Photometric Refinement of Depth Maps for Multi-albedo Objects

Avishek Chatterjee, Venu Madhav Govindu
Indian Institute of Science, Bengaluru 560012, India.

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 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.

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. Then for a Lambertian surface, the brightness (B_{ip}) at the pixel p in the i-th image is given by

$$B_{ip} = \lambda_i \mathbf{l}_i{}^T \mathbf{n}_p \alpha_p = \mathbf{L}_i{}^T \mathbf{N}_p \tag{1}$$

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. If we have k lighting conditions and m pixels, then we 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

$$\mathbb{B} = \mathbb{L}^T \mathbb{N} \tag{2}$$

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 shadow, specularity, inter-reflection etc. cause $\mathbb{B}$ to have rank more than 3. We first robustly estimate a rank 3 factorization of the observed brightness matrix using an iterative reweighting method as

$$\mathbb{B} \approx \mathbf{V}\mathbb{X} \tag{3}$$

This factorization is not unique. A general factorization of $\mathbb{B}$ can be written as

$$\mathbb{B} \approx \mathbf{V}\mathbf{A}\mathbf{A}^{-1}\mathbb{X} \tag{4}$$

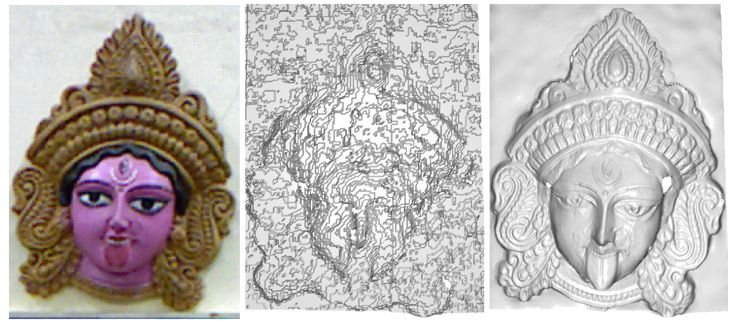
where $\mathbf{A}$ can be any non-singular 3×3 matrix. Thus from Equations 2 and 4 we can say that there exists an $\mathbf{A}$ such that

$$\mathbb{L} = \mathbf{A}^T \mathbf{V}^T \tag{5}$$
$$\mathbb{N} = \mathbf{A}^{-1}\mathbb{X} \tag{6}$$

Hence our problem is one of estimating the requisite matrix $\mathbf{A}$ which allows us to recover $\mathbb{N}$. We solve for $\mathbf{A}$ using the raw depth map from the 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 $\widehat{\mathbf{N}} = \begin{bmatrix} \widehat{\mathbf{n}}_1 & \cdots & \widehat{\mathbf{n}}_m \end{bmatrix}$.

For a surface with constant albedo, we can assign the relative albedo $\alpha_p = 1, \forall p$. This allows us to replace $\mathbb{N}$ in Equation 6 with the Kinect's estimated normal matrix $\widehat{\mathbf{N}}$. Subsequently, we solve for $\mathbf{A}$ by robustly solving the linear system of equations given by Equation 6.

But the problem becomes challenging for a surface with arbitrarily varying albedo. The norms of the columns of $\mathbb{N}$ i.e. the albedo for individual pixels are no longer equal. The estimated normals from Kinect's depth map i.e. $\widehat{\mathbf{N}}$ are not scaled by the albedos of the corresponding pixels. Thus we can not simply replace $\mathbb{N}$ in Equation 6 with the Kinect's estimated normal matrix $\widehat{\mathbf{N}}$. Therefore, we solve for $\mathbf{A}$ in the following manner.



(a) Color image (b) Raw scan (c) Our reconstruction

Figure 1: Refined depth map estimates of a painted terracotta plaque of Kali using our approach.

Let $\mathbf{x}_p$ denote the normalized p-th column of $\mathbb{X}$, i.e. corresponding to the p-th pixel. Then we have

$$\mathbf{A}\alpha_p\widehat{\mathbf{n}}_p \parallel \mathbf{x}_p \Rightarrow \mathbf{A}\widehat{\mathbf{n}}_p \parallel \mathbf{x}_p \quad \forall p \tag{7}$$

where the symbol $\parallel$ denotes that the vectors $\mathbf{A}\widehat{\mathbf{n}}_p$ 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

$$e_p = \left\| \mathbf{x}_p - \frac{A\widehat{\mathbf{n}}_p}{\|A\widehat{\mathbf{n}}_p\|} \right\| \tag{8}$$

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

$$\mathbf{A}^* = \underset{\mathbf{A}}{\operatorname{argmin}} \sum_{p=1}^{m} \rho\left(e_p\right) \tag{9}$$

where $\rho(\cdot)$ is a robust cost function. The optimization problem of Equation 9 is solved using gradient descent. The approach of solving for $\mathbf{A}$ is shown to be convergent.

Given the Kinect's depth map and estimated surface normals, we fuse them to obtain a high quality reconstruction of the imaged surface. In our work, we use a depth-normal fusion method similar to the one originally proposed in "Efficiently combining positions and normals for precise 3d geometry" by Nehab *et. al.* in ACM Transactions on Graphics (TOG), 2005.

We experimentally demonstrate the value of our approach by presenting highly accurate three-dimensional reconstructions of a wide variety of objects. As an illustration, in Figure 1 we show the depth estimates obtained using this approach for a painted terracotta plaque of the Hindu goddess Kali (Figure 1). As shown in Figure 1(b) the raw Kinect depth maps are highly noisy and the resultant refinement obtained using our method (Figure 1(c) is quite substantial.

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 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.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.