Photometric Refinement of Depth Maps for Multi-albedo Objects

Avishek Chatterjee   Venu Madhav Govindu
Indian Institute of Science
Bengaluru 560012, India
avishek|venu@ee.iisc.ernet.in

Abstract

In this paper, we propose a novel uncalibrated photometric method for refining depth maps of multi-albedo objects obtained from consumer depth cameras like Kinect. Existing uncalibrated photometric methods either assume that the object has constant albedo or rely on segmenting images into constant albedo regions. The method of this paper does not require the constant albedo assumption and we believe it is the first work of its kind to handle objects with arbitrary albedo under uncalibrated illumination. We first robustly estimate a rank 3 approximation of the observed brightness matrix using an iterative reweighting method. Subsequently, we factorize this rank reduced brightness matrix into the corresponding lighting, albedo and surface normal components. The proposed factorization is shown to be convergent. We experimentally demonstrate the value of our approach by presenting highly accurate three-dimensional reconstructions of a wide variety of objects. Additionally, since any photometric method requires a radiometric calibration of the camera used, we also present a direct radiometric calibration technique for the infra-red camera of the structured-light stereo depth scanner. Unlike existing methods, this calibration technique does not depend on a known calibration object or on the properties of the scene illumination used.

1. Introduction

Depth maps from structured-light stereo depth cameras like Kinect suffer from very high noise and poor resolution. As a result, fine-scale (high frequency) structural details of an object are not captured by such depth cameras. The requisite high frequency details can be obtained through methods that make radiometric measurements such as shape-from-shading [14, 25, 7] or photometric stereo [21], which can then be fused with the low quality depth maps to obtain high quality shape estimates. Since the literature on shape-from-shading and photometric stereo is enormous, in the following we restrict our discussion of the relevant literature to our specific context of fusing depth and radiometric estimates of normals [23, 9, 18, 10, 5]. These methods can broadly be classified into two categories: shape-from-shading approaches that use only one intensity image [23, 9, 5] and photometric stereo methods that rely on observations under varying lighting conditions [18, 10]. While acquiring a single image is simpler and faster, shape-from-shading suffers from intrinsic ambiguities including multiple solutions even for known lighting [3, 8] and is generally restricted to objects of constant albedo. Often it is difficult to distinguish between changes in albedo from changes in the surface normal using only a single image. While the methods in [23, 5] segment observed images into regions of constant albedo, this is very restrictive since many objects have continuously varying albedo or have small regions with a different albedo that makes the segmentation task difficult or impossible.

By using multiple images acquired under different lighting conditions, photometric stereo can overcome the ambiguities suffered by shape-from-shading. In this paper, we propose a novel photometric method for recovering the surface normals of multi-albedo objects that are then fused with the depth estimates from Kinect to result in a depth refinement method that accurately recovers the 3D object shape. In our work, we use the infra-red (henceforth IR) camera of Kinect for radiometric measurements as was done in [5, 10]. While [18, 10] also use the principles of photometric stereo for normal estimation, their methods make the restrictive assumption that the objects have a single albedo. While the normal estimation step of [10] is independent of albedo, their lighting estimation step assumes constant albedo. Thus, for multi-albedo objects the normal estimates and final refined depth estimates are poor. [18] uses the uncalibrated photometric stereo method proposed in [11]. While solving the linear ambiguity in the factorization of observed radiometric brightness, [18] uses a constant albedo assumption. In contrast, in our approach detailed in Section 3.2 we make no such restrictive assumption. While solving for the linear ambiguity of factorization of bright-
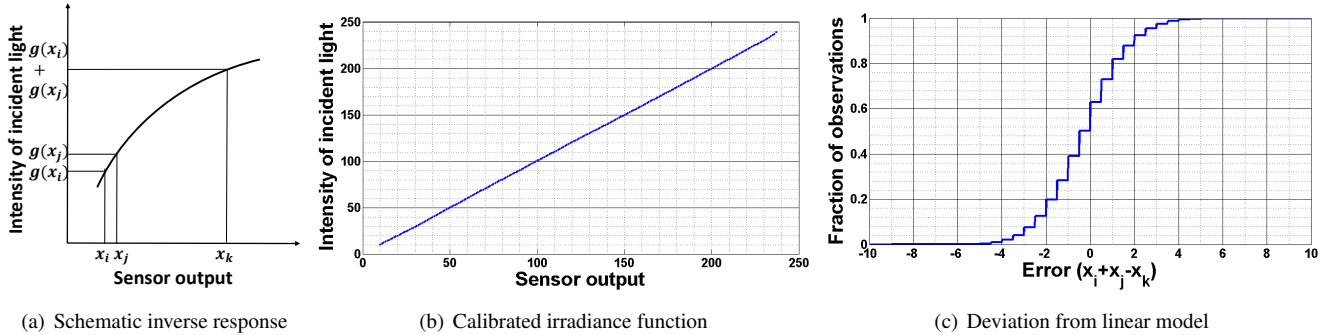
Figure 1. Radiometric calibration of IR camera. See the text for more details.

ness matrix, instead of assuming constant albedo, we minimize a novel cost which is a function of the normals estimated from the raw Kinect depth map and those obtained using photometric stereo. This minimization is carried out iteratively. We analytically prove in Section 3.5 that our iterative method is convergent in the noise free case and demonstrate that it is convergent for real data under a wide range of conditions. Finally, as mentioned in Section 4, we use a fusion technique that combines depth and normal estimates to create high quality 3d shape representations of objects.

2. Radiometric Calibration of Kinect's IR Camera

Radiometric calibration of the IR camera is an important prerequisite for any photometric estimation that relies on radiometric measurements. In [5], radiometric calibration is done by comparing predicted and observed brightness of a Lambertian sphere under controlled illumination. This approach demands that a perfectly Lambertian surface with known geometry be used in combination with a well calibrated illumination source. To overcome the limitations of such an approach, we propose a radiometric calibration method which does not depend on knowledge of either the reflectance of the observed surface or the illumination source. In the schematic function shown in Figure 1(a), the horizontal axis represents the measured brightness (i.e. pixel intensity measured in the IR camera) and the vertical axis represents the actual amount of light received by the IR camera, i.e. $g(\cdot)$ is the inverse response function of the camera, see entry 'Radiometric Calibration' in [15]. Thus, recovering $g(\cdot)$ implies that for any measured pixel intensity x we can directly recover the true irradiance $g(x)$.

In our calibration exercise, we seek to recover $g(\cdot)$ as a non-parametric function. Let the measured brightness or pixel intensities lie in the range $[x_1, \cdots, x_n]$, typically $[0, \cdots, 255]$ in steps of 1. Thus by radiomet-

ric calibration we seek to recover the unknown values $\{g(x_1), \cdots, g(x_n)\}$. We illuminate an arbitrary scene with two light sources. Consider a pixel with observed intensity equal to x_i and x_j for the first and the second light sources respectively and equal to x_k for both the sources of light turned on simultaneously. Therefore, since the irradiance when both lights are on is the sum of the irradiances of each individual light sources, we have

$$g(x_k) = g(x_i) + g(x_j) \qquad (1)$$

Thus, Equation 1 represents a linear relationship between the three unknowns among the n unknowns of the nonparametric function $g(x)$. We generate enough such constraints by imaging scenes with substantial variations of reflectance and solve the resulting homogeneous set of equations in a least squares sense. While there are other radiometric calibration methods such as [6] based on varying exposure times, we chose to use the method described above as it is difficult to control the exposure time for Kinect's IR camera. In our experiments, to reduce the effect of sensor noise, images are averaged over 10 consecutive observations under the same illumination and intensities near the saturation ranges are discarded. In Figure 1(b) we see that the radiometric calibration of Kinect's IR camera is linear, implying that like Equation 1, $x_k = x_i + x_j$ should also hold. In Figure 1(c) we depict the fraction of observations vs. error $x_i + x_j - x_k$ and note that most of the observations have a deviation within the value of ± 4 for a dynamic range of 256 values.

3. Lighting, Normal and Albedo Estimation

In our method, we assume that the objects are Lambertian.[1] We also assume that the IR light sources are directional in nature, which is ensured by placing the sources far

---

[1]Although, real-world Lambertian surfaces are rare, we show in our experiments that our method works well with many non-metallic materials which also have a substantial amount of specular reflectance.

enough from the objects being scanned. Let the lighting intensity and direction corresponding to the i-th image be denoted by λ_i and the 3×1 unit norm vector $\mathbf{l}_i$ respectively. Let the albedo and the surface normal of the point observed at pixel p be denoted by α_p and the 3×1 unit norm vector $\mathbf{n}_p$ respectively. For a Lambertian surface, the brightness (B_{ip}) at the pixel p in the i-th image is given by

$$B_{ip} = \lambda_i \mathbf{l}_i{}^T \mathbf{n}_p \alpha_p = \mathbf{L}_i{}^T \mathbf{N}_p \qquad (2)$$

where $\mathbf{L}_i = \lambda_i \mathbf{l}_i$ is the intensity scaled lighting direction corresponding to the image i and $\mathbf{N}_p = \alpha_p \mathbf{n}_p$ is the albedo scaled normal vector of the point viewed at pixel p [21, 14]. [1] finds this linear model of brightness inadequate due to reflection of light from various directions in an indoor scene and proposes a brightness model based on second order spherical harmonics which is used in [23]. Nevertheless, the linear model of brightness is found accurate enough for depth-guided photometry in [18, 10, 5].

If we have k lighting conditions and m observed pixels, then define $\mathbb{L}(3 \times k) = \begin{bmatrix} \mathbf{L}_1 & \cdots & \mathbf{L}_k \end{bmatrix}$. and $\mathbb{N}(3 \times m) = \begin{bmatrix} \mathbf{N}_1 & \cdots & \mathbf{N}_m \end{bmatrix}$. We also define a $k \times m$ brightness matrix $\mathbb{B}$ such that $\mathbb{B}(i, p)$ is the brightness of pixel p in image i. Thus the brightness constraints for all the pixels in all images can be summarized as

$$\mathbb{B} = \mathbb{L}^T \mathbb{N} \qquad (3)$$

Our problem is to factorize an observed brightness matrix $\mathbb{B}$ into $\mathbb{L}$ and $\mathbb{N}$. The rank of $\mathbb{B}$ should not be more than 3 since both $\mathbb{L}$ and $\mathbb{N}$ are of atmost rank 3. But noise in images, outliers like shadows, specularities, inter-reflections etc. cause $\mathbb{B}$ to have rank more than 3. In [11], $\mathbb{B}$ is approximated with its closest rank 3 matrix in a least square sense using SVD. But since the least square approximation is not robust to outliers we propose a robust rank reduction method.

3.1. Rank 3 approximation of Brightness Matrix

The observed brightness matrix $\mathbb{B}$ can be modeled as a combination of a rank 3 matrix $\mathbb{B}^*$, a noise matrix $\mathbf{e}$ and a sparse outlier matrix $\mathbf{E}$, i.e.

$$\mathbb{B} = \mathbb{B}^* + \mathbf{e} + \mathbf{E} \qquad (4)$$

[22] ignores the noise term $\mathbf{e}$ and minimizes a weighted sum of the nuclear norm of $\mathbb{B}^*$ and the ℓ_1 norm of $\mathbf{E}$ using the Augmented Lagrangian method (ALM) [16]. Alternatively, this problem can be solved using GoDEC [26], which is able to take the noise component into account. In our experiments, we have found that the approximation matrix generated by ALM or GoDEC often has a rank

less than 3. This occurs since in many practical scenarios, $\mathbb{N}$ does not span all directions in $\mathbb{R}^3$ equally. It is often the case that most of the points on the observed part of a surface have their normals pointing towards the camera. Similarly, $\mathbb{L}$ also fails to span $\mathbb{R}^3$ well since light sources are typically placed close to the Z axis to avoid shadows due to self-occlusion. Therefore, while eliminating outliers from $\mathbb{B}$, despite their expensive computations involved, ALM or GoDEC often remove useful information and output a matrix with rank less than 3, which cannot be used for further processing. We now describe our method that mitigates this problem and provides a robust rank 3 estimate $\mathbb{B}^*$.

Given an observation matrix $\mathbb{B}$ ($k \times m$ and $m \gg k$), our task is to find 3 dimensional orthonormal basis $\mathbf{V} \in \mathbb{R}^{k \times 3}$ such that the columns of $\mathbb{B}$ mostly lie in the subspace spanned by $\mathbf{V}$. Let $\mathbb{B}_p$ be the p-th column of $\mathbb{B}$. We define the error e_p as the magnitude of the component of $\mathbb{B}_p$ that is orthogonal to the subspace spanned by $\mathbf{V}$ i.e.

$$e_p = \|\mathbb{B}_p - \mathbf{V}\mathbf{V}^T\mathbb{B}_p\| \qquad (5)$$

Then we minimize $\sum_{p=1}^{m} \rho(e_p)$ to find the optimal $\mathbf{V}$. If we choose $\rho(e) = e^2$ in our cost function, the solution is equivalent to the least squares method which is not robust to outliers. If instead we use the absolute norm $\rho(e) = |e|$, then $\rho(e)$ is not differentiable at $e = 0$. Therefore, we use the Huber cost function defined as:

$$\rho(e) = \begin{cases} e^2 & \forall |e| \leq \alpha \\ 2\alpha|e| - \alpha^2 & \forall |e| > \alpha \end{cases} \qquad (6)$$

which ensures that $\rho(e)$ is C^1 continuous as well as robust and convex. To optimize this function, we use the iteratively reweighted least squares (IRLS) method [12, 19] which results in a simple method that is efficient and more importantly robust to outliers. The gradient of this robust function is given by

$$\begin{aligned} \nabla(\rho(e)) &= \begin{cases} \nabla(e^2) & \forall |e| \leq \alpha \\ \nabla\left(2\alpha\sqrt{e^2} - \alpha^2\right) & \forall |e| > \alpha \end{cases} \\ &= \begin{cases} \nabla(e^2) & \forall |e| \leq \alpha \\ \frac{\alpha}{\sqrt{e^2}}\nabla(e^2) & \forall |e| > \alpha \end{cases} \\ &= w^2\nabla(e^2) \begin{cases} w = 1 & \forall |e| \leq \alpha \\ w = \sqrt{\frac{\alpha}{|e|}} & \forall |e| > \alpha \end{cases} \end{aligned} \qquad (7)$$

We can equate the gradient of our cost function to zero for the optimal solution as

$$\nabla\left(\sum_{p=1}^{m} \rho(e_p)\right) = \sum_{p=1}^{m} w_p^2 \nabla(e_p{}^2) = 0 \qquad (8)$$
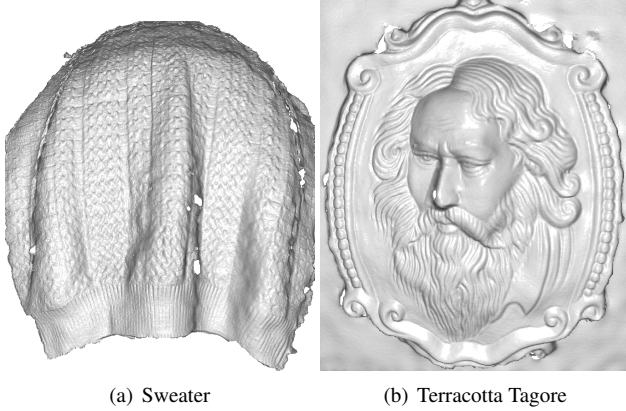
(a) Sweater   (b) Terracotta Tagore

Figure 2. Our estimated 3d depth reconstructions of some constant albedo objects.

For a given w, solving Equation 8 is equivalent to minimizing $\sum_{p=1}^{m} (w_p e_p)^2$ which is a weighted least squares problem.

We iteratively compute – until convergence – weights w_p based on the errors e_p $\forall p$ and then keeping w_p fixed, we solve the weighted least squares problem – which in this case is the problem of principal component analysis – as illustrated in Algorithm 1.

Algorithm 1 Rank 3 Approximation of $\mathbb{B}$ using Robust Principal Component Analysis.

Input: $\mathbb{B}$: Observed Brightness Matrix ($k \times m$)
Outputs: $\mathbb{B}^*$ ($k \times m$), $\mathbf{V}$ ($k \times 3$), $\mathbb{X}$ ($3 \times m$), where $\mathbb{B}^* = \mathbf{V}\mathbb{X}$ is the rank 3 approximation of $\mathbb{B}$.
Initialize: $w_p = 1$ $\forall p = 1, 2, \cdots, m$

    1. $\mathbf{V}$ = Leading 3 principal components of $\mathbb{B}\mathbf{W}$, where $\mathbf{W}$ is a diagonal weighting matrix such that $\mathbf{W}(p,p) = w_p$
    2. $\mathbf{X} = \mathbf{V}^T\mathbb{B}$ and $\mathbb{B}^* = \mathbf{V}\mathbb{X}$
    4. e_p=p-th column of $(\mathbb{B} - \mathbb{B}^*)$ $\forall p = 1, 2, \cdots, m$
    5. Compute w_p using Equation 7 $\forall p = 1, 2, \cdots, m$
    6. If not converged, Goto State 1

We note here that all the columns of brightness matrix $\mathbb{B}$ are first normalized before being used as input in Algorithm 1. This ensures that the value of e_p is bounded below 2 which both ensures robustness and also prevents pixels with higher albedo from contributing more to the error and thereby dominating the solution.

3.2. Factorization of Brightness Matrix

From Algorithm 1 we have

$$\mathbb{B} \approx \mathbb{B}^* = \mathbf{V}\mathbb{X} \tag{9}$$

but this factorization is not unique. A general factorization of $\mathbb{B}^*$ can be written as

$$\mathbb{B}^* = \mathbb{L}^T\mathbb{N} = \mathbf{V}\mathbf{A}\mathbf{A}^{-1}\mathbb{X} \tag{10}$$

where $\mathbf{A}$ can be any non-singular 3×3 matrix. Therefore, there exists an $\mathbf{A}$ such that

$$\mathbb{L} = \mathbf{A}^T\mathbf{V}^T \tag{11}$$
$$\mathbb{N} = \mathbf{A}^{-1}\mathbb{X} \tag{12}$$

Hence our problem is one of estimating the requisite matrix $\mathbf{A}$ which allows us to recover $\mathbb{N}$. In [11], $\mathbf{A}$ is solved by using Equation 12 by asking the user to pick 6 pixels with the same albedo or 6 images with the same lighting intensity. In contrast, we solve for $\mathbf{A}$ using the raw depth map from Kinect as a guide. We generate a rough estimate of the surface normals $\widehat{\mathbf{n}}_p$ at every pixel from the raw depth map and collect them in a $3 \times m$ matrix, as $\widehat{\mathbf{N}} = \left[\begin{array}{ccc} \widehat{\mathbf{n}}_1 & \cdots & \widehat{\mathbf{n}}_m \end{array} \right]$.

While the surface normals ($\mathbb{N}$) are linear transformation of $\mathbb{X}$, (see Equation 12), all linear transformations of $\mathbb{X}$ are not valid surface normals. This is because of the requirement that any valid set of surface normals should be integrable or curl free [13]. However, even the imposition of this constraint does not yield a unique linear transformation $\mathbf{A}$. Rather, for every $\mathbf{A}$ that gives an integrable solution, there is a three parameter family of transformations of $\mathbf{A}$ that also satisfies the integrability constraint. This is the well-known bas-relief ambiguity [2, 4].

In this paper, we ignore the integrability constraint while solving for $\mathbf{A}$. We note that since $\mathbf{A}^{-1}\mathbb{X}$ should be aligned with the normals estimated from Kinect's depth map, we have a sufficient number of constraints to solve for $\mathbf{A}$, although such an $\mathbf{A}$ does not necessarily satisfy the integrability constraint. However, since in our approach we solve for a surface that carries out a fusion of both Kinect's depth estimate as well as the normal map, we need not impose the integrability constraint during the estimation of the normals.

Before we develop our solution for multi-albedo factorization, we illustrate the use of $\widehat{\mathbf{N}}$ to solve the simpler problem of constant albedo that was addressed in [18, 10].

3.3. Factorization for Constant Albedo Surfaces

For a surface with constant albedo, we can assign the relative albedo $\alpha_p = 1$, $\forall p$. This allows us to replace $\mathbb{N}$

4

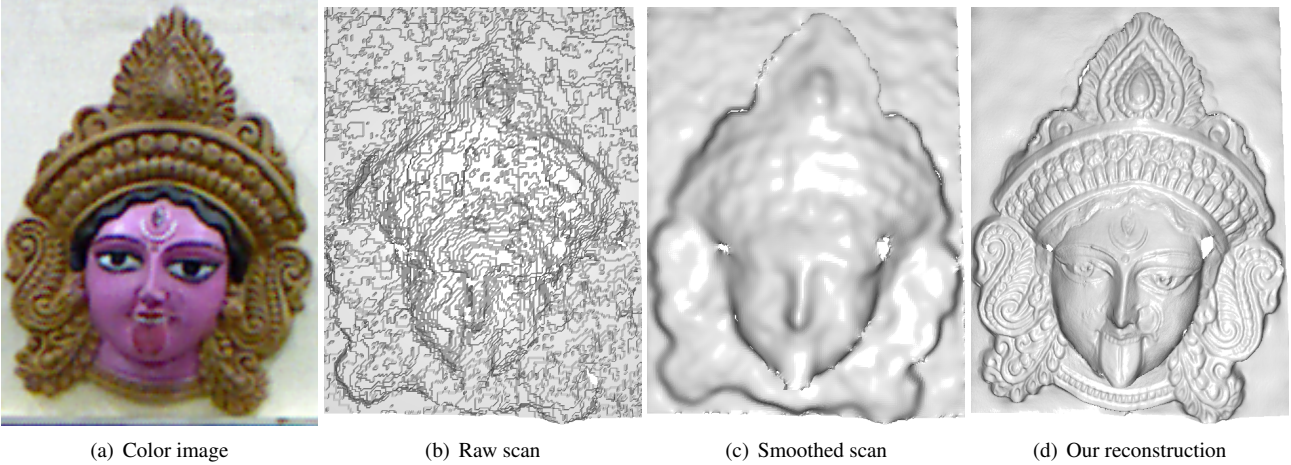| (a) Color image | (b) Raw scan | (c) Smoothed scan | (d) Our reconstruction |

Figure 3. Refined depth map estimates of a painted terracotta plaque of Kali using our approach.

in Equation 12 with Kinect's estimated normal matrix $\widehat{\mathbf{N}}$. Thus, we need to solve for $\mathbf{A}$ using Equation 12. In [18] $\mathbf{A}$ was solved in a least squares sense as

$$\mathbf{A} = \left(\left(\widehat{\mathbf{N}}^T \right)^{\dagger} \widehat{\mathbb{X}}^T \right)^T \qquad (13)$$

where $\dagger$ denotes the pseudo inverse operation. But the estimates of $\widehat{\mathbf{n}}_p$ near high curvature regions including depth edges of an object are not reliable since Kinect depth maps are highly noisy. Thus, a least squares solution is affected by the poor or incorrect estimates of $\widehat{\mathbf{n}}_p$ near depth edges. To solve for $\mathbf{A}$ robustly, we define $e_p = \|p\text{-th column of}(\mathbf{A}\widehat{\mathbf{N}} - \mathbb{X})\|$ and minimize $\sum_{p=1}^{m} \rho(\mathbf{e}_p)$, where $\rho(\cdot)$ is defined in Equation 6. This minimization is performed using the iteratively reweighted least square method detailed in Algorithm 2.

Algorithm 2 Robust cost minimization using IRLS

Initialize: $w_p = 1 \;\; \forall p$

   1. $\mathbf{A} = \left(\left(\mathbf{W}\widehat{\mathbf{N}}^T \right)^{\dagger} \left(\mathbf{W}\mathbb{X}^T \right) \right)^T$, where $\mathbf{W}$ is a diagonal weighting matrix such that $\mathbf{W}(p,p) = w_p$

   2. Compute errors e_p and weights w_p for all pixels p

   3. If not converged, Goto State 1

In Figure 2 we show our estimated 3d reconstructions of (a) a wollen sweater and (b) a terracotta plaque depicting the Indian poet Rabindranath Tagore. As can be seen, our method very accurately recovers high frequency shape details .

3.4. Factorization for Multi-Albedo Surfaces

We can now proceed to solve the more general problem of recovering $\mathbf{A}$ for objects with multiple or arbitrarily varying albedos. Let $\mathbf{x}_p$ denote the normalized p-th column of $\mathbb{X}$, i.e. corresponding to the p-th pixel. Then we have

$$\mathbf{A}\alpha_p\widehat{\mathbf{n}}_p \parallel \mathbf{x}_p \;\;\Rightarrow\;\; \mathbf{A}\widehat{\mathbf{n}}_p \parallel \mathbf{x}_p \;\; \forall p \qquad (14)$$

where α_p is the albedo for the p-th pixel and the symbol $\parallel$ denotes that the vectors $\mathbf{A}\alpha\widehat{\mathbf{n}}_p$ and $\mathbf{x}_p$ are parallel for all pixels p. We define the error e_p to be a measure of the angle between the vectors $\mathbf{A}\widehat{\mathbf{n}}_p$ and $\mathbf{x}_p$ as

$$e_p = \left\| \mathbf{x}_p - \frac{A\widehat{\mathbf{n}}_p}{\|A\widehat{\mathbf{n}}_p\|} \right\| \qquad (15)$$

The optimal value of $\mathbf{A}$ can now be robustly estimated as

$$\mathbf{A}^* = \underset{\mathbf{A}}{\operatorname{argmin}} \sum_{p=1}^{m} \rho(e_p) \qquad (16)$$

The optimization problem of Equation 16 is solved using gradient descent. Once we find the optimal value of $\mathbf{A}$, the surface normals can be recovered for every pixel.

In Figure 3 and 4 we show the depth estimates obtained using this approach for two multi-albedo objects: a painted terracotta plaque of the Hindu goddess Kali (Figure 3) and a plastic figure of the Hindu deity Ganesh (Figure 4). Both these models are quite specular in nature. As shown in Figure 3(b) and 4(a) , the raw Kinect depth maps are highly noisy and the resultant refinements obtained using our method (Figure 3(d) and 4(b)) is quite substantial.
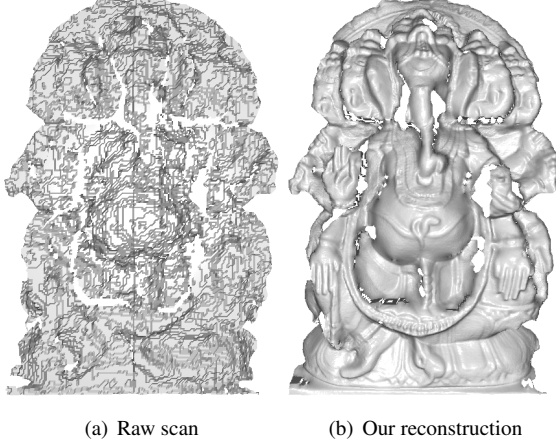
(a) Raw scan   (b) Our reconstruction

Figure 4. Refined depth map estimates of Ganesh using our approach.
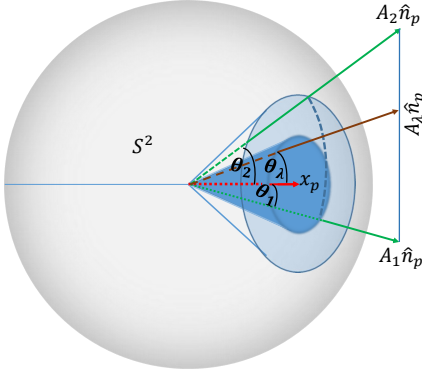

Figure 5. Convergence properties of our multi-albedo factorization method. Please see text for details.

3.5. Convergence of Multi-Albedo Factorization

In this subsection we present a proof for the convergence of the gradient descent method of Section 3.4 that we use to solve for **A** in the multi-albedo scenario. Specifically, we prove that the cost function in Equation 16 is monotonically non-decreasing away from the global minima in the noise free scenario. Therefore, as long as the corruption of the observed brightness matrix is not too large, our gradient descent method is guaranteed to converge to the optimal solution.

Lemma 3.1 *Let* $\mathbf{A}_\lambda$ *be a convex combination of two possible values of* **A** *denoted by* $\mathbf{A}_1$ *and* $\mathbf{A}_2$ *respectively, i.e.* $\mathbf{A}_\lambda = (1 - \lambda)\mathbf{A}_1 + \lambda\mathbf{A}_2$ *with* $0 \leq \lambda \leq 1$. *Then,*

$$\rho\left(e_p\left(\mathbf{A}_\lambda\right)\right) \leq \max\left(\rho\left(e_p\left(\mathbf{A}_1\right)\right), \rho\left(e_p\left(\mathbf{A}_2\right)\right)\right) \quad (17)$$

In Figure 5 the vectors $\mathbf{x}_p$, $\mathbf{A}_1\widehat{\mathbf{n}}_p$, $\mathbf{A}_2\widehat{\mathbf{n}}_p$ and $\mathbf{A}_\lambda\widehat{\mathbf{n}}_p$ are drawn from the center of the unit sphere S^2. In addition,

we also show the cones of revolution of $\mathbf{A}_1\widehat{\mathbf{n}}_p$ and $\mathbf{A}_2\widehat{\mathbf{n}}_p$ around $\mathbf{x}_p$. Suppose, $\mathbf{A}_1\widehat{\mathbf{n}}_p$, $\mathbf{A}_2\widehat{\mathbf{n}}_p$ and $\mathbf{A}_\lambda\widehat{\mathbf{n}}_p$ are at angles θ_1, θ_2 and θ_λ respectively from $\mathbf{x}_p$. Without loss of generality, let $\theta_2 > \theta_1$.

Proof As λ increases from 0 to 1, the point $\mathbf{A}_\lambda\widehat{\mathbf{n}}_p$ moves in a straight line in $\mathbb{R}^3$ from $\mathbf{A}_1\widehat{\mathbf{n}}_p$ to $\mathbf{A}_2\widehat{\mathbf{n}}_p$. Correspondingly, the projection of $\mathbf{A}_\lambda\widehat{\mathbf{n}}_p$ onto S^2 i.e $\frac{\mathbf{A}_\lambda\widehat{\mathbf{n}}_p}{\|\mathbf{A}_\lambda\widehat{\mathbf{n}}_p\|}$ moves along a geodesic arc on S^2 shown as a dashed blue line. Since $\mathbf{A}_1\widehat{\mathbf{n}}_p$ and $\mathbf{A}_2\widehat{\mathbf{n}}_p$ both lie inside the convex cone of revolution generated by $\mathbf{A}_2\widehat{\mathbf{n}}_p$ around $\mathbf{x}_p$, any convex combination of $\mathbf{A}_1\widehat{\mathbf{n}}_p$ and $\mathbf{A}_2\widehat{\mathbf{n}}_p$ i.e. $(1 - \lambda)\mathbf{A}_1\widehat{\mathbf{n}}_p + \lambda\mathbf{A}_2\widehat{\mathbf{n}}_p$ must also lie inside the cone of revolution of $\mathbf{A}_2\widehat{\mathbf{n}}$ around $\mathbf{x}_p$ $\forall\lambda \in [0, 1]$. Thus, $\theta_\lambda \leq \theta_2 = \max(\theta_1, \theta_2)$. Now since $\rho(e_p(\mathbf{A}_\lambda))$ is a monotonically increasing function of θ_λ, we can see that $\rho(e_p(\mathbf{A}_\lambda)) \leq \rho(e_p(\mathbf{A}_2)) = \max(\rho(e_p(\mathbf{A}_1)), \rho(e_p(\mathbf{A}_2)))$. In turn this implies that Equation 17 always holds true. ∎

Theorem 3.2 $\sum\limits_{p=1}^{m} \rho\left(e_p\right)$ *is monotonically non-decreasing away from the global minima in the noise free case.*

Proof Summing Equation 17 over all pixels we get,

$$\sum_p \rho\left(e_p\left(\mathbf{A}_\lambda\mathbf{x}_p\right)\right) \leq \sum_p \max\left(\rho\left(e_p\left(\mathbf{A}_1\right)\right), \rho\left(e_p\left(\mathbf{A}_2\right)\right)\right)$$
$$(18)$$

If we consider the noise free case and if $\mathbf{A}_1$ is the true solution, i.e. $\rho\left(e_p\left(\mathbf{A}_1\right)\right) = 0$ $\forall p$, then

$$\sum_p \rho\left(e_p\left(\mathbf{A}_\lambda\right)\right) \leq \sum_p \rho\left(e_p\left(\mathbf{A}_2\right)\right) \quad (19)$$

Therefore, our cost is monotonically non-decreasing from the true solution $\mathbf{A}_1$. ∎

We note that [20] uses a similar approach to solve for global translation estimation in the context of structure-from-motion.

4. Fusion of Depth and Normal Estimates

Given the depth map of Kinect and estimated surface normals, we need a principled method to fuse them to obtain a high quality reconstruction of the imaged surface. In our work, we use a depth-normal fusion method that was originally proposed in [17] and extended to Kinect's context in [24]. This approach was also used in [9, 18, 10] and the reader may refer to [17, 24, 10] for details as we cannot provide them here due to space constraints.
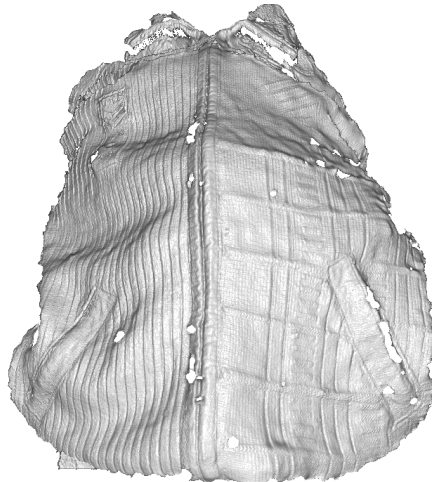
(a) Woolen Jacket
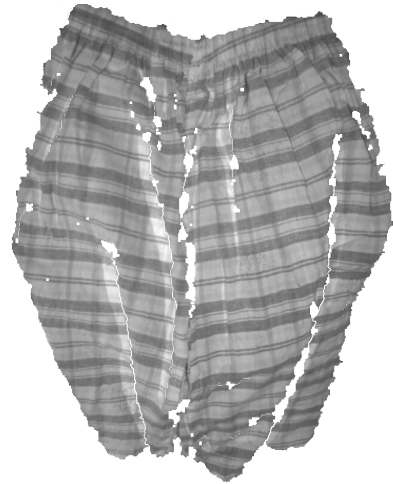
(b) Cotton Shorts
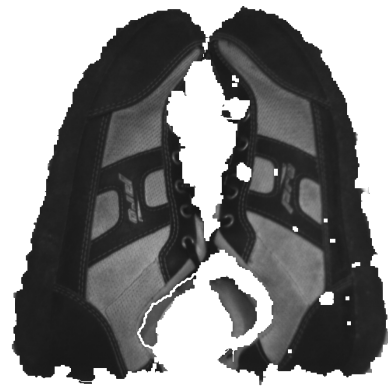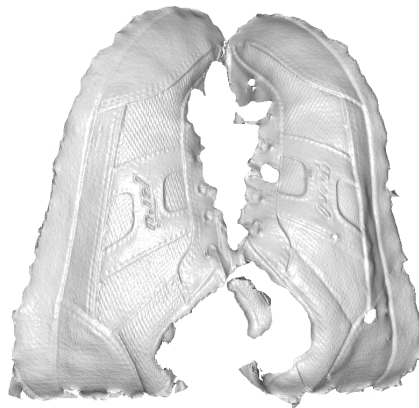
(c) Shoe

Figure 6. Reconstruction of some complex multi-albedo objects. Left: IR image, Middle: our 3d reconstruction and Right: our estimated relative albedo.

(a) Wooden Plate

Figure 7. Reconstruct of some objects. Left: an infrared image, Middle: our 3d reconstruction and Right: our estimate of relative albedo.

5. Results

In addition to the results for constant and multi-albedo objects shown in Figures 2, 3 and 4 respectively, in this section we present additional results for complex multi-albedo objects. In Figures 6 we show our results for a woolen jacket, a pair of cotton shorts and a pair of shoes. From the IR image of the jacket it is seen that the albedo variations are highly correlated with depth edges, e.g. the vertical depth stripes of the jacket. Nevertheless, we are able accurately recover both the 3d structure and the relative albedo. In the example of the striped shorts (middle row), we can observe that the striped albedo variation is perfectly recovered and the presence of this varying albedo pattern is not misinterpreted as shape variation. In addition, the high frequency details for the elastic waistband of the shorts are also correctly recovered. As can be seen in the bottom row of Figure 6, despite the large amount of variation in albedo for the pair of shoes, we are able to accurately recover their 3d shape, including the fine-scale details of the stitches. In Figure 7, we depict the results for a very challenging object, i.e. a carved wodden plate that exhibits a continuous variation of albedo owing to the textured grain of the wood used. In such a case, any approach that relies on segmenting albedo into constant albedo regions will fail. In our case, we are able to accurately recover the highly varying albedo pattern and we can see that our 3d shape reconstruction is not affected at all by this albedo variation. We also note that our reconstruction works despite the fact that this wodden plate is moderately specular. Finally, in Table 5, we summarise the properties of all the objects that we have accurately reconstructed and presented as results in this paper. We believe that the wide variation of the properties of these objects is evidence for the efficacy of our 3d reconstruction approach.

Object	Fig. Num.	Material	Multi-Albedo	Reflectance
Sweater	2(a)	Wool	✗	Lambertian
Tagore	2(b)	Terracotta	✗	Lambertian
Kali	3	Painted-Terracotta	✓	Specular
Ganesh	4(b)	Plastic	✓	Highly Specular
Jacket	6(a)	Wool	✓	Lambertian
Shorts	6(b)	Cotton	✓	Lambertian
Shoe	6(c)	Mixed	✓	Mildly Specular
Plate	7	Wood	✓	Moderate Specular

Table 1. Applicability of our method on a wide range of objects with varying albedo and other material properties. Our method works accurately in all these cases.

6. Conclusions

In this paper, we present a novel approach of refining the depth estimates of multi-albedo objects. Our method is able to factorize the observed brightness matrix into appropriate illumination, albedo and surface normal components. Our approach works without requiring any pre-calibrated illumination configuration or segmentation of albedo and we have presented an extensive set of results to demonstrate its accuracy for a wide variety of objects.

Acknowledgments

8

References

[1] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general unknown lighting. International Journal of Computer Vision, 72(3):239–257, 2007.

[2] P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille. The bas-relief ambiguity. International journal of computer vision, 35(1):33–44, 1999.

[3] M. J. Brooks. Two results concerning ambiguity in shape from shading. In AAAI, pages 36–39, 1983.

[4] M. Chandraker. Bas-relief ambiguity. In Computer Vision, pages 43–46. Springer, 2014.

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

[6] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '97, pages 369–378, 1997.

[7] 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.

[8] A. Ecker and A. D. Jepson. Polynomial shape from shading. In Computer Vision and Pattern Recognition (CVPR), pages 145–152. IEEE, 2010.

[9] Y. Han, J. Lee, and I. Kweon. High quality shape from a single rgb-d image under uncalibrated natural illumination. In International Conference on Computer Vision (ICCV), pages 1617–1624. IEEE, 2013.

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

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

[12] P. W. Holland and R. E. Welsch. Robust regression using iteratively reweighted least-squares. Commun. Statist. Theory Methods, 9, 1977.

[13] B. K. P. Horn. Robot vision. MIT press, 1986.

[14] B. K. P. Horn. Obtaining Shape from Shading Information. MIT Press, 1989.

[15] K. Ikeuchi. Computer Vision: A Reference Guide. Springer, 2014.

[16] Z. Lin, M. Chen, and Y. Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. arXiv preprint arXiv:1009.5055, 2010.

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

[18] J. Park, S. N. Sinha, Y. Matsushita, , Y. Tai, and I. Kweon. Multiview photometric stereo using planar mesh parameterization. In International Conference on Computer Vision (ICCV), pages 1161–1168. IEEE, 2013.

[19] R. Szeliski. Computer Vision: Algorithms and Applications. Springer-Verlag, 2010.

[20] K. Wilson and N. Snavely. Robust global translations with 1dsfm. In Computer Vision–ECCV 2014, pages 61–75. Springer, 2014.

[21] R. J. Woodham. Photometric method for determining surface orientation from multiple images. Optical engineering, 19(1):191139–191139, 1980.

[22] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma. Robust photometric stereo via low-rank matrix completion and recovery. In Computer Vision–ACCV 2010, pages 703–717. Springer, 2011.

[23] L. Yu, S. K. Yeung, Y. Tai, and S. Lin. Shading-based shape refinement of RGB-D images. In Computer Vision and Pattern Recognition, pages 1415–1422. IEEE, 2013.

[24] Q. Zhang, M. Ye, R. Yang, Y. Matsushita, B. Wilburn, and H. Yu. Edge-preserving photometric stereo via depth fusion. In Computer Vision and Pattern Recognition (CVPR), pages 2472–2479. IEEE, 2012.

[25] 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.

[26] T. Zhou and D. Tao. Godec: Randomized low-rank & sparse matrix decomposition in noisy case. In International Conference on Machine Learning (ICML), pages 33–40, 2011.