lighting. For natural lighting, the diffuse and ambient shading were recovered using first and second order spherical harmonics in order to have two models for comparison. In both lighting scenarios the surface normals were calculated from the ground truth depth map. The specular lighting is recovered using Eqs. (3) and (7), where the IR lighting direction $\vec{l}_p$ is calculated using the camera-projector calibration parameters. In the natural lighting scene we use the relevant normalized coefficients of the first and second order spherical harmonics in order to compute the general lighting direction. From the results in Table 1 we can infer that the specular irradiance can be accurately estimated in our proposed lighting model as opposed to the natural lighting (NL SH1/2) where estimation errors are much larger. The reason for large differences is that, as opposed to our lighting model, under natural illumination there are usually multiple light sources that cause specularities whose directions cannot be recovered accurately. An example of this can be seen in Figure 4.

To measure the depth reconstruction accuracy of the proposed method we performed experiments using both synthetic and real data. In the first experiment, we used the

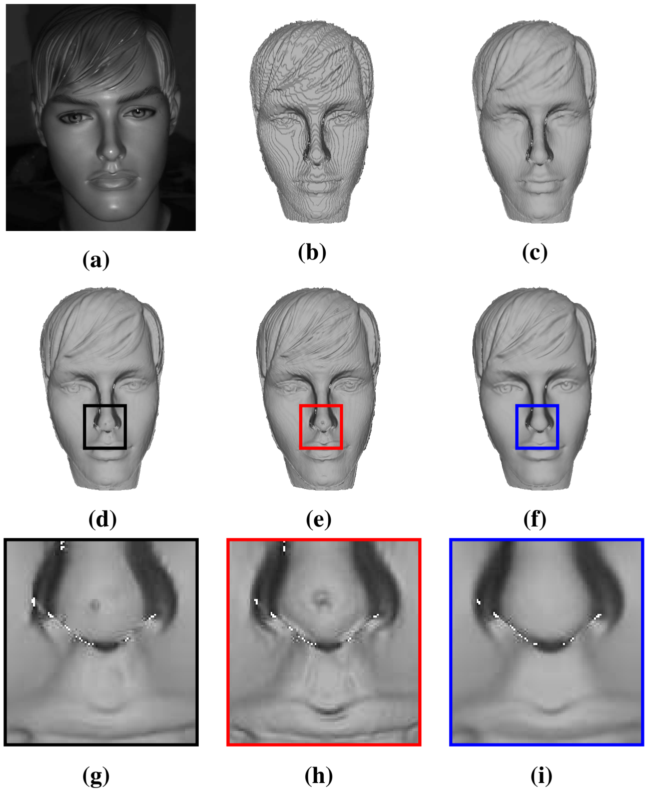

Figure 7: Results for the lab conditions experiment, (a) Input IR image. (b) Initial Depth. (c) Result after bilateral smoothing. (d)-(f) Reconstructions of Wu *et al*., Or - El *et al*. and the proposed method, respectively. (g)-(i) Magnifications of a specular region.

$3D$ models with mixed diffuse and specular shaders and rendered their IR image and ground truth depth maps in Blender. We then quantized the ground truth depth map to 1.5mm units in order to simulate the noise of a depth sensor. We applied our method to the data and defined the reconstruction error as the absolute difference between the result and the ground truth depth maps. We compared our method's performance with the methods proposed in [12, 22]. The comparisons were performed in the specular regions of the objects according to the ground truth specularity maps. The results are shown in Table. 2. A qualitative evaluation of the accuracy when the method is applied to the synthetic data can be seen in Figures. 5 and 6.

In the second experiment we tested our method under laboratory conditions using a structured-light $3D$ scanner to capture the depth of several objects. The camera-projector system was calibrated according to the method suggested in [25]. We reduced the number of projected patterns in order to obtain a noisy depth profile. To approximate an IR lighting scenario, we used a monochromatic projector and camera with dim ambient illumination.

We also tested the algorithm with an Intel Real-Sense depth scanner, using the IR image and depth map as inputs. The camera-projector calibration parameters were acquired from the Real-Sense SDK platform. Although no accurate ground-truth data was available for these experiments, we note that while all methods exhibit sufficient accuracy in diffuse areas, the proposed method is the only one that performs qualitatively well in highly specular areas as can be seen in Figures 7 and 8.

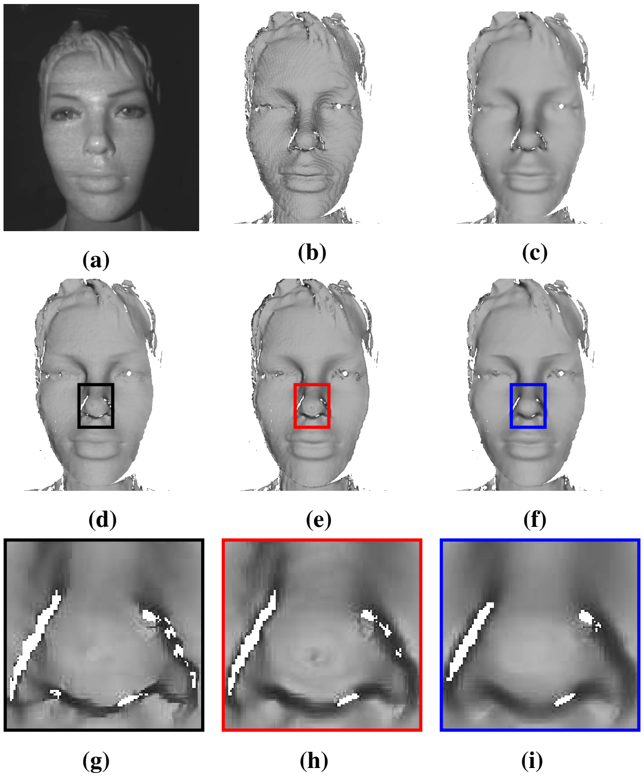

Figure 8: Results from Intel's Real-Sense depth scanner, (a) Input IR image. (b) Initial Depth. (c) Result after bilateral smoothing. (d)-(f) Reconstructions of Wu *et al*., Or - El *et al*. and the proposed method, respectively. (g)-(i) Magnifications of a specular region.

6. Conclusions

We presented a new framework for depth refinement of specular objects based on shading cues from an IR image. To the best of our knowledge, the proposed method is the first depth refinement framework to explicitly account for specular lighting. An efficient optimization scheme enables our system to produce state of the art results at real-time rates.

Acknowledgments

We wish to thank Alon Zvirin for his help with the experiments. This research was supported by European Communitys FP7-ERC program grant agreement no. 267414. G.R is partially funded by VITALITE Army Research Office Multidisciplinary Research Initiative program, award W911NF-11-1-0391.

References

[1] J. T. Barron and J. Malik. Shape, illumination, and reflectance from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1670–1687, 2015. 2

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

[3] 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

[4] A. Chatterjee and V. M. Govindu. Photometric refinement of depth maps for multi-albedo objects. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 933–941, 2015. 2

[5] G. Choe, J. Park, Y.-W. Tai, and I. So Kweon. Exploiting shading cues in kinect IR images for geometry refinement. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3922–3929, 2014. 2

[6] 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. 1

[7] 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 (ICCV)*, pages 1617–1624, 2013. 1, 2, 5

[8] S. Haque, A. Chatterjee, and V. M. Govindu. High quality photometric reconstruction using a depth camera. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2283–2290, 2014. 2

[9] A. Kadambi, V. Taamazyan, B. Shi, and R. Raskar. Polarized 3D: High-quality depth sensing with polarization cues. In *IEEE International Conference on Computer Vision*, pages 3370–3378, 2015. 2

[10] A. Kar, S. Tulsiani, J. Carreira, and J. Malik. Category-specific object reconstruction from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1966–1974. 2015. 2

[11] 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

[12] R. Or El, G. Rosman, A. Wetzler, R. Kimmel, and A. M. Bruckstein. RGBD-Fusion: Real-time high precision depth recovery. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5407–5416, 2015. 1, 2, 5, 8

[13] T. Ping-Sing and M. Shah. Shape from shading using linear approximation. *Image and Vision Computing*, 12(8):487–498, 1994. 6

[14] 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. 4

[15] S. R. Richter and S. Roth. Discriminative shape from shading in uncalibrated illumination. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1128–1136, 2015. 2

[16] G. Rosman, A. M. Bronstein, M. M. Bronstein, X.-C. Tai, and R. Kimmel. Group-valued regularization for analysis of articulated motion. In *NORDIA workshop, European Conference on Computer Vision (ECCV)*, pages 52–62. Springer, 2012. 5

[17] A. Roussos and P. Maragos. Tensor-based image diffusions derived from generalizations of the total variation and beltrami functionals. In *IEEE International Conference on Image Processing (ICIP)*, pages 4141–4144. IEEE, 2010. 5

[18] N. Sochen, R. Kimmel, and R. Malladi. A general framework for low level vision. *IEEE Transactions on Image Processing*, 7(3):310–318, 1998. 5

[19] C. Ti, R. Yang, J. Davis, and Z. Pan. Simultaneous time-of-flight sensing and photometric stereo with a single tof sensor. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4334–4342, 2015. 2

[20] A. Wetzler and R. Kimmel. Efficient beltrami flow in patch-space. In *Scale Space and Variational Methods in Computer Vision (SSVM)*, pages 134–143, 2011. 5

[21] C. Wu and X.-C. Tai. Augmented lagrangian method, dual methods, and split bregman iteration for ROF, vectorial TV, and high order models. *SIAM J. Img. Sci.*, 3:300–339, July 2010. 4, 6

[22] C. Wu, M. Zollhöfer, M. Nießner, 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. 1, 2, 5, 8

[23] 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 (CVPR)*, 2013. 2

[24] 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. 1

[25] S. Zhang and P. S. Huang. Novel method for structured light system calibration. *Optical Engineering*, 45(8):083601–1–083601–8, 2006. 8