

Near Light Correction for Image Relighting and 3D Shape Recovery

Xiang Huang^{*}, Marc Walton^{*}, Greg Bearman[†] and Oliver Cossairt^{*}

^{*}Northwestern University, Evanston, IL, USA

[†]ANE Image, Pasadena, CA, USA

xianghuang@gmail.com

Abstract—In this paper, we propose a near-light illumination model for image relighting and 3D shape recovery. Classic methods such as used by the popular RTI software from Cultural Heritage Imaging assume that lighting is infinitely far away from the scene. However, this constraint is impossible to achieve in practice: light sources cannot be too far away from the scene due to space and power constraints. This causes non-uniform illumination artifacts due to variations in the distance between the light source and points in the scene. We correct this effect to provide much more uniformly lit images that yield more appealing image for relighting applications. Furthermore, we use our near-light model for more accurate photometric stereo calculations of surface normals, eliminating the “potato-chip” shaped surface reconstruction error that results from violating the far-light assumption. We verify our model with both free-form capture using hand-held flash as the illumination source, and capture using LED lights mounted on a dome shaped surface.

Keywords—*Reflectance Transformation Imaging (RTI), Polynomial Texture Map (PTM), HSH, Near Light Calibration, Photometric Stereo, Shape from Shading, 3D Surface Shape Reconstruction, Image Relighting*

I. INTRODUCTION

The appearance of a work of art results from the physical interaction of light with its constituent materials. To formally capture a complete record of appearance one should measure all the essential components of its light-transport function: the total interaction of the materials and microstructure comprising a work of art with all possible incoming/illuminating light rays, measured by all possible outgoing/observable light rays [1]. Embedded in the light-transport function are each fixed fraction of incident light the artwork will absorb, reflect, refract, scatter, and transmit from its surface, for all possible incident-light locations, directions, wavelengths, and polarizations. It also contains the directional fractions of incident light that will scatter beneath the object’s surface and re-emerge in different directions from various neighboring points.

While these measurements are theoretically possible, the enormous size of such data and the time required to capture them is largely impractical. As a result, the light-transport function is typically undersampled by taking only a few images of an artwork with varying lighting conditions rather than directly measuring all the possible interactions between light and object. One example of such a technique is Reflectance Transformation Imaging (RTI), originally known as Polynomial Texture Mapping (PTM). Malzbender [2] first proposed RTI as a way to interactively change the lighting conditions of a digital image. By interpolating multiple images of an object, each with different illumination angles from a fixed camera position, an ‘active photo’ may be produced with easy controls that encourage exploration to see vanishingly-subtle features.

In the last decade the art conservation community has become increasingly interested in using RTI for more closely

examining artworks through relighting. As a relighting technique, RTI has provided visually compelling ways to interactively explore surface relief and discover subtle surface features otherwise missing or indiscernible in ordinary photos or by direct visual inspection [3], [4]. The freely available viewer software from Cultural Heritage Imaging (CHI) [5] can also exaggerate surfaces, pixel-by-pixel, to depict the topography more clearly, and to compute estimates of surface normal vectors via photometric stereo or from the PTM interpolation equation itself. However, while current RTI methods offer conservators a powerful exploratory tool, the many systematic approximations inherent to the technique limit its use to qualitative assessments of appearance.

Quantitative determination of surface normals could be very useful for art conservation. For instance, surface normals measured over time can detect and map shape changes to an object within a museum environment. More accurate surface normals might help determine the sequences of brush strokes or ink application to a work of art to help understand how an artist made the object. As one step towards quantitative surface normal estimations, here we address a fundamental limitation of the RTI model; the assumption that the whole object is lit from the same illumination angle with the same illumination intensity across the entire field of view. This requirement is rarely met in real-life experimental conditions because the light would need to be placed infinitely far away from the object – and can’t be satisfied due to space constraints of most workspaces, and a very powerful light to illuminate from such a large distance. The mismatch between the lighting model and real experimental conditions has been documented to produce erroneous surface normal estimations, a “potato-chip” shape estimation error when the surface normals are integrated, and non-uniform illumination effects in relighting that we call the ‘spot-light’ effect [6], [7], [8].

A. Contributions

In this paper, we present an approach to remove dependence on this far light assumption. Using our technique we relight images and perform surface reconstructions using lighting in close proximity to the object, either a matrix of LED lights mounted to a small half meter dome or a handheld flash light source positioned about one meter away. Our new method offers the following contributions:

- Our method allows lighting direction estimation *without the need for a mirror ball placed in the scene*, as implemented in most in RTI experimental setups. Our method requires just a flat matte surface (approximately Lambertian) such as a color-checker calibration target to be placed in the scene.
- Our method estimates not only the *2D* lighting direction but *the true 3D position of light sources*.
- By estimating 3D light positions, we can more accurately solve for the surface normal and albedo for

all points in the scene. With our near-light model, the recovered 3D surface shape is exempt from the “potato-chip” shape errors encountered in traditional RTI captures.

- By compensating each pixel’s image intensity according to its distance to lighting (often called ‘flat-fielding’) we remove the spot-light effect (brighter-center, darker-border) often encountered in RTI captures.
- We provide freely available software written in Matlab that implements our algorithm on image sets captured for RTI to create standard RTI files that can be viewed in the RTI viewer freely available from CHI [5].

II. PREVIOUS WORK

There are two computation methods for representing RTI images using the CHI RTI Builder and Viewer software suites [5]. The polynomial texture mapping (PTM) version, originally proposed by Malzbender [2], uses a polynomial basis function for light intensity interpolation, and the hemispherical harmonics (HSH) version [9] reduces the directional bias. Surface normals are calculated differently for each representation. The PTM version fits the pixel intensity to a local bi-quadratic function of light angles and then finds the normal by setting the derivative to zero, which has the effect of finding the direction of the brightest pixel [10]. For the HSH version, three lighting directions are chosen to generate a set of three relit photographs from the fitted HSH fitted data. Conventional photometric stereo is then applied to using the chosen lighting directions and synthesized pixel intensities. CHI software users can create normals maps in RTI Viewer for both computation methods and export those images at JPG files for later use.

Integrating the RTI-derived normals to create a lofted 3D surface results in large scale surface cupping (our “potato-chip” surface) [7], [8], especially at the edges compared to the center. To correct for the large scale bending, MacDonald *et al.* [8] measured additional depths of a few surface points to bound the surface during reconstruction. While this solution is feasible, it requires user intervention and its accuracy is difficult to characterize. Instead, we seek a more principled solution that does not involve further interaction or measurement of the object.

As MacDonald *et al.* [8] originally observed, violation of the far-light assumption will result in normals that are inaccurate for very low spatial frequencies and this situation is quite common for RTI capture. As shown in Figure 1, a typical RTI freeform capture set up can easily mis-estimate light angles with errors that span 19 degree from the left image border to the right. Errors in the estimated light direction will produce incorrect normal estimation, which will in turn produce errors in integrated depth. In this paper, we correct these errors by using a near-light model.

Photometric stereo is a well established computer vision technique that often used to recover surface shape from image intensity. The original formulation by Horn [11] assumed lights are infinitely far away, the camera is orthographic, and the object surface is Lambertian and convex (*i.e.* no shadows or inter-reflections). Since photometric stereo was originally introduced, several researchers have sought to generalize the technique for more practical camera, surface and lighting models. Belhumeur *et al.* [12] discovered that with an orthographic camera model and uncalibrated lighting, the object’s surface can be uniquely determined to within a bas-relief ambiguity. Papadimitri and Favaro *et al.* [13] recently pointed out that this ambiguity is resolved under the perspective camera model.

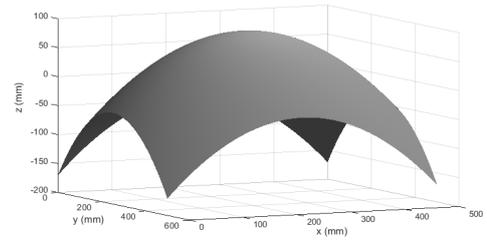


Fig. 1: “Potato-chip” shaped surface calculated from typical RTI data. In this example the object is 50 cm wide, smooth and uniformly flat, and the light is 150 cm above the object. The center of the object has an incident illumination angle of 90 degrees, but the left and right border of the central scan line have incident illumination angles of 80.5 degrees and 99.5 degrees respectively, with an error range of $19 = 2 \times \text{atan}(25/150)$ degrees from left to right. The erroneous lighting angles cause mis-estimated surface normals in RTI, which causes the “potato-chip” shaped surface.

Several researchers have also sought to relax the Lambertian reflectance assumption and incorporate effects such as specular highlights and shadows. New techniques have been introduced based on non-Lambertian reflectance models [14], [15], [16], or sophisticated statistic methods to automatically filter non-lambertian effects [17], [18], [19]. However, less attention has been paid to relaxing assumptions on the lighting model. Several other researchers [20], [21] recently investigated removing the far-light assumption to improve the accuracy of photometric stereo. Others [22] further considers non-isotropic illuminations.

In this paper, we devise a near-light model and introduce a new fast optimization method to solve an energy minimization problem and obtain more accurate light position and surface normal estimates. Our method corrects for inaccuracies in the conventional RTI capture process, producing more accurate relighting results, surface normal estimates, and 3D models from captured data, and yet it removes the troublesome requirement to capture a ‘mirror ball’ image within each artwork photograph.

III. IMAGE INTENSITY FORMATION MODEL OF NEAR LIGHT SOURCE

We consider the physical model of light transport as shown in Figure 2. The light rays are emitted from a source, reflect from an object surface, then finally reach a camera. The pixel intensities measured at the camera sensor depend on three components: the light source, the object shape and reflectance, and the camera pixel’s intensity response and lens vignetting. Each are discussed individually.

Both RTI and conventional photometric stereo assume each light source is infinitely far away, *e.g.* sun light, and illuminates from direction $\hat{l} = (\hat{l}_x, \hat{l}_y, \hat{l}_z)$, where the ‘hat’ atop \hat{l} denotes normalized (unit-length) vector. This far-light assumption is widely used to simplify image formation equations: light incident at any scene point arrives with the same angle and power. However, in practice it is not possible to place lights far from the scene due to space and power constraints. Thus, a near-light model is necessary to more accurately describe the non-uniform light distribution incident on the scene. Given an isotropic point light source with power e at position $\mathbf{l} = (l_x, l_y, l_z)$ (note the ‘hat’-free \mathbf{l} is **not** a unit vector), and each scene point $\mathbf{p} = (x_p, y_p, z_p)$ that illuminates from direction $\mathbf{l} - \mathbf{p}$, the irradiance will fall by the square of distance:

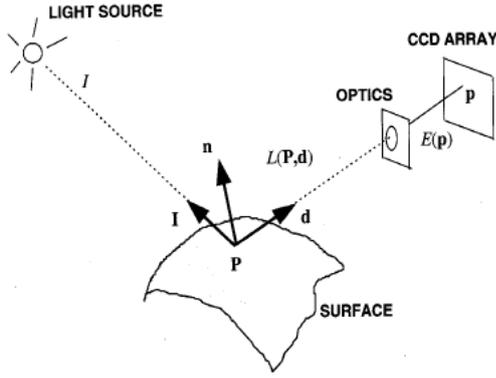


Fig. 2: Light transport from the point light source, reflected by the object surface, and sensed by the camera. Figure courtesy of Trucco and Verri. [23]

$\frac{e}{\|l-p\|^2}$. Unlike the distant light model, the light direction and energy is not uniform across all scene points.

Light incident on an object's surface is reflected according to surface material's bidirectional reflectance distribution function (BRDF). As in conventional photometric stereo, we assume the surface is Lambertian. In the far-light model, reflected light intensity falls by cosine of lighting angle: $R_p = \hat{n} \cdot \hat{l} e a_p = \hat{n}^T \hat{l} e a_p$, where $\hat{n} = (\hat{n}_x, \hat{n}_y, \hat{n}_z)$ denotes the unit-length surface normal and a_p denotes albedo, the reflectivity at point p . In our near-light model, the reflected light energy takes the more complicated form:

$$R_p = \hat{n}_p^T \frac{(l-p)}{\|l-p\|^3} e a_p.$$

We refer to the quantity

$$f(l_k, p) \triangleq \hat{n}_p^T \frac{(l-p)}{\|l-p\|^3} \quad (1)$$

as the light pencil, which depends entirely on the scene geometry, namely the object shape and light position. The light pencil consists of two terms. The first term $\hat{n}_p^T \frac{(l-p)}{\|l-p\|^3}$ represents the cosine result of the dot-product between the light angle and surface normal as in the far-light model. The second component $\frac{1}{\|l-p\|^2}$ describes the squared distance fall off of light energy.

Assuming a linear camera model, the intensity measured at a pixel I_p is proportional to the amount of light reflected at the corresponding point R_p :

$$I_p = \hat{n}_p^T \frac{(l-p)}{\|l-p\|^3} e a_p \eta_p = \hat{n}_p^T \frac{(l-p)}{\|l-p\|^3} e a'_p. \quad (2)$$

The fractional coefficient η_p is due to camera vignetting and depends on pixel location: defines as 1.0 at the center of the image and smaller at the border. Vignetting is partly caused by a natural illumination fall-off known as the "cosine fourth" law: the vignetting coefficient is proportional to the four power of the cosine of the angle between the principal ray and optical axis. Vignetting can also be caused by blocking of lighting rays from apertures within a compound lens, and often stronger for large-aperture and wide-angle lenses. For a fixed camera position, albedo and vignetting terms cannot be not separated, we can only estimate their combined effect $a'_p = a_p \eta_p$.

IV. AN OPTIMIZATION SCHEME FOR NEAR LIGHT POSITION AND SURFACE NORMAL ESTIMATION

The near light model in Equation 2 of Section III describes the pixel intensity given the light position and power, and the object shape and albedo. In this section, we solve the inverse problem: given the observed pixel intensity, compute the light and object parameters using our near-light model.

We gather a series of images with fixed camera position illuminated from different light positions. We do a least-square fitting using the model for pixel intensities from Equation 2 and the measured intensity. A solution for lighting positions is found by minimizing the following energy function for all N pixels in K images:

$$E(l_k, e_k, \hat{n}_p, p, a'_p) = \sum_{p,k} \left(\frac{\hat{n}_p^T (l_k - p)}{\|l_k - p\|^3} e_k a'_p - I_{pk} \right)^2. \quad (3)$$

In the energy minimization, we have $N \cdot K$ observed pixel values, and $4K + 3N$ independent parameters to solve. More specifically, we need to solve $3K$ parameters for lighting positions l_k , K parameters for lighting intensity e_k , $2N$ parameters for surface normals with unit-norm \hat{n}_p , and N parameters for combined albedo and vignetting a'_p . Note that the 3D surface position p can be integrated from estimated surface normals \hat{n}_p [24]. In the case of captured rgb color images, we have $3N \cdot K$ observed pixel values and $6K + 5N$ independent parameters ($3K$ instead of K parameters for rgb light intensity, $3N$ instead of N parameters for rgb albedo times vignetting).

Directly solving the energy minimization problem defined in Equation 4 is generally prohibitive, as it is non-convex with millions of parameters $4K + 3N$, where the pixel number N for most modern cameras is in the range of 10 million. We use an alternating minimization formulation that iteratively solves two subproblems: one small scale non-convex optimization to estimate light position given albedo and normal, and another linear least squares optimization to find albedo and normal given 3D light position. The steps are as follows:

- 1) Given the combined albedo and vignetting a'_p , surface normal n_p and 3D scene point p , we calibrate light position l_k and power e_k .
- 2) Given the lighting position l_k , power e_k , and 3D scene point p , we compute the albedo a'_p and normal n_p . The albedo and normal can be computed for each pixel by solving a least squares problem similar to conventional photometric stereo:

$$\arg \min_{a'_p, \hat{n}_p} \sum_{p,k} \left(a'_p \hat{n}_p^T \frac{l_k - p}{\|l_k - p\|} - I'_{pk} \right)^2, \quad (4)$$

where the distance compensated pixel intensity

$$I'_{pk} = I_{pk} \frac{\|l_k - p\|^2}{e_k}. \quad (5)$$

After computing the normal, the surface shape p can be integrated from the normal using the method from Agrawal *et al.* [24].

- 3) We iteratively solve problems 1) and 2) similar to the work by Papadimitri *et al.* [21].

The details of performing step 2) are similar to conventional photometric stereo. Step 3) is similar to the iterative procedure introduced by Papadimitri *et al.* [21]. In step 1),

Papadimitri *et al.* [21] propose to initialize the scene as a planar object $n_p = [0, 0, 1]$ with constant effective albedo $a'_p = 1$. They perform an initial brute force search within a one meter cubic volume with a 10cm step to determine light positions in order to get an initial estimate for the light positions. However, we found in practice that this method tends to fail because the effective albedo $a'_p = a_p \eta_p$ isn't constant due to vignetting, which results in an erroneous estimation of light position.

Instead, we propose a new optimization procedure for estimating the 3D light positions that eliminates dependence on albedo and vignetting (*i.e.* the term a'_p in Equation 4). Rather than performing a brute force search for 3D lighting positions, we derive an objective function that can be solved using gradient decent or quasi-Newton's optimization method and is found to be robust to initialization value. As a result, our light estimation is much more robust to sources of noise and modeling error common in RTI captures. To derive our new objective function, we first transform Equation 2 for the point p under lighting k to

$$\begin{aligned} a'_p &= I_{pk} \frac{\|\mathbf{l}_k - \mathbf{p}\|^3}{\hat{\mathbf{n}}_p^T (\mathbf{l}_k - \mathbf{p})} e_k^{-1} \\ &= I_{pk} g(\mathbf{l}_k, \mathbf{p}) e_k^{-1}, \end{aligned}$$

where the inverse of light pencil $f(\mathbf{l}_k, \mathbf{p})^{-1}$ is defined as

$$g(\mathbf{l}_k, \mathbf{p}) \triangleq \frac{\|\mathbf{l}_k - \mathbf{p}\|^3}{\hat{\mathbf{n}}_p^T (\mathbf{l}_k - \mathbf{p})}, \quad (6)$$

which purely depends on the geometric properties of the scene: the light position and object shape.

Equation $a'_p = I_{pk} g(\mathbf{l}_k, \mathbf{p}) e_k^{-1}$ holds for all $k = 1, 2, \dots, K$ lights for a given point p . Those K measurements are equal in theory, but slightly different from each other in practice due to noise and modeling error. We find the best lighting and surface shape parameters that minimize the variance of the expected a'_p from each lighting instance k , for each pixel, by minimizing the following objective function:

$$\begin{aligned} D &= \frac{1}{N} \sum_p \frac{\sum_k \left(I_{pk} g(\mathbf{l}_k, \mathbf{p}) e_k^{-1} - \frac{1}{K} \sum_m I_{pm} g(\mathbf{l}_m, \mathbf{p}) e_m^{-1} \right)^2}{\sum_k (I_{pk} g(\mathbf{l}_k, \mathbf{p}) e_k^{-1})^2} \\ &= 1 - \frac{1}{NK} \sum_p \frac{\left(\sum_k I_{pk} g(\mathbf{l}_k, \mathbf{p}) e_k^{-1} \right)^2}{\sum_k (I_{pk} g(\mathbf{l}_k, \mathbf{p}) e_k^{-1})^2}. \end{aligned} \quad (7)$$

The variance of a'_p is normalized by its total power to eliminate bias towards small albedo values and closer lights. The above objective function $D(\mathbf{l}_k, e_k, \hat{\mathbf{n}}_p, \mathbf{p})$ has minimum value of 0 and maximum value of $1 - 1/K$. It reaches a minimum of 0 when $i_{pm} g(\mathbf{l}_m, \mathbf{p}) e_m^{-1} = i_{pn} g(\mathbf{l}_n, \mathbf{p}) e_n^{-1}$ for all $m, n \in 1, 2, \dots, K$. It reaches a maximum of $1 - 1/K$ when the ratio $I_{pk} g(\mathbf{l}_k, \mathbf{p}) e_k^{-1} / \max(I_{pk} g(\mathbf{l}_k, \mathbf{p}) e_k^{-1})$ for all $k \in 1, 2, \dots, K$ has $K - 1$ zeros, which occurs when one of the lights is at actually at finite distance from the scene, but erroneously computed to be a very far distance away.

With the new objective function, the light position and intensity can be updated by the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton optimization method, with the gradient as

$$\begin{aligned} \frac{\partial D}{\partial \mathbf{l}_k} &= \frac{2}{NK} \sum_p Q_{pk} \left(\frac{3(\mathbf{l}_k - \mathbf{p})}{\|\mathbf{l}_k - \mathbf{p}\|^2} - \frac{\hat{\mathbf{n}}_p}{\hat{\mathbf{n}}_p^T (\mathbf{l}_k - \mathbf{p})} \right). \\ \frac{\partial D}{\partial e_k^{-1}} &= \frac{2}{NK} \sum_p Q_{pk} e_k. \\ Q_{pk} &\triangleq \left(\frac{\sum_m I_{pm} g(\mathbf{l}_m, \mathbf{p}) e_m^{-1}}{\left(\sum_m (I_{pm} g(\mathbf{l}_m, \mathbf{p}) e_m^{-1})^2 \right)^{1/2}} I_{pk} g(\mathbf{l}_k, \mathbf{p}) e_k^{-1} - 1 \right) \\ &\quad \times \frac{\sum_m I_{pm} g(\mathbf{l}_m, \mathbf{p}) e_m^{-1}}{\left(\sum_m (I_{pm} g(\mathbf{l}_m, \mathbf{p}) e_m^{-1})^2 \right)^{1/2}} I_{pk} g(\mathbf{l}_k, \mathbf{p}) e_k^{-1} \end{aligned} \quad (8)$$

In theory, all pixels in the captured images can be used to perform this optimization. In practice, we use only a small number (generally less than five percent of all pixels), generated by downsampling the image. As a result, 3D light estimation can be performed in only seconds on a personal desktop with Intel®Xeon®E5-1650 3.5 GHz CPU.

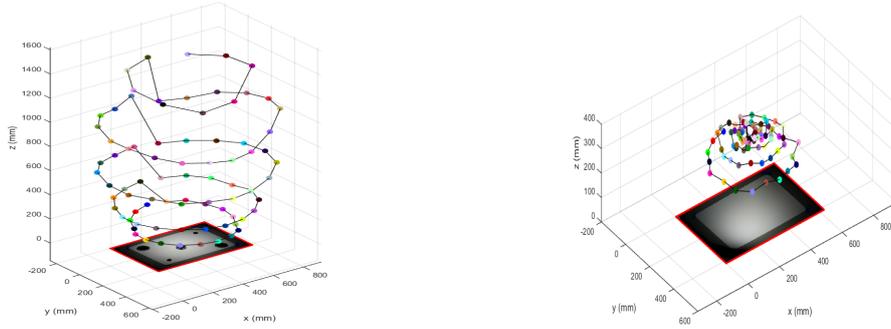
Note that our new framework still requires initial values for the combined albedo and vignetting and surface shape in step 1), and light position and power in step 2). We initially assume the scene is flat, so that $\hat{\mathbf{n}}_p = (0, 0, 1)^T$, and $p = (u, v, 0)$ where (u, v) are pixel coordinates. We have found that our algorithm converges for a large variety of initial light positions. In all our experiments, we initialized all lights to have the same power and position $\mathbf{l}_k = [w/2, h/2, \max(w, h)]$, where w, h are the width and height of the image in pixels. While our algorithm does not explicitly require surface shape to be known ahead of time, in practice, many scenes often have a relatively flat surface as background that can be used to further constrain the lighting optimization.

V. EXPERIMENTAL RESULTS

We now show how our new near light model can improve the results of image relighting in RTI and also the surface normal reconstruction. Calibrated light positions are shown in Figure 3 for a lighting dome with Canon 5D Mark III camera and a prime lens of 20mm. For reference, comparisons are shown using triangulation with multiple spherical balls placed in the scene. The 3D light positions for the dome are captured faithfully with our technique while significant errors are present when triangulating with multiple mirror balls.



Fig. 4: (a) Freeform capture (b) dome capture



(a) Calibrated light position from 5 mirror spheres (b) Our calibrated light position from one piece of paper

Fig. 3: A Comparison of 3D light position estimation using triangulation from multiple mirror balls (left) and the method proposed in this paper (right). In the left figure, the light positions of 81 dome lights obtained by least square error triangulation from five mirror spheres placed in the center, top left, top right, bottom left, bottom right of the scene. In the right figure, the light positions of the same setup is estimated from just a piece of white matte printing paper. The light positions from the mirror balls are subject to large triangulation errors for lights near the top of the dome. Our technique, however, produces highly accurate estimates of 3D light position.

A. Image Relighting using PTM or HSH with Near Light

Once we have accurately estimated the 3D location of light sources during RTI capture, we can use this to generate more accurate relighting results. The PTM or HSH image relighting methods assume images are captured under distant light so that each pixel is lit from same lighting angle with same energy. If images are captured using near lights but the 3D location of light sources is known, it is still possible to generate accurate relighting results using these methods.

The PTM and HSH techniques relight an image under a novel lighting direction by interpolating basis images captured from a fixed camera and a set of known lighting directions. PTM stores six coefficients $\mathbf{c}_p = [c_0, c_1, c_2, c_3, c_4, c_5]^T$ for each pixel, and computes pixel intensity I_p from a novel illumination direction $\hat{\mathbf{l}}$ as $I_p = c_0 l_x^2 + c_1 l_y^2 + c_2 l_x l_y + c_3 l_x + c_4 l_y + c_5 = \mathbf{h}(\hat{\mathbf{l}})^T \mathbf{c}_p$, where the biquadratic polynomial $\mathbf{h}(\hat{\mathbf{l}}) = [l_x^2, l_y^2, l_x l_y, l_x, l_y, 1]$. HSH is similar to PTM, but uses Hemispherical Harmonics functions instead of polynomial functions as basis function $\mathbf{h}(\hat{\mathbf{l}})$. After capturing a set of images with K known lighting direction (K larger than the number of coefficients in \mathbf{c}_p), the coefficients of HSH or PTM can be computed by solving an over-determined linear system of equations:

$$I_p = c_0 l_x^2 + c_1 l_y^2 + c_2 l_x l_y + c_3 l_x + c_4 l_y + c_5 = \mathbf{h}(\hat{\mathbf{l}})^T \mathbf{c}_p$$

$$\begin{bmatrix} \mathbf{h}(\hat{\mathbf{l}}_1) \\ \mathbf{h}(\hat{\mathbf{l}}_2) \\ \vdots \\ \mathbf{h}(\hat{\mathbf{l}}_K) \end{bmatrix} \mathbf{c}_p = \begin{bmatrix} I_{p1} \\ I_{p2} \\ \vdots \\ I_{pK} \end{bmatrix} \quad (9)$$

With near light sources, captured images suffer from a non-uniform illumination artifact that produces a “spot light” effect: regions of the object further away from the light source will receive weaker illumination and appear darker. For example, in the dome capture setup used in this paper, scene points near the border receive as little as 10% of the illumination received at the center.

To correct for non-uniform illumination, we compute a corrected or ‘re-lit’ image I' so that each pixel I'_p appears to be lit from a distant light source as shown in Figure 5. We

first correct for the distance-squared attenuation to pixels in all K captured images. The corrected images are then used to fit the coefficients of a relighting interpolation function in a similar way to Equation 9. However, unlike PTM or HSH where each pixel has the same interpolation lighting matrix H , we create a pixel dependent H matrix by replacing light direction $\hat{\mathbf{l}}_k$ with $\frac{(\mathbf{l}_k - \mathbf{p})}{\|\mathbf{l}_k - \mathbf{p}\|}$, as shown in Equation 10.

$$\begin{bmatrix} \mathbf{h}\left(\frac{(\mathbf{l}_1 - \mathbf{p})}{\|\mathbf{l}_1 - \mathbf{p}\|}\right) \\ \mathbf{h}\left(\frac{(\mathbf{l}_2 - \mathbf{p})}{\|\mathbf{l}_2 - \mathbf{p}\|}\right) \\ \vdots \\ \mathbf{h}\left(\frac{(\mathbf{l}_K - \mathbf{p})}{\|\mathbf{l}_K - \mathbf{p}\|}\right) \end{bmatrix} \mathbf{c}_p = \begin{bmatrix} I'_{p1} \\ I'_{p2} \\ \vdots \\ I'_{pK} \end{bmatrix} \quad (10)$$

Figure 5 shows a comparison between relighting results with and without our near-light correction for two RTI captures of prints by the artist Paul Gauguin, housed within the permanent collection at the Art Institute of Chicago. The images were captured using our dome illumination, which produces a severe spotlight effect in the raw captured data. After 3D light position estimation and near-light correction is applied, the relit images appear uniformly lit and are generally much more visually pleasing.

B. Surface Normal and Shape Reconstruction

RTI measures only one light direction at a place near the object using a mirror ball, and applies the same light direction for any point of the object; this does not match the near-light model for close illumination. The mismatching light angles cause erroneous surface normal estimates which yield “potato-chip” surface shape when integrated.

To determine the accuracy of our technique in measuring 3D surface shape, we use flat matte paper as a ground truth planar surface. Figure 6 shows comparisons between raw data captured using our lighting dome, and corrected data using our 3D light estimation technique. Before correction, 3D surface reconstructions exhibit a severe bending, resulting in a “potato chip” like appearance due to large global errors in surface shape estimation. After our correction is applied, the 3D surface appears nearly flat, indicating it has been reconstructed with much higher quality.

We further test 3D reconstruction quality for a capture of a woodblock print by Paul Gauguin housed at the Art

Institute of Chicago, as shown in Figure 7. A similar bending is 3D reconstruction without correcting for the near-light effect shows the familiar “potato” chip bending so that surface details are difficult to resolve. However, after applying our near-light correction, the 3D reconstruction of the woodblock appears much flatter. As a result, minute details in the carving of the woodblock can now be discerned from the reconstruction.

VI. CONCLUSION

We have developed a novel automatic method to estimate the 3D location of light sources for captured photometric stereo and RTI data. We have shown how to incorporate our results into a near-light model for more accurate 3D surface reconstructions and show results for both ground truth data, as well as RTI captures of woodblocks and prints from the Art Institute of Chicago’s collection of the work of Paul Gauguin. Our work may have direct impact on cultural heritage imaging applications by providing further steps towards inexpensive and accurate 3D surface measurement techniques that are accessible to curators and conservation scientists around the world. Our 3D light estimation and RTI generation software will be released as a freely available MATLAB package that will integrate seamlessly into the workflow of any curators or conservators currently using RTI software. The input to our software will be a set of photos from a typical RTI capture setup and the output will be an RTI file directly loadable in RTI Viewer software.

VII. FUTURE WORK

In the future, we want to investigate more accurate physical models for light sources. Currently we assume lights are isotropic and light power is emitted uniformly across all angles. In practice, emitted power usually exhibits some degree of angular dependence especially for systems such as the dome lighting used for experiments in this paper. Some recent works model the non-isotropic effect of light sources [25], [22]. We are also interested in developing methods to perform accurate 3D surface estimation using more sophisticated models of material reflectance. In addition, we would like to incorporate non-linear global illumination effects such as shadows and multi-bounce inter-reflections within the scene.

VIII. ACKNOWLEDGEMENT

This project was undertaken at the Northwestern University / Art Institute of Chicago Center for Scientific Studies in the Arts (NU-ACCESS). NU-ACCESS is funded through a generous grant from the Andrew W. Mellon Foundation. Supplemental support is provided by the Materials Research Center, the Office of the Vice President for Research, the McCormick School of Engineering and Applied Science and the Department of Materials Science and Engineering at Northwestern University.

REFERENCES

- [1] K. Kutulakos, “Light Transport Analysis for 3D Photography,” *Sixth International Conference on 3-D Digital Imaging and Modeling (3DIM 2007)*, no. 3dim, pp. 413–418, 2007.
- [2] T. Malzbender, D. Gelb, and H. Wolters, “Polynomial texture maps,” in *Proceedings of SIGGRAPH 2001, Annual Conference Series*. New York, New York, USA: ACM Press, 2001, pp. 519–528. [Online]. Available: <http://portal.acm.org/citation.cfm?doi=383259.383320>
- [3] M. Mudge, T. Malzbender, C. Schroer, M. Lum, R. Art, and R. Fields, “New Reflection Transformation Imaging Methods for Rock Art and Multiple-Viewpoint Display,” in *VAST*, 2006.
- [4] G. Earl, K. Martinez, and T. Malzbender, “Archaeological applications of polynomial texture mapping : analysis , conservation and representation,” *Journal of Archaeological Science*, pp. 1–11, 2010. [Online]. Available: <http://dx.doi.org/10.1016/j.jas.2010.03.009>
- [5] “Cultural Heritage Imaging: Reflectance Transformation Imaging (RTI),” 2013. [Online]. Available: <http://culturalheritageimaging.org/Technologies/RTI/index.html>
- [6] L. Macdonald and S. Robson, “Spatial Calibration of an Illumination Dome Conventional copystand,” Tech. Rep., 2011.
- [7] M. Elfarargy, A. Rizq, M. Rashwan, B. Alexandrina, and P. O. Box, “3D Surface Reconstruction Using Polynomial Texture Mapping,” in *Lectures Notes in Computer Science*, no. 1, 2013, pp. 353–362.
- [8] L. W. Macdonald, “Representation of Cultural Objects by Image Sets with Directional Illumination,” no. 45 cm, pp. 43–56, 2015.
- [9] S. Y. Elhabian, H. Rara, and A. a. Farag, “Towards accurate and efficient representation of image irradiance of convex-Lambertian objects under unknown near lighting,” *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1732–1737, 2011.
- [10] G. Palma, M. Corsini, P. Cignoni, R. Scopigno, and M. Mudge, “Dynamic shading enhancement for reflectance transformation imaging,” *Journal on Computing and Cultural Heritage*, vol. 3, no. 2, pp. 1–20, 2010.
- [11] B. K. P. Horn, “Obtaining shape from shading information,” *The Psychology of Computer Vision*, pp. 115–155, 1975. [Online]. Available: <http://dl.acm.org/citation.cfm?id=93877>
- [12] P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille, “The Bas-Relief Ambiguity,” *IJCV*, vol. 35, no. 1, pp. 33–44, 1999.
- [13] T. Papadhimetri and P. Favaro, “A new perspective on uncalibrated photometric stereo,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1474–1481, 2013.
- [14] a. Hertzmann and S. Seitz, “Shape and materials by example: a photometric stereo approach,” *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 1, pp. 1–8, 2003.
- [15] N. Alldrin, T. Zickler, and D. Kriegman, “Photometric stereo with non-parametric and spatially-varying reflectance,” *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2008.
- [16] D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz, “Shape and spatially-varying BRDFs from photometric stereo,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 6, pp. 1060–1071, 2010.
- [17] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma, “Robust photometric stereo via low-rank matrix completion and recovery,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6494 LNCS, pp. 703–717, 2011.
- [18] S. Ikehata, D. Wipf, Y. Matsushita, and K. Aizawa, “Robust photometric stereo using sparse regression,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, no. 1, 2012, pp. 318–325.
- [19] M. Zhang, “Robust surface normal estimation via greedy sparse regression,” Ph.D. dissertation, 2014. [Online]. Available: <http://summit.sfu.ca/item/13719>
- [20] A. Wetzler, R. Kimmel, A. M. Bruckstein, and R. Mecca, “Close-Range Photometric Stereo with Point Light Sources,” in *2014 2nd International Conference on 3D Vision*, 2014, pp. 115–122. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7035816>
- [21] T. Papadhimetri, P. Favaro, and U. Bern, “Uncalibrated Near-Light Photometric Stereo,” in *Proceedings of the British Machine Vision Conference*, 2014, pp. 1–12.
- [22] Y. Quéau and J.-d. Durou, “Some Illumination Models for Industrial Applications of Photometric Stereo,” 2015, p. QCAV.
- [23] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*. Prentice Hall PTR, 1998.
- [24] A. Agrawal, R. Raskar, and R. Chellappa, “What is the Range of Surface Reconstructions from a Gradient Field ?” in *ECCV*, 2006, pp. 578–591.
- [25] J. Park, S. N. Sinha, and Y.-w. Tai, “Calibrating a Non-isotropic Near Point Light Source using a Plane,” in *CVPR*, 2014, pp. 2267–2274.

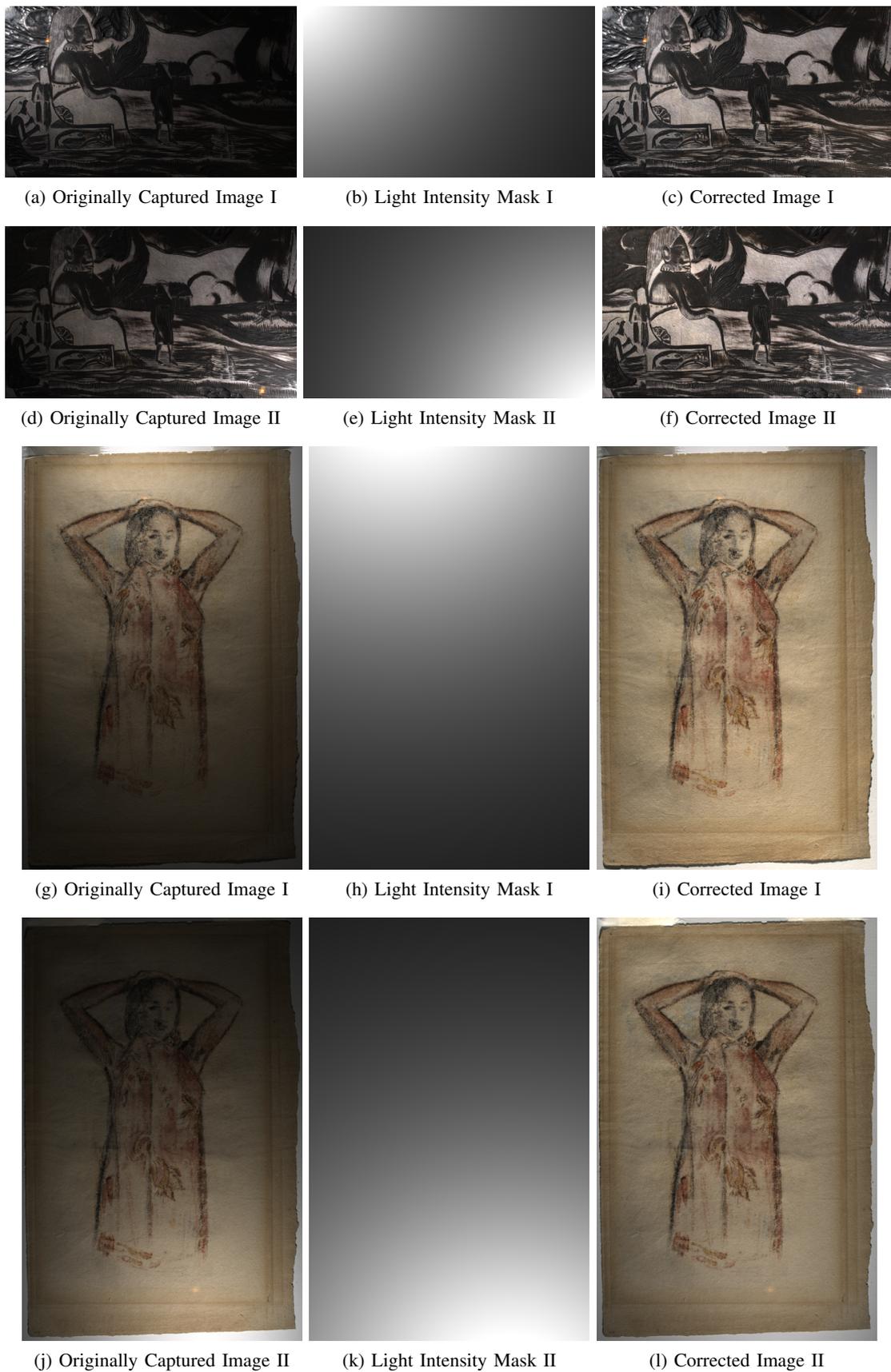


Fig. 5: Relighting comparisons for two works by Paul Gauguin housed at the Art Institute of Chicago, a woodblock (top), and a transfer print (bottom). The concession number for the woodblock is 1940-91, and 2002-237 for the print. A comparison is shown between the raw captured images (left) and after the near-light correction technique introduced in this paper (right). We use the calibrated light position to compute the light attenuation due to the distance squared fall-off. The inverse of this attenuation mask (middle) is used to produce relit images with even illumination (right). The corrected images look uniformly lit and more visually pleasing.

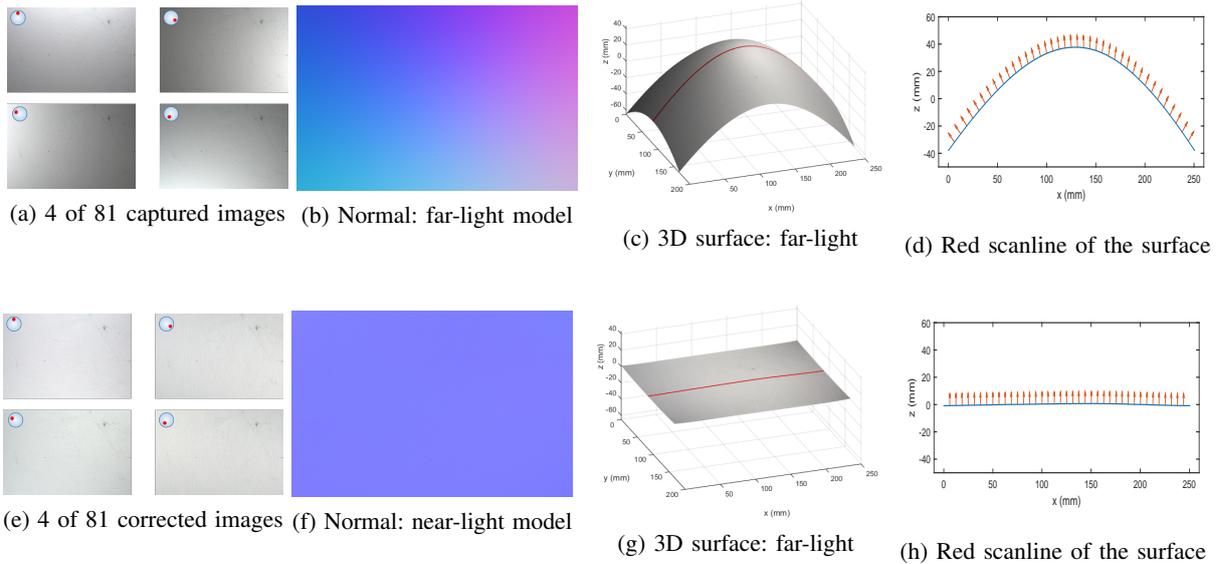


Fig. 6: Experimental comparison between far-light and near-light models for surface normal and 3d shape reconstruction. We use a flat matte paper as a ground truth test object. The first column shows the captured and corrected images. The position of the red dot in the circle approximates lighting direction. The second column shows the surface normal. The ground truth should be uniform, but the normal from the distant light model has a large error especially at the border. The third column shows the reconstructed 3d surface. The last column shows a horizontal scan line of the recovered depth map. The near-light model results are close to ground truth, while the far-light model results in a large potato chipped distortion.

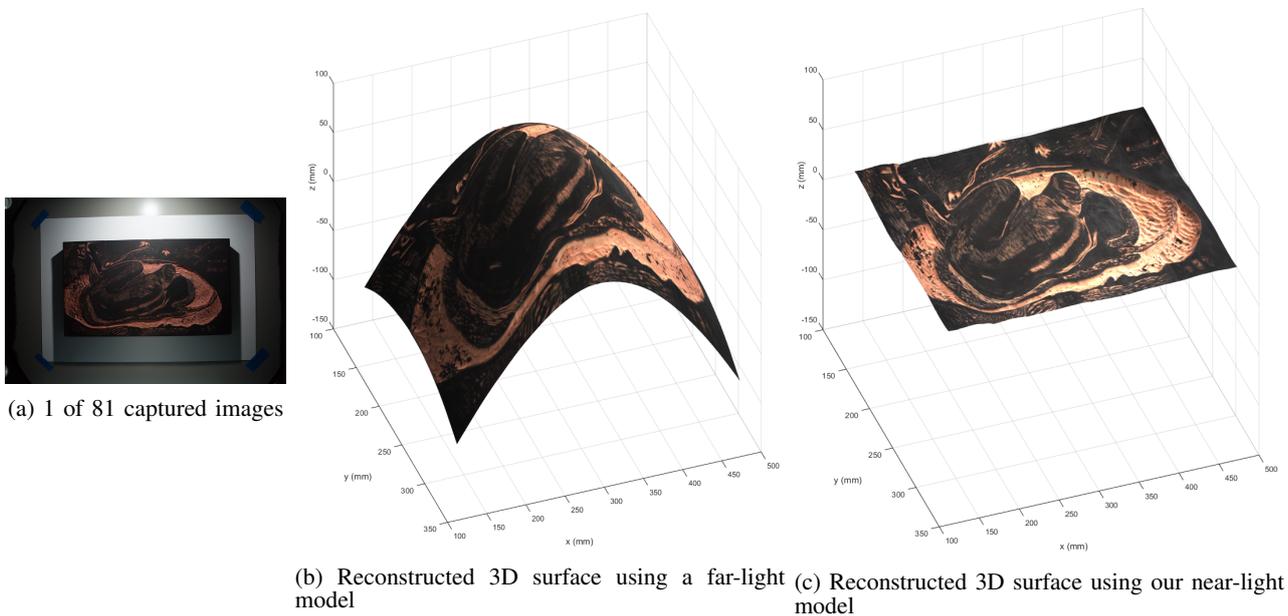


Fig. 7: Comparison of far-light and near-light models for 3d shape reconstruction. The work of art shown is a woodblock produced by artist Paul Gauguin, housed at the Art Institute of Chicago. The concession number of the artwork is 2238-56. (a) One of the captured images showing the near-light illumination effect. (b) A 3D reconstruction without correcting for the near-light effect. The woodblock appears to be bent like a “potato chip” so surface details are difficult to resolve. (c) A 3D reconstruction using our method to correct for the near-light effect. The woodblock is now flat and details of the carving can be discerned from the reconstruction.